

Research Article

SCRD-Net: A Deep Convolutional Neural Network Model for Glaucoma Detection in Retina Tomography

Hua Wang ^{1,2,3,4} Jingfei Hu ^{1,2,3,4} and Jicong Zhang ^{1,2,3,4}

¹School of Biological Science and Medical Engineering, Beihang University, Beijing, China

²Hefei Innovation Research Institute, Beihang University, Hefei, China

³Beijing Advanced Innovation Centre for Biomedical Engineering, Beihang University, Beijing, China

⁴Beijing Advanced Innovation Centre for Big Data-Based Precision Medicine, Beihang University, Beijing, China

Correspondence should be addressed to Jicong Zhang; jicongzhang@buaa.edu.cn

Received 22 February 2020; Revised 8 March 2021; Accepted 28 March 2021; Published 10 April 2021

Academic Editor: Roberto Natella

Copyright © 2021 Hua Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Early and accurate diagnosis of glaucoma is critical for avoiding human vision deterioration and preventing blindness. A deep-neural-network model has been developed for the diagnosis of glaucoma based on Heidelberg retina tomography (HRT), called “Seeking Common Features and Reserving Differences Net” (SCRD-Net) to make full use of the HRT data. In this work, the proposed SCR-Net model achieved an area under the curve (AUC) of 94.0%. For the two HRT image modalities, the model sensitivities were 91.2% and 78.3% at specificities of 0.85 and 0.95, respectively. These results demonstrate a significant improvement over earlier results. In addition, we visualized the network outputs to develop an interpretation of the learned mechanism for discriminating glaucoma and normal images. Thus, the SCR-Net can be an effective diagnostic indicator of glaucoma during clinical screening. To facilitate SCR-Net utilization by the scientific community, the code implementation will be made publicly available.

1. Introduction

Glaucoma is a chronic clinical eye disease that can lead to irreversible vision loss [1]. According to Quigley and Brnman [2], glaucoma is the second-leading cause of blindness worldwide, and it disproportionately affects women and Asians. Furthermore, the number of people (aged 40–80 years) with glaucoma worldwide was 64.3 million in 2013, and this number is expected to increase to 79.6 million in 2020 and 111.8 million in 2040 [3]. The emergence of pathological tissues in the optic nerve head (ONH) can lead to the dysfunction of the retinal ganglion cells (RGC) [4]. To prevent this irreversible damage, early diagnosis of glaucoma is essential.

Heidelberg retina tomography (HRT) has been widely used to diagnose glaucoma by optic disc topography measurements and parameter calculations [5]. Glaucoma can be clinically confirmed by an ophthalmologist according to qualitative and quantitative indicators for the pathological

structural features of the ONH and the retinal nerve fiber layer (RNFL) in HRT data [6]. On the one hand, the qualitative evaluation indicators include the contour of the neuroretinal rim, optic disc hemorrhage, parapapillary atrophy, barred circumferential vessels, and RNFL appearance. On the other hand, quantitative evaluation indices include the optic disc size (or vertical disc diameter), cup/disc ratio (in the vertical direction), rim/disc ratio, and retinal nerve fiber layer height (RNFLH).

Conventional machine learning methods have already been developed to improve the diagnostic performance of computer-aided glaucoma detection systems. For example, Wollstein et al. [7] used Moorfields regression analysis (MRA) for ONH evaluation in glaucoma conditions. However, this analysis of the ONH region requires a manual contour annotation, which can have large errors and impose a heavy workload on doctors. While the calculation of the glaucoma probability score (GPS) [8] does not require manual annotation, the GPS-based methods produce limited

features and show low effectiveness in nonstandard ONH areas. The shape abnormality score (SAS) [9] represents a glaucoma diagnosis method that uses wavelets for feature extraction, performs Gaussian process optimization, and employs a support vector machine (SVM) for classifying HRT images. Although the SAS method shows good diagnostic performance, it shows low sensitivities of 0.59 and 0.77 for specificities of 0.95 and 0.85, respectively. Asaoka et al. [10] combined multiple HRT parameters using a random-forest classifier to diagnose glaucoma and achieved an area under the curve (AUC) of 0.96, outperforming the MRA and GPS methods. Bizios et al. [11] demonstrated that the input parameters of optical coherence tomography (OCT) for measuring the retinal nerve fiber layer thickness (RNFLT) have a large impact on the glaucoma diagnostic performance. Errors and limitations may result from the dependence of these input parameters on the image quality and doctor's expertise [7]. Particularly, in most conventional methods, a doctor manually extracts features and classifies samples, and hence, the diagnosis is subject to the limitations of expert knowledge and skills [12–14].

Recently, deep learning methods have been successfully applied in multiple medical imaging applications [15–19]. These methods can effectively capture latent representations and ultimately improve detection and classification outcomes. Furthermore, deep learning methods eliminate diagnostic variations and improve diagnosis through earlier detection and, in turn, early treatment. For detecting ophthalmic diseases, recent studies focused on the segmentation and classification of the optic disc (or cup) patterns [20–22]. Numerous classification schemes have been proposed for diabetic retinopathy [23, 24], hemorrhage detection [25], optic disc abnormality detection [26], age-related macular degeneration (AMD) [27, 28], diabetic macular edema (DME) [29], multiclassification [30, 31], and other conditions [32–34]. Most studies on deep learning for glaucoma diagnosis have exploited certain features of digital fundus images [35–39], such as the visual field [40]. The robustness of these approaches is not fully verified. Chen et al. [41] presented a simple deep convolutional neural network (DCNN) framework to capture hidden glaucoma-related features. For the two HRT image modalities, the framework resulted in AUC values of 0.831 and 0.887, respectively. Improved network design [42] increased the AUC values to 0.838 and 0.898, respectively. Abbas et al. [43] selected regions of interest (ROI) from the green channel of fundus images, extracted features from these ROIs, processed the features with a convolutional neural network (CNN), and optimized the performance with a deep belief network (DBN) [44]. Through validation on multiple datasets, this model scored, on average, a sensitivity of 84.5%, specificity of 98.01%, accuracy of 99%, and precision of 84%. Nonetheless, this model has not been tested on large-scale glaucoma datasets to confirm its practical applicability. Chai et al. [45] designed a two-branch deep learning architecture that can simultaneously analyze whole retinal images and local optic disc regions.

In this paper, we demonstrate the effectiveness of a deep-learning algorithm for glaucoma detection based on HRT

data and propose a model called “Seeking Common Features and Reserving Differences Net” (SCRD-Net) for early-stage glaucoma diagnosis. The proposed model supports the decisions of clinical diagnosis and alleviates the time and effort spent by doctors on complex and repetitive work.

The main contributions of this work are summarized as follows:

- (i) SCR-Net makes full use of the available data by incorporating two features of HRT data, namely, the light reflection intensity and topographic values. By entering these features separately, as shown in Figure 1, SCR-Net can extract elementary and key information, capture subtle differences, and highlight critical details of the glaucoma and normal images.
- (ii) For the glaucoma detection evaluation on the ORIGA and SCES datasets, SCR-Net achieved sensitivities of 0.783 and 0.912, with corresponding specificities of 0.95 and 0.85. These sensitivity results on the two datasets are, respectively, better by 0.193 and 0.142 compared with other state-of-the-art methods. Meanwhile, the AUC achieved by SCR-Net is 0.05 higher than those of previous methods.
- (iii) We visualize the salient areas in the original retinal images to provide evidences for or against a specific class. This visualization illustrates the learned features and helps accelerate the adoption of interpretable black-box deep learning classifiers in medical applications.

2. Methods

2.1. Data Sources. In our experiments, two datasets were used. The first one is from the QEII Eye Care Centre at the Queen Elizabeth II Health Sciences Centre in Halifax, Canada [46]. The dataset contains 1,439 images of glaucoma patients and 560 images of normal subjects. The second dataset was collected at Queen's Medical Centre in Nottingham, UK [47, 48]. The dataset contains 188 images of glaucoma patients and 882 images of normal subjects. In the two datasets previously mentioned, 106 glaucoma patients and 76 normal subjects were collected for both left and right eyes, while others only collected one eye. For the sake of unification, each participant chose only one eye in our experiment. Therefore, from the two datasets, we selected only 1,521 images with varying degrees of glaucoma and 1,366 normal images for the experimental. In addition, in order to better verify the generalization ability of the proposed model, we divided this collection into randomly selected training, validation, and test sets with 60%, 10%, and 30% of the images, respectively.

While the collected data consists of measurements of light reflection intensity and topographic values, the data comes from multiple centers with different qualities, and some measurement parts were not collected due to machine limitations.

As previously mentioned, the HRT data contains two types of features [49]: light reflection intensity and

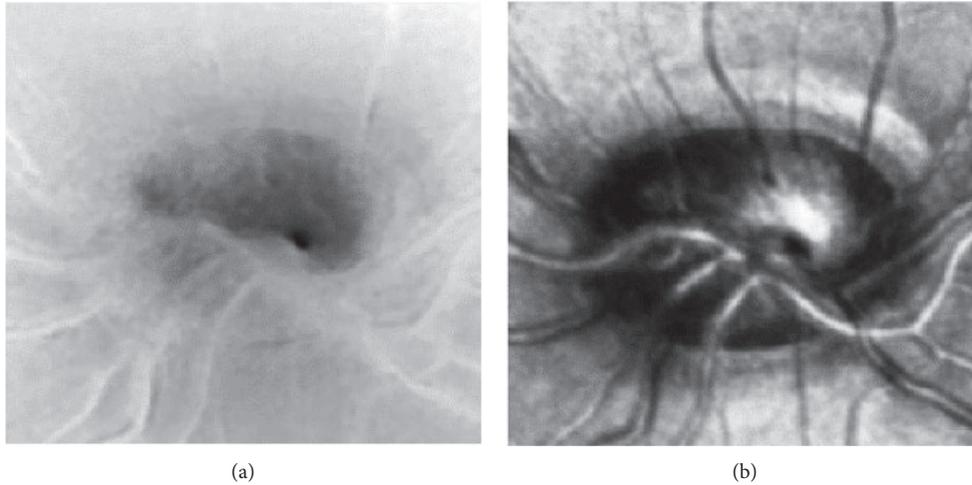


FIGURE 1: HRT data for a glaucoma patient: (a) topographic data and (b) reflection data.

topographic values. The topographic values indicate the height of the fundus ONH and other relevant information of the surrounding terrain, such as the light position, height, and gradient. Figure 1(a) shows the location and morphological information of the optic nerve fibers and blood vessels around the optic disc, while Figure 1(b) shows the light reflection intensity, indicating the light absorption of the fundus retinal reflex.

2.2. Data Preprocessing. We use the ophthalmic picture quality screening method based on the mean pixel height standard deviation (MPHSD) [50] to select the highest-quality HRT images with MPHSD values smaller than 50. Image normalization makes the data more uniform, and the model becomes more robust to multicenter data variations. We replaced the missing fundus field values with the mean value of the collected fundus field data. This data completion step efficiently exploits the CNN advantages to overcome data deficiency due to instrument limitations. We also augmented the data samples, though only in the horizontal direction, by mainly considering the eye alignment.

2.3. Model. Many medical image processing techniques use highly effective deep learning models such as CNNs to assist in the diagnosis of diseases. In this work, we combine several classical CNN models with Xavier initialization to train and test a deep learning architecture for glaucoma diagnosis from HRT data. The proposed architecture is called “Seeking Common Features and Reserving Differences Net” (SCRD-Net). A block diagram of this architecture is shown in Figure 2.

2.3.1. Parallel Processing of Underlying Characteristics. The proposed model was designed while taking into consideration the differences between natural and medical images. Indeed, for medical images and especially HRT images, we would need to retain low-level information. This requirement is satisfied as follows. As shown in Figure 3, a

single-modality network considers as an input either the reflectance or the topographical image components of the HRT data. A merge network processes the two types of components together. A split-merge model involves two independent parallel structures in the bottom layers (blue blocks in Figure 3) for the reflectance and topographical image components of the HRT data. The split-merge model not only avoids the information loss of single-modality networks that use images of a single HRT feature but also reduces the modality confusion that can occur in merge networks. The split-merge model facilitates the extraction of separate reflectance and morphological features of the optic nerve fibers and blood vessels around the optic disc, where the topographical features indicate the height of the fundus optic nerve and the gradient information. The features extracted from both modalities are indispensable for glaucoma diagnosis, and our model well recognizes such features for both modalities.

2.3.2. Seeking Common Ground. After extracting low-level information of both reflectance and topographical features in the bottom layers of the split-merge network model, we fuse the two types of features in the middle layers, as shown by the yellow blocks of the split-merge model in Figure 3. In this work, we mainly use the pretrained inception [51] deep model because its multiscale feature extraction and fusion operations are well suited for the characteristics of different tissue scales in the ONH region. These operations are also comparable to human cognition processes like extracting common features between two objects such as a desk and a coffee table, including the numbers of legs and surfaces. The process of extracting the common features of different modalities that can be attributed to the same class also agrees with doctors’ diagnosis of glaucoma.

2.3.3. Reserving Differences. The fusion of low-level information in the middle layers is effective at extracting common contributions between different HRT features. However, this

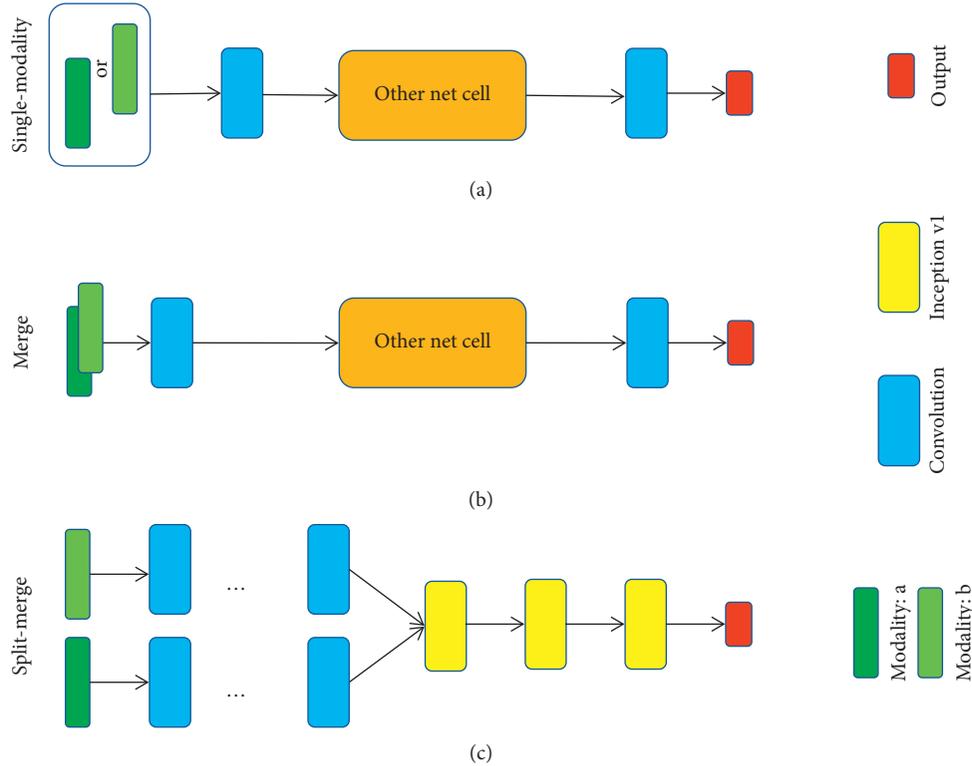


FIGURE 2: Structures of three kinds of deep networks. (a) Singlemodality network: data samples of a single feature are processed; (b) merge network: data samples of two features are processed together; (c) split-merge network: data samples of two features are processed separately in the bottom layers and jointly in the middle layers.

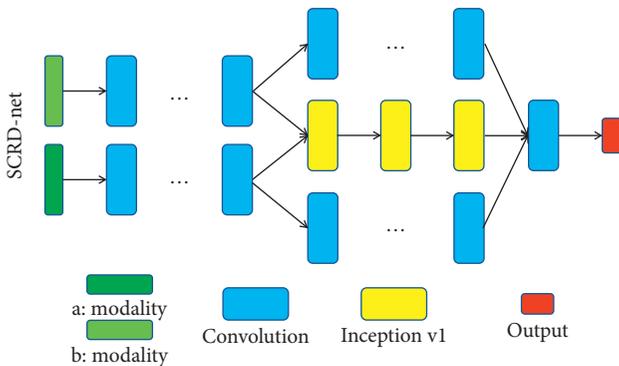


FIGURE 3: Overall architecture of the proposed SCRD-Net.

fusion ignores some individual contributions to the details. Therefore, as shown in Figure 2, in addition to the split-merge model of Figure 3, we enforce the low-level information of the reflectance and topographical image modalities to flow into the top layer of the model as output features. The weighted parameter distribution (in the network equations) of the three parts (i.e., the split-merge path, reflection path, and topographical path) enhances the model sensitivity to glaucoma images, which is a clinically desirable property. In addition, the low-level features that are parallel to the structural features comprehensively improve the network capabilities. Moreover, the blue part is half as deep as the yellow part. Therefore, the blue part can also transfer

the gradient and accelerate the convergence. This design is considered more suitable for small samples and is similar to the auxiliary loss in GoogLeNet [51].

2.4. Highlighted Area Analysis. DCNNs are generally black-box classifiers, and we cannot know which factors play a decisive role in forming the final network outcome. However, in medical diagnosis, it is extremely important to know the image regions that play such a role because different pathological regions require different clinical treatments. To ensure the clinical applicability of our method, we investigate here the visualization of our predictions that result as the response of a deep neural network to a specific input. When classifying images, our visualization method highlights areas in a given input image that provide evidence for or against a certain class. This visualization method is based on [44]. Specifically, this method assumes high dependence of a pixel's value on neighboring pixels. Furthermore, the conditional probability of a pixel, given its neighborhood, does not depend on the position of the pixel in an image. Therefore, the conditional probability $P(c|x_i)$ of a feature x_i is approximated using the marginal distribution $P(x_i)$, where x_i represents a small patch of the image with the center and the data belongs to class c . Then, the instant-specific theory given by Robnik-Šikonja and Kononenko [53] is used to obtain the decision weights in the prediction model. Thus, the region that makes the largest probability contribution is visualized, and we can see where the main

deciding factors have been concentrated. Consequently, we can infer the area where the pathological tissues are most likely located.

3. Experiments and Results

3.1. Evaluation Criteria. In this work, the performance of the glaucoma detectors is evaluated using the metrics of accuracy (ACC), precision (Pr), and area under the receiver operating characteristics (ROC) curve (AUC). The ROC curve highlights the tradeoff between the sensitivity (SE) or the true positive rate (TPR) and the specificity or the true negative rate (TNR). These measures are defined as follows:

$$\begin{aligned} \text{TPR} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{TNR} &= \frac{\text{TN}}{\text{TN} + \text{FP}}, \end{aligned} \quad (1)$$

where TP and TN are the numbers of the true positives and the true negatives, respectively, while FP and FN are the numbers of the false positives and the false negatives, respectively. 0.05-SE denotes the sensitivity at a specificity of 0.95, and 0.15-SE denotes the sensitivity at a specificity of 0.85. This sensitivity to high specificity is necessary to meet clinical needs.

3.2. Performance Indicators and Comparison Baseline. We compared SCRD-Net with several related traditional methods [7–9] and mainstream deep networks in the ImageNet competition [51, 54–57, 60], as shown in Table 1. It can be seen that our proposed model has achieved the best results among all compared methods on all evaluation criteria except 0.15-SE. In particular, the sensitivity at a specificity of 0.95 has increased by 17.1% compared with the latest model. Parallel processing of individual modalities and multimodality fusion can help in capturing common features while realizing differences. Furthermore, the two side paths in the top network layers can easily transmit middle layer information to the top layers and reserve particular features from the two modalities. Experimental results demonstrate that the proposed network structure is suitable and efficient for extracting features from retinal images.

In addition, we have drawn ROC curves of different methods, as shown in Figure 4. It is obvious from the figure on the left that our method is superior to MRA, GPS, and SAS. From the figure on the right, we can see that the SCRD-Net shows the best sensitivity when the specificity is greater than 0.9. When the specificity is less than 0.9, the curves of SCRD-Net are relatively flat. Therefore, not only can SCRD-Net effectively extract the common features between the different modalities, but it can also obtain individual features.

3.3. Ablation Experiments. We input data as shown in the models of Figure 3. The results show that SCRD-Net not only has high accuracy but also improves the sensitivity at a

TABLE 1: Comparison of our model with state-of-the-art network models.

	Method	ACC	AUC	0.05-SE	0.15-SE	SE
<i>Traditional</i>	MRA [7]	—	—	—	0.650	—
	GPS [8]	—	0.800	0.420	0.610	—
	SAS [9]	—	0.890	0.590	0.770	—
<i>Deep learning</i>	Googlenet [51]	0.860	0.925	0.694	0.857	0.846
	Resnet [54]	0.846	0.895	0.635	0.835	0.862
	Densnet [55]	0.857	0.925	0.590	0.895	0.851
	Vgg-16 [56]	0.859	0.913	0.698	0.888	0.808
	Alexnet [57]	0.843	0.896	0.644	0.813	0.879
	Efficientnet [60]	0.858	0.929	0.612	0.924	0.898
<i>Ours</i>	SCRD-Net	0.884	0.940	0.783	0.912	0.910

specificity of 0.95. This high sensitivity is necessary to meet clinical needs.

As summarized in Table 2, if we consider HRT data of only one modality, the results are roughly similar to those obtained using the SAS method. As shown in Table 2, when the light reflection intensity or topographic values are processed, we obtain sensitivities of only 47.4% and 53.4%, respectively, for a specificity of 0.95. From the detection results using each of the reflection intensity and topographic values, we can observe that the sensitivities at specificities of 0.95 and 0.85 are complementary. Therefore, we need to consider combining the data of the two modalities. Indeed, we merged the light reflection intensity and the topographic values as two channels of an RGB image. Compared with the above mentioned experiments, this channel merging greatly improves performance. In particular, the glaucoma detection sensitivity increased by more than 10% at a specificity of 0.95. This is consistent with the fact that ophthalmologists consider the images of both modalities during diagnosis. However, if we consider the two modalities as two channels of an RGB image, we may ignore the individual characteristics of each modality. Therefore, we use the split-merge model to emphasize the independence of the features. Using this model, the sensitivity was improved by an additional 6% at a specificity of 0.85 and by 8.4% at a specificity of 0.95. As shown in Figure 5, we can see that the sensitivity of SCRD-Net is higher than other methods at high specificity. These results show that it is beneficial to treat different model features independently. Moreover, HRT data with multiple modalities plays a crucial role in clinical diagnosis [58]. More importantly, our proposed SCRD-Net model achieved a sensitivity of 78.3% at a specificity of 0.95, demonstrating an improvement of 7.3% when the structural and textural features are combined. Furthermore, the results verify that SCRD-Net can well handle similar multimodal data patterns.

3.4. Highlighted Area Results. In the visualization of the normal and glaucoma-affected retinal images (Figures 6 and 7, resp.), the proposed classifier makes correct predictions for the normal and glaucoma cases as shown by the evidence for and against each decision (marked in red and blue, resp.)

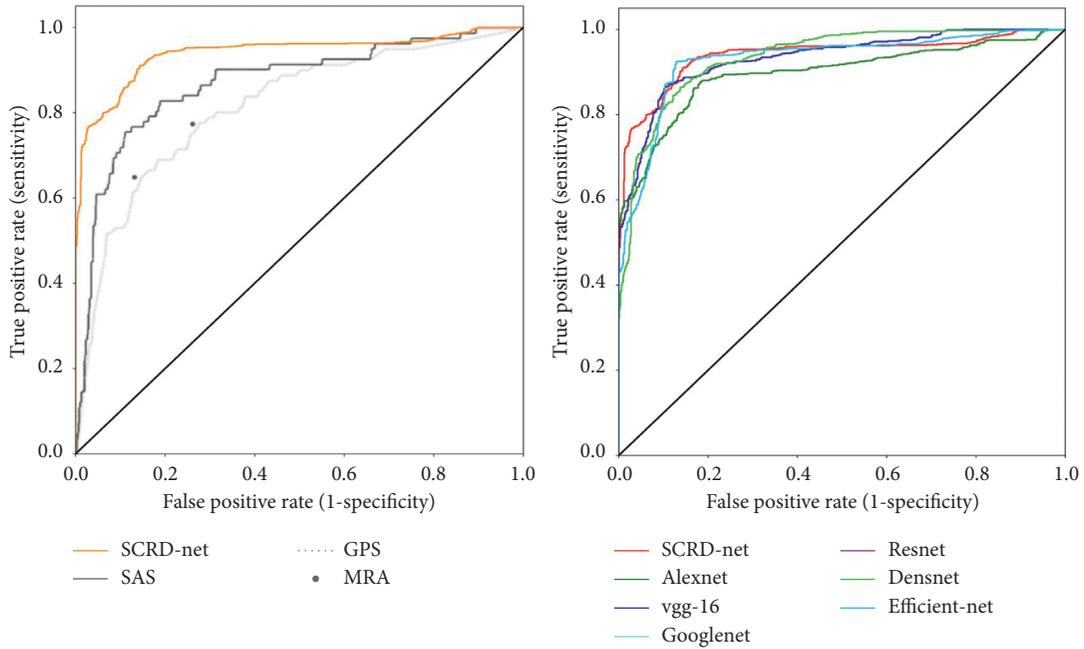


FIGURE 4: Comparison of ROCs of different glaucoma detection methods. Our SCRD-Net model is compared with other methods based on the structure of HRT data, with the aim of achieving a high sensitivity at a specificity of 0.95.

TABLE 2: Comparison of our SCRD-Net method with other methods based on the HRT data composition.

Method	ACC	AUC	0.05-SE	0.15-SE	SE
Topographic	0.828	0.898	0.534	0.795	0.855
Reflection	0.837	0.882	0.474	0.810	0.841
Merge	0.852	0.911	0.650	0.837	0.811
Split-Merge	0.889	0.930	0.710	0.921	0.910
SCRD-Net	0.884	0.940	0.783	0.912	0.920

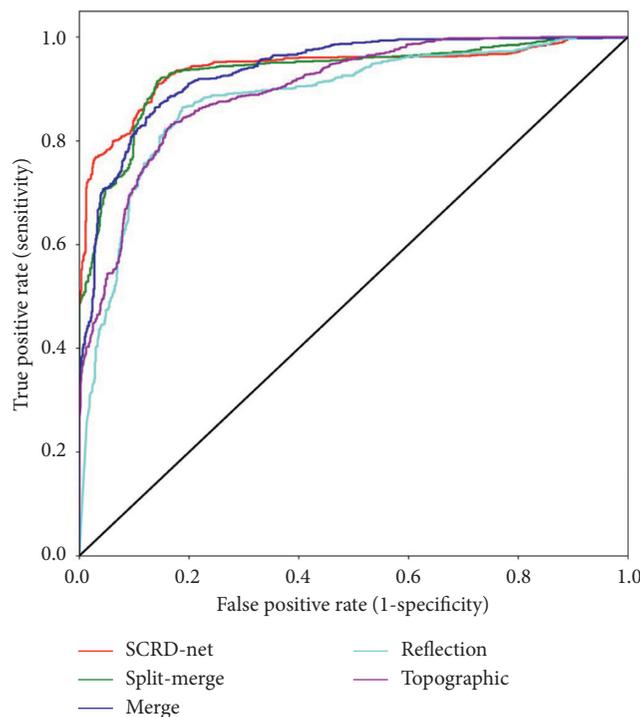


FIGURE 5: Comparison of ROCs of different network models. Our model is compared with different methods based on the structure of HRT data, with the aim of achieving a high sensitivity at a specificity of 0.95.

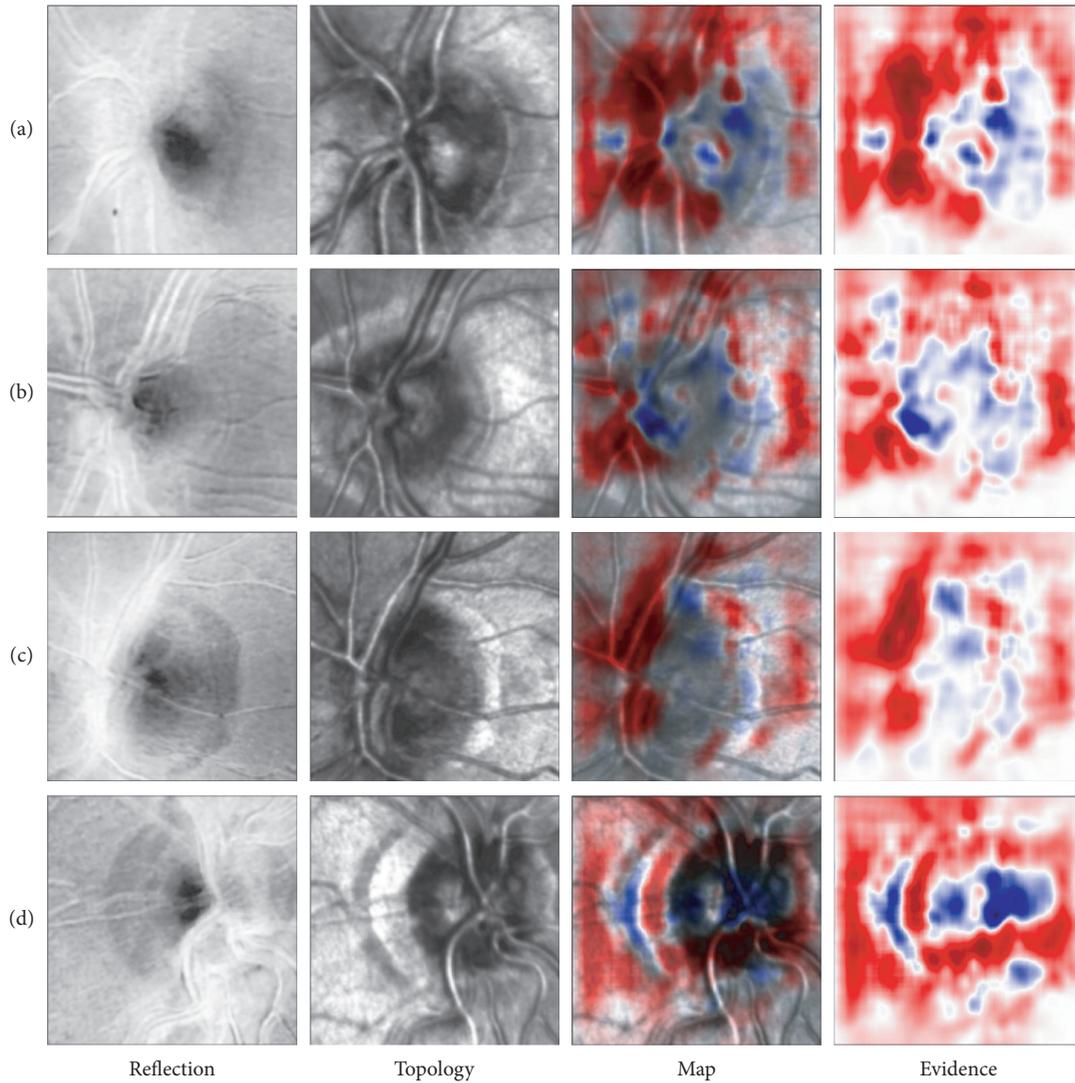


FIGURE 6: Visualization of normal eye patterns in retinal images.

at the output layer. We can see that the supporting evidence (in red) is located around the optic cup for the glaucoma images, whereas the supporting evidence (in red) is located on the optic disc for the normal images.

The red and blue signals differ in strength, size, and location in the retinal images with and without glaucoma. The red signals are more intense and concentrated mostly in the temporal, temporal-superior, or temporal-inferior regions around the ONH or within the ONH in the glaucoma images. The blue signals present a more homogeneous pattern in the ONH and the temporal side. However, these signals are more diverse and even mixed with the red signals in glaucoma images.

Figures 6(a)–6(d) represent four different samples of the normal eye, where (a) is a normal round optic disc with a cup-to-disc ratio of 0.35, and the parapapillary RNFLs are normally distributed. (b) is a normal optic disc with parapapillary atrophy, and the rim width and RNFL are normally distributed. (c) is an eye with a tilted optic disc and a cup-to-disc ratio of 0.5. The rim satisfies the ISNT rule [59], and the

RNFL appears normal. The ISNT rule is as follows. In normal discs, the neuroretinal rim (NRR) is said to have a characteristic configuration with the rim width being the thickest in the inferior region, followed by the superior sector, nasal area, and, finally, temporal sector. (d) is a highly myopic eye with a normal RNFL, and large parapapillary atrophy is present.

Figures 7(e)–7(h) represent four different samples with glaucoma, where (e) is an eye with a large cup-to-disc ratio of about 0.7. There is a significant notch in the inferior-temporal region and a corresponding RNFL loss. (f) is an eye with a large round optic disc and enlarged cup with a cup-to-disc ratio of 0.75. The rim is mildly reduced circumferentially with inferiorly predominant. The RNFLs are thinner in comparison with those of a normal eye. Note that the lamina cribrosa pores are visualized. (g) is a highly myopic eye with glaucoma. The optic disc is tilted and rotated. Prominent parapapillary atrophy is present mainly in the temporal region. The cup-to-disc ratio is 0.75, and the rim is significantly narrow in the inferior quadrant. (h) is

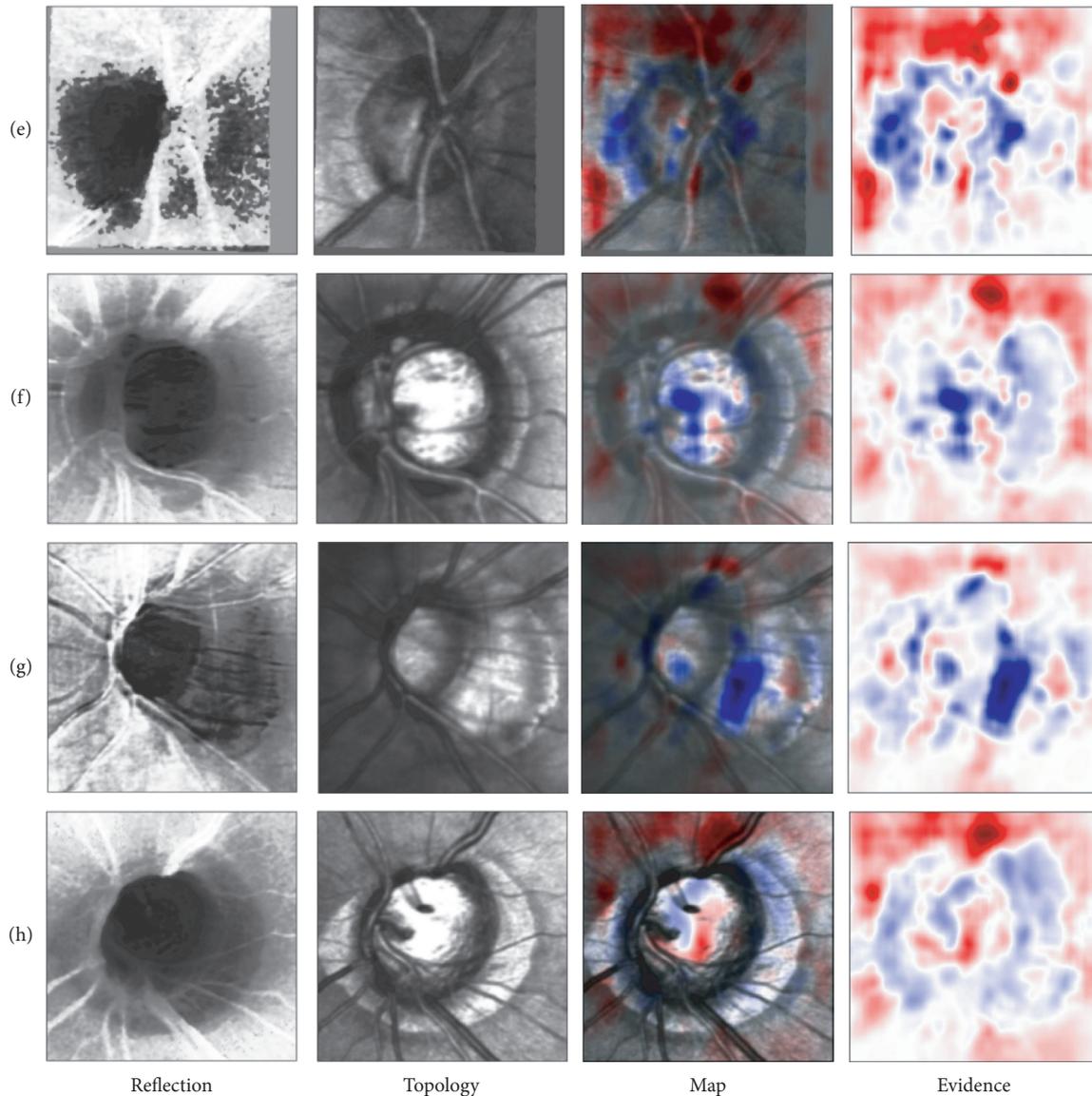


FIGURE 7: Visualization of eyes with glaucoma.

another highly myopic eye with glaucoma. The rim is narrow in every quadrant, and parapapillary atrophy occurred circumferentially. The optic cup is large and deep, and an RNFL defect is present.

4. Discussion

This study presented a framework for the diagnosis of glaucoma based on the inception-V1 model (seeking common ground) that can capture discriminative features in multiple scales to better characterize hidden and key patterns related to glaucoma. We also used two side paths to reserve differences, identify low-level information for the light reflection intensity and topographic values, avoid confusion among the different characteristics, and increase detection sensitivity.

The algorithm for glaucoma and normal image classification is, to a certain extent, similar to some human cognition

processes. For example, when looking at a desk and coffee table, humans will grasp the common morphological information of these two objects, such as the common number of legs. This is an example of global features. Moreover, humans grasp different details, such as the color, height, width, and length of each object to avoid confusion between a desk and a coffee table. These details are examples of local individual features that lead to a high detection sensitivity. Therefore, combining the two aspects of the features can help in better exploitation of the HRT data and understanding the intrinsic differences between glaucoma and normal eye images.

In addition, the low-level information that is hidden (or latent) in the retinal images is important for the glaucoma diagnosis process. These images contain not only high-level geometric shapes that natural images have but also other features including fine texture, border, and contour details. In general, our results show that low-level features are also necessary for the HRT-based diagnosis of glaucoma.

Glaucoma causes vision impairments such as optic neuropathy and visual field defects. Morphologically, internal changes of the ONH might involve disc cupping, retinal rim loss, prelamina tissue loss, thinning and deformation of the lamina cribrosa, changes in the lamina cribrosa pores, and enlargement of the scleral canal opening. In the parapapillary region, impairments include RNFL defects, retinal vessel thinning, and parapapillary atrophy. For mild and moderate glaucoma, defects generally occur in the superior-temporal or inferior-temporal regions. These are characteristic patterns of the disease. Although the signals for differentiation could not be specified to indicate a specific morphological change, signal differentiation might be a potential learning opportunity for identifying glaucoma-induced changes.

5. Conclusions

In this work, we have proposed a glaucoma diagnosis deep learning model called Seeking Common Features and Reserving Differences Net (SCRD-Net) to make full use of the reflectance and topography features of HRT data. The model has two side paths to reserve differences, extract elementary and key information, and capture subtle differences and critical details of both HRT modalities. Based on the experimental results, SCR-Net showed an area under the curve of 94.0%. The sensitivities using the two HRT modalities are 91.2% and 78.3% at specificities of 85.0% and 95.0%, respectively. These results demonstrate a significant improvement over previous results. However, the computations using Telsa K80 in the network are too slow. For future work, we will investigate approaches for accelerating the computations. Furthermore, we plan to probe the pathological basis of glaucoma, especially in the highlighted areas, and then develop a robust computer-aided diagnosis system.

Data Availability

The Heidelberg retina tomography (HRT) data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Beijing Natural Science Foundation (grant no. Z200024); the National Key Research and Development Program of the Ministry of Science and Technology of China (grant no. 2016YFF0201002); the University Synergy Innovation Program of Anhui Province (grant no. GXXT-2019-044); and Hefei Innovation Research Institute, Beihang University.

References

- [1] J. B. Jonas, W. M. Budde, and S. Panda-Jonas, "Ophthalmoscopic evaluation of the optic nerve head," *Survey of Ophthalmology*, vol. 43, no. 4, pp. 293–320, 1999.
- [2] H. A. Quigley and A. T. Broman, "The number of people with glaucoma worldwide in 2010 and 2020," *British Journal of Ophthalmology*, vol. 90, no. 3, pp. 262–267, 2006.
- [3] Y.-C. Tham, X. Li, T. Y. Wong, H. A. Quigley, T. Aung, and C.-Y. Cheng, "Global prevalence of glaucoma and projections of glaucoma burden through 2040," *Ophthalmology*, vol. 121, no. 11, pp. 2081–2090, 2014.
- [4] C. F. Burgoyne, J. Crawford Downs, A. J. Bellezza, J.-K. Francis Suh, and R. T. Hart, "The optic nerve head as a biomechanical structure: a new paradigm for understanding the role of IOP-related stress and strain in the pathophysiology of glaucomatous optic nerve head damage," *Progress in Retinal and Eye Research*, vol. 24, no. 1, pp. 39–73, 2005.
- [5] F. S. Mikelberg, C. M. Parfitt, N. V. Swindale, S. L. Graham, S. M. Drance, and R. Gosine, "Ability of the heidelberg retina tomograph to detect early glaucomatous visual field loss," *Journal of Glaucoma*, vol. 4, no. 4, pp. 242–247, 1995.
- [6] S. Bhartiya, R. Gadia, H. S. Sethi, and A. Panda, "Clinical evaluation of optic nerve head in glaucoma," *Current Journal of Glaucoma Practice with DVD*, vol. 4, pp. 115–132, 2010.
- [7] G. Wollstein, D. F. Garway-Heath, and R. A. Hitchings, "Identification of early glaucoma cases with the scanning laser ophthalmoscope11The authors have no proprietary interest in the development or marketing of this or a competing instrument," *Ophthalmology*, vol. 105, no. 8, pp. 1557–1563, 1998.
- [8] N. V. Swindale, G. Stjepanovic, A. Chin, and F. S. Mikelberg, "Automated analysis of normal and glaucomatous optic nerve head topography images," *Investigative Ophthalmology & Visual Science*, vol. 41, pp. 1730–1742, 2000.
- [9] H. Zhu, A. Poostchi, S. A. Vernon, and D. P. Crabb, "Detecting abnormality in optic nerve head images using a feature extraction analysis," *Biomedical Optics Express*, vol. 5, no. 7, pp. 2215–2230, 2014.
- [10] R. Asaoka, A. Iwase, T. Tsutsumi et al., "Combining multiple HRT parameters using the 'Random Forests' method improves the diagnostic accuracy of glaucoma in emmetropic and highly myopic eyes," *Investigative Ophthalmology & Visual Science*, vol. 55, no. 4, p. 2482, 2014.
- [11] D. Bizios, A. Heijl, J. L. Hougaard, and B. Bengtsson, "Machine learning classifiers for glaucoma diagnosis based on classification of retinal nerve fibre layer thickness parameters measured by Stratus OCT," *Acta Ophthalmologica*, vol. 88, no. 1, pp. 44–52, 2010.
- [12] C. Sinthanayothin, V. Kongbunkiat, S. Phoojaruenchanachai, and A. Singalavanija, "Automated screening system for diabetic retinopathy," *International Symposium on Image and Signal Processing and Analysis*, vol. 2, pp. 915–920, 2003.
- [13] S. Samanta, S. S. Ahmed, M. M. Salem, S. S. Nath, N. Dey, and S. S. Chowdhury, "Haralick features based automated glaucoma classification using back propagation neural network." in *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA)*, Cham, Switzerland, November 2015.
- [14] E. Borque, A. Ferreras, V. Polo, J. M. Larrosa, E. Alias, and F. M. Honrubia, "Evaluation of four new discriminant functions for HRT II in glaucoma diagnosis," *Archivos De La Sociedad Española De Oftalmología*, vol. 83, pp. 349–356, 2008.

- [15] A. Esteva, B. Kuprel, R. A. Novoa et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, p. 115, 2017.
- [16] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougialakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.
- [17] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013: 16th International Conference, Nagoya, Japan, September 22–26, 2013, Proceedings, Part II*, K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, Eds., Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [18] L. Cao, X. J. Li, J. Zhan, and W. F. Chen, "Automated lung segmentation algorithm for CAD system of thoracic CT," *Journal of Medical Colleges of PLA*, vol. 23, pp. 215–222, 2008.
- [19] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.
- [20] H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao, "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation," *IEEE Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1597–1605, 2018.
- [21] R. Srivastava, J. Cheng, D. W. K. Wong, and J. Liu, "Using deep learning for robustness to parapapillary atrophy in optic disc segmentation," in *Proceedings of the IEEE International Symposium on Biomedical Imaging*, New York, NY, USA, April 2015.
- [22] A. Sevastopolsky, "Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network," *Pattern Recognition and Image Analysis*, vol. 27, no. 3, pp. 618–624, 2017.
- [23] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional neural networks for diabetic retinopathy," *Procedia Computer Science*, vol. 90, pp. 200–205, 2016.
- [24] V. Gulshan, L. Peng, M. Coram et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, p. 2402, 2016.
- [25] G. M. Van, G. B. Van, C. Hoyng, T. Theelen, and C. Sanchez, "Fast convolutional neural network training using selective data sampling: application to hemorrhage detection in color fundus images," *IEEE Transactions on Medical Imaging*, vol. 35, pp. 1273–1284, 2016.
- [26] H. S. Alghamdi, H. L. Tang, S. A. Waheeb, and T. Peto, "Automatic optic disc abnormality detection in fundus images: a deep learning approach," in *Proceedings of the Ophthalmic Medical Image Analysis Third International Workshop*, pp. 17–24, Athens, Greece, October 2016.
- [27] C. Lee, A. Rokem, and A. Lee, "Deep learning is effective for the classification of OCT images of normal versus age-related macular degeneration," *Ophthalmology Retina*, vol. 1, no. 4, pp. 322–327, 2017.
- [28] L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search," *Biomedical Optics Express*, vol. 8, no. 5, pp. 2732–2744, 2017.
- [29] M. Awais, H. Müller, T. B. Tang, and F. Meriaudeau, "Classification of SD-OCT images using a deep learning approach," in *Proceedings of the IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, Kuching, Malaysia, September 2017.
- [30] G. Lim, D. S. W. Ting, C. Y. Cheung et al., "Deep learning system for screening of diabetic retinopathy, glaucoma and age-related macular degeneration using retinal photographs: the DEEP EYE Study," *Investigative Ophthalmology & Visual Science*, vol. 58, p. 683, 2017.
- [31] D. S. W. Ting, C. Y.-L. Cheung, G. Lim et al., "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes," *JAMA*, vol. 318, no. 22, pp. 2211–2223, 2017.
- [32] S. T. Devarakonda, K. K. Vupparaboina, A. Richhariya, J. Chhablani, and S. Jana, "Automated detection of retinal disorders from OCT images using artificial neural network," in *Proceedings of the India Conference 2017*, Cambridge, MA, USA, February 2017.
- [33] T. Schlegl, S. M. Waldstein, H. Bogunovic et al., "Fully automated detection and quantification of macular fluid in OCT using deep learning," *Ophthalmology*, vol. 125, no. 4, pp. 549–558, 2018.
- [34] K. Gopinath, S. B. Rangrej, and J. Sivaswamy, "A deep learning framework for segmentation of retinal layers from OCT images," in *Proceedings of the Asian Conference on Pattern Recognition 2017*, Nanjing, China, November 2017.
- [35] B. Al-Bander, W. Al-Nuaimy, M. A. Al-Tae, and Y. Zheng, "Automated glaucoma diagnosis using deep learning approach," in *Proceedings of the International Multi-Conference on Systems, Signals & Devices 2017*, Marrakech, Morocco, March 2017.
- [36] A. Cerentini, D. Welfer, M. D. O. Cordeiro, C. H. Pereira, and G. N. Dotto, "Automatic identification of glaucoma using deep learning methods," *Medinfo 2017 Precision Healthcare Through Informatics*, vol. 245, 2017.
- [37] U. Raghavendra, H. Fujita, S. V. Bhandary, A. Gudigar, J. H. Tan, and U. R. Acharya, "Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images," *Information Sciences*, vol. 441, 2018.
- [38] H. Fu, J. Cheng, Y. Xu et al., "Disc-aware ensemble network for glaucoma screening from fundus image," *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2493–2501, 2018.
- [39] W. Wang, H. Zhu, Y. X. Wang, and D. Garway-Heath, "A transparent deep learning method for diagnosis of early glaucoma with optical coherence tomography," *Investigative Ophthalmology & Visual Science*, vol. 58, p. 4000, 2017.
- [40] H. Muhammad, T. J. Fuchs, N. C. De et al., "Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects," *Journal of Glaucoma*, vol. 26, p. 1, 2017.
- [41] X. Chen, Y. Xu, D. W. K. Wong, T. Y. Wong, and J. Liu, "Glaucoma detection based on deep convolutional neural network," in *Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering In Medicine and Biology Society (EMBC)*, pp. 715–718, Milano, Italy, September 2015.
- [42] X. Chen, Y. Xu, S. Yan, D. W. K. Wong, T. Y. Wong, and J. Liu, "Automatic feature learning for glaucoma detection based on deep learning," in *Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference*, pp. 669–677, Munich, Germany, October 2015.

- [43] Q. Abbas, "Glaucoma-deep: detection of glaucoma eye disease on retinal fundus images using deep learning," *International Journal of Advanced Computer Science & Applications*, vol. 8, 2017.
- [44] G. E. Hinton, S. Osindero, and Y. W. Teh, *A Fast Learning Algorithm for Deep Belief Nets*, MIT Press, Cambridge, MA, USA, 2006.
- [45] Y. Chai, L. He, Q. Mei, H. Liu, and L. Xu, "Deep learning through two-branch convolutional neuron network for glaucoma diagnosis." in *Proceedings of the Smart Health: International Conference, ICSH 2017*, Hong Kong, China, June 2017.
- [46] B. A. Ford, P. H. Artes, T. A. McCormick, M. T. Nicoleta, R. P. Leblanc, and B. C. Chauhan, "Comparison of data analysis tools for detection of glaucoma with the Heidelberg Retina Tomograph," *Ophthalmology*, vol. 110, no. 6, pp. 1145–1150, 2003.
- [47] S. A. Vernon, M. J. Hawker, G. Ainsworth, J. G. Hillman, H. K. MacNab, and H. S. Dua, "Laser scanning tomography of the optic nerve head in a normal elderly population: the Bridlington eye assessment project," *Investigative Ophthalmology & Visual Science*, vol. 46, no. 8, pp. 2823–2828, 2005.
- [48] M. J. Hawker, S. A. Vernon, C. L. Tattersall, and H. S. Dua, "Detecting glaucoma with RADAAR: the Bridlington eye assessment project," *British Journal of Ophthalmology*, vol. 90, no. 6, pp. 744–748, 2006.
- [49] P. H. Artes and D. P. Crabb, "Estimating normative limits of Heidelberg Retina Tomograph optic disc rim area with quantile regression," *Investigative Ophthalmology & Visual Science*, vol. 51, no. 1, pp. 355–361, 2010.
- [50] N. G. Strouthidis, E. T. White, V. M. Owen, T. A. Ho, J. H. Tan, C. J. Hammond et al., "Factors affecting the test-retest variability of Heidelberg retina tomograph and Heidelberg retina tomograph II measurements," *British Journal of Ophthalmology*, vol. 89, no. 11, pp. 1427–1432, 2005.
- [51] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Boston, MA, USA, June 2015.
- [52] L. M. Zintgraf, T. S. Cohen, T. Adel, and M. Welling, "Visualizing deep neural network decisions: prediction difference analysis," 2017, <https://arxiv.org/abs/1702.04595>.
- [53] M. Robnik-Sikonja and I. Kononenko, "Explaining classifications for individual instances," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 5, pp. 589–600, 2008.
- [54] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Vegas, NV, USA, June 2016.
- [55] G. Huang, Z. Liu, L. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, Honolulu, HI, USA, July 2017.
- [56] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, <https://arxiv.org/abs/1409.1556>.
- [57] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proceedings of the International Conference on Neural Information Processing Systems*, pp. 1097–1105, Lake Tahoe, Nevada, December 2012.
- [58] N. E. Benzebouchi, N. Azizi, A. S. Ashour et al., "Multi-modal classifier fusion with feature cooperation for glaucoma Diagnosis," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 31, 2019.
- [59] J. B. Jonas, G. C. Gusek, and G. O. Naumann, "Optic disc, cup and neuroretinal rim size, configuration and correlations in normal eyes," *Investigative Ophthalmology & Visual Science*, vol. 29, pp. 1151–1158, 1988.
- [60] M. Tan and Q. V. Le. EfficientNet, "Rethinking model scaling for convolutional neural Networks," in *Proceedings of the 36th International Conference on Machine Learning*, Long Beach, CA, USA, June 2019.