Automated Glaucoma Diagnosis Using Deep and Transfer Learning: Proposal of a System for Clinical Testing

Mohammad Norouzifard[1], Ali Nemati[2], Hamid GholamHosseini[1], Reinhard Klette[1], Kouros Nouri-Mahdavi[3], and Siamak Yousefi[4]

[1] School of Engineering, Computer, and Mathematical Sciences
Auckland University of Technology (AUT), Auckland, New Zealand
[2] School of Engineering and Technology, University of Washington, Tacoma, USA
[3] Department of Ophthalmology, University of California Los Angeles, Los Angeles, USA
[4] Department of Ophthalmology, University of Tennessee Health Science Center, Memphis, USA

***Abstract*—We developed a deep learning algorithm for identifying glaucoma on optic nerve head (ONH) photographs. We applied transfer learning to overcome overfitting on the small training sample size that we employed. The transfer learning framework that was previously trained on large datasets such as ImageNet, uses the initial parameters and makes the approach applicable to small sample sizes. We then classified the input ONH photographs as "normal" or "glaucoma".**

The proposed approach achieved a validation accuracy of 92.3% on a dataset of 277 ONH photographs from normal eyes and 170 ONH photographs from eyes with glaucoma. In order to re-test the accuracy and generalizability of the proposed approach, we re-tested the algorithm using an independent dataset of 30 ONH photographs. The re-test accuracy was 80.0%.

***Index Terms*—Glaucoma diagnosis, Deep learning, Image classification, Transfer learning, VGG19, Inception-ResNet-V2**

I. Introduction

Glaucoma, the second leading cause of blindness in the world, is a group of optic neuropathy disorders that lead to loss of vision if left untreated [1], [2]. It is estimated that there will be approximately 80 million people worldwide affected by glaucoma by 2020 [1]. In 2010, glaucoma affected more than 2.7 million Americans age 40 and older, which is approximately 2% of the population [4]. Glaucoma is the third cause of blindness in New Zealand [5]. According to the census of the Glaucoma New Zealand website, glaucoma is the leading cause of blindness in New Zealand and it is estimated that approximately 91,000 New Zealanders have the disease but are not aware of it [6].

Because of the rapid increase in aging populations, accurate diagnosis is critical for making treatment decisions to preserve vision and maintain quality of life [3]–[8]. Stereoscopic disc photos provide an appropriate record of the optic nerve, independent of the specialized viewing instrument [5], [6], [9], [11], [15], [16].

Stereoscopic disc photos remain one of the most widely-used and accepted methods for documentation of the optic nerve head [9], [10]. However, due to its subjective nature, assessment of optic disc photographs for presence of glaucoma is labor-intensive and prone to interpretation errors. From a clinical perspective, many eye care specialists prefer to have access to more objective analyses for glaucoma diagnosis. Five rules for assessment of fundus stereo-photographs to identify glaucoma and monitor its progression over time have been described by Fingret et al. [11], [12].

Recent advances in artificial intelligence and a significant growth in available data have enhanced identification of ocular disorders including glaucoma diagnosis. In particular, deep learning techniques can identify highly complex patterns to detect various ocular pathologies [13], [14].

Identifying *glaucomatous optic neuropathy* (GON) based on ONH photographs is one of the standard methods used for glaucoma diagnosis [12]. This process is labor-intensive and biased by reader variations. In this paper, we propose an automated technique based on deep learning and transfer learning that can differentiate between normal eyes and those with glaucoma using ONH photographs. We selected the regions of interest within the ONH photographs, namely regions which included the cup. In fact, the *cup-to-disc ratio* (CDR) is one

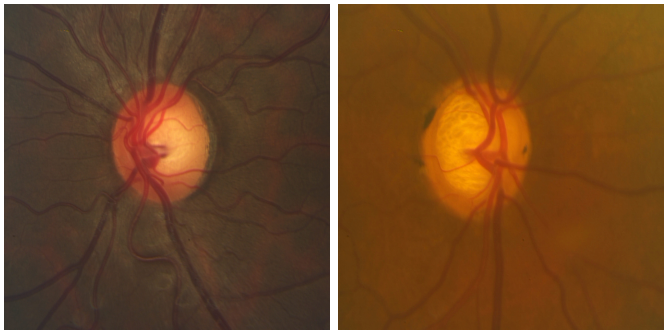


Fig. 1: ONH photographs from UCLA dataset. *Left:* ONH photographs from a abnormal eye, *Right:* ONH photographs from an eye with glaucoma.

major parameter for identifying glaucoma [11]; see Fig. 1.

Convolutional neural networks (CNNs) have been widely used for image segmentation and classification [17]–[20]. Transfer learning is widely implemented in developing deep learning frameworks to address restrictions due to the limited number of input samples as well as computational resources for running deep learning techniques. Transfer learning employs the weights and parameters that were learned from previous large labeled datasets and applies them to the new task [21].

The remainder of this paper is organized into sections as follows: in section II we will describe the datasets, in section III we will explain the deep learning frameworks, Inception-ResNet-V2 and VGG19. In section IV we will present the results and finally, we will conclude the paper in section V.

II. Materials

We used two independent datasets as described below, from universities in the USA and in Germany/Czech Republic.

The first dataset was obtained from patients with normal eyes and those with glaucoma who visited the glaucoma clinic at the *University of California Los Angeles* (UCLA). Therefore, the ONH photographs in this dataset have a diagnostic label of either normal or glaucoma. The collection of these ONH photographs followed the tenets of the Declaration of Helsinki, Health Insurance Portability and Accountability Act guidelines; the Human Research Protection Program approved these studies. Written informed consent was obtained from all study participants.

The UCLA dataset includes 447 fundus images; 170 fundus images from eyes with glaucoma and 277 images from normal eyes. Eyes were defined as glaucomatous or glaucoma suspect if there was evidence of localized or diffuse neuroretinal rim loss or retinal nerve fiber layer loss based on the review of the ONH photographs by a glaucoma specialist (KNM) regardless of the visual field findings. Otherwise, eyes were considered as normal. The IRB at UCLA approved the original study and all the patients consented prospectively. All procedures adhered to the Tenets of the Declaration of Helsinki.

The second dataset is the *high-resolution fundus* (HRF) dataset that is publicly available. This dataset is provided by the pattern recognition laboratory in the Department of Ophthalmology of the *Friedrich-Alexander University Erlangen-Nuremberg* in Germany in collaboration with the Department of Biomedical Engineering at the *Brno University of Technology*, in Czech Republic [22]. This dataset included 15 ONH photographs from the normal eyes and 15 ONH photographs of eyes with glaucoma.

III. Methodology

We use deep learning along with transfer learning to detect glaucoma from ONH photographs. Transfer learning (transferring the pre-trained parameters and weights to a new deep learning model) is a state-of-the-art machine learning technique that is used widely to train deep learning approaches.

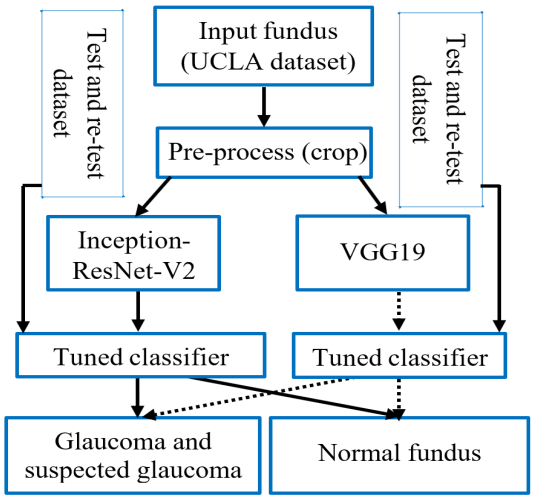


Fig. 2: Overall approach for automated glaucoma diagnosis

Using this approach, deep learning frameworks are trained using a large standard, like ImageNet [26], then weights and parameters are saved to be used for another task. Thus, we transferred the pre-trained parameters as the initial setting of the deep learning framework and then tuned the parameters using the ONH photographs from the UCLA dataset for glaucoma diagnosis.

Basically, it is challenging to train a robust model with a limited number of input images without transfer learning. In another words, transfer learning can serve as prior knowledge for measuring the difference between glaucomatous signs manifested in ONH photographs of eyes with glaucoma and those from normal eyes.

The block diagram in Fig. 2 represents our proposed approach to automatically diagnose glaucoma using ONH photographs. Training and validation data of UCLA dataset were entered as input images. Then, all images were cropped as the pre-processing step. Afterwards, data were fed into VGG19 and Inception-ResNet-V2 models, then hyperparameters (features on deep learning models that should be initialized such as learning rate, and batch size, and so forth) were tuned to have accurate and tuned classifiers. Finally, the test and retest datasets were entered into the tuned classifiers to detect normal and glaucoma eyes.

In this work, we use two deep learning models; VGG19 and Inception-ResNet-V2, which are discussed below.

ONH photographs were entered into both VGG19 and InceptionResNet-V2 models. We kept the initial parameters of these two models and trained them using UCLA ONH photographs. Different layers of these two models extract features (related to glaucoma symptoms) at different resolutions.

We used 447 ONH photographs from the UCLA dataset that included 277 images from normal eyes and 170 images from eyes with glaucoma. We randomly selected 70% of the images for training, 25% for the validation and 5% for testing.

We also used the HRF data for re-testing the model and assessing its generalizability. The region of interest of each

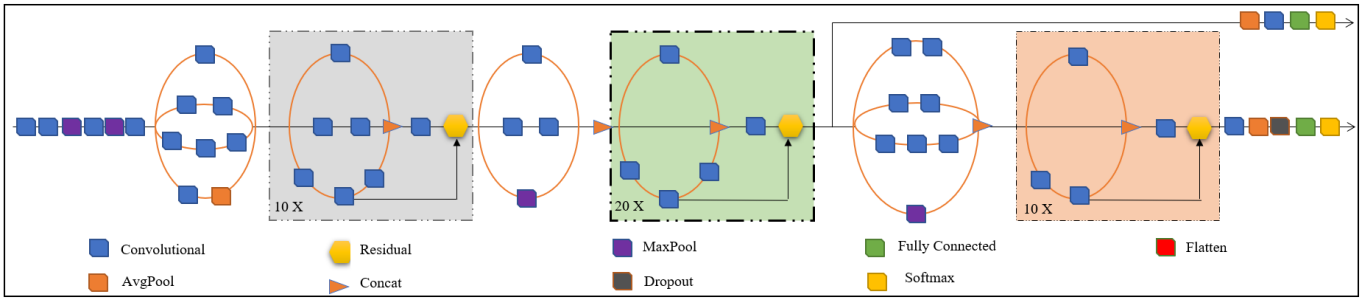


Fig. 3: Schematic block diagram of Inception-ResNet-V2 [25].

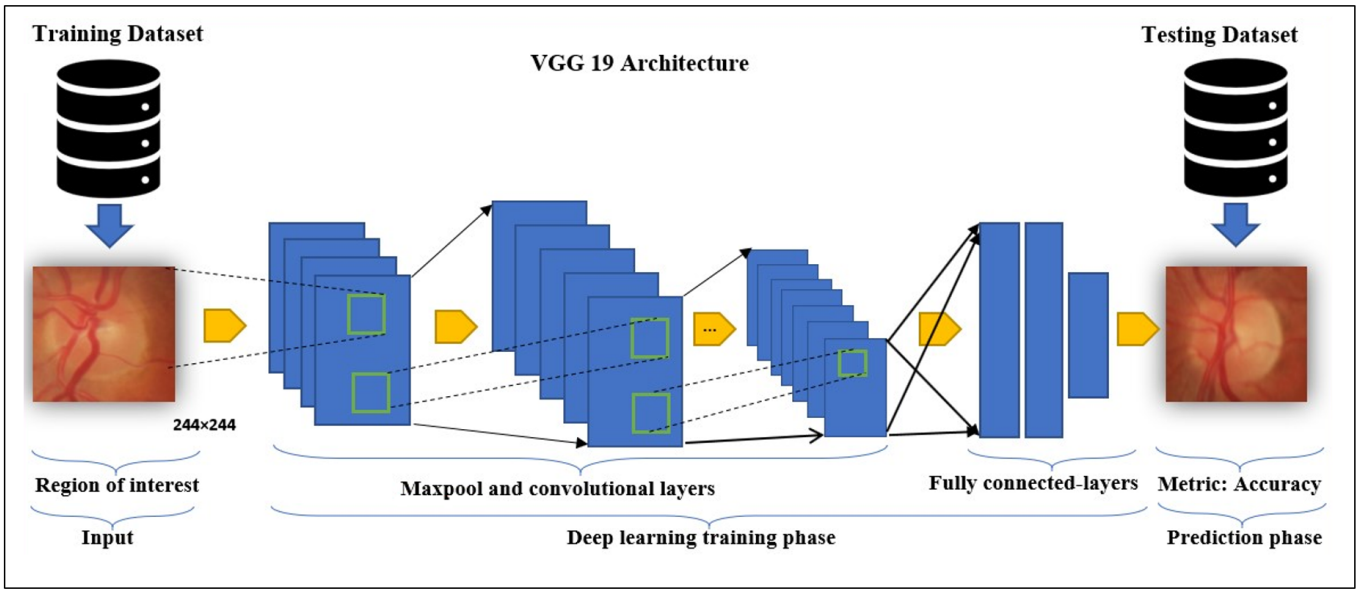


Fig. 4: Schematic block diagram of VGG19 [29].

image (optic disc) was cropped manually and fed to the models.

A. *InceptionResNet-V2*

InceptionResNet-V2 is a very deep convolutional network (825 layers [23]) that has been employed in different image recognition tasks in recent years. InceptionResNet has multiple layers including input, output, convolutional, pooling, residual, concatenate, dropout, and fully connected layers. The default image input size is 299×299 in color format [24].

In our study, we required most of the parameters to be trainable and only a small fraction were selected as default. In order to optimize the training computational complexity, we used a cloud-based *graphics processing unit* (GPU).

Figure 3 shows details of InceptionResNet layers.

B. *VGG19*

VGG19 [27] has been widely used for different applications. As its name implies, VGG19 has 19 layers, with 16 convolutional layers and three fully connected layers [28].

VGG19 accepts a default input size of 244×244 for a color image. In this research, the input size is modified to 299×299 and the number of layers is extended to 25 layers to address the overfitting problem. Similar to InceptionResNet, most of the parameters are trainable and a small fraction is kept as default.

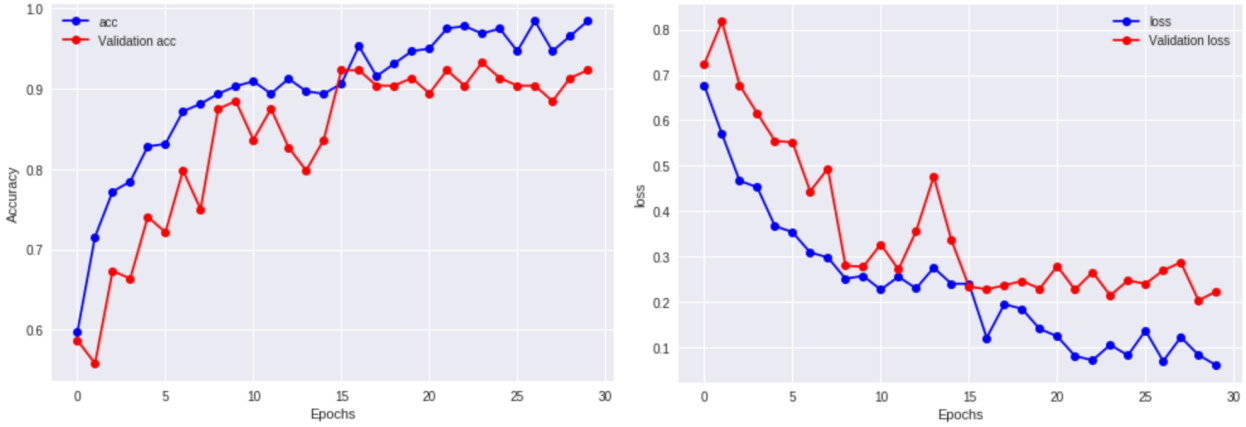
VGG19 has three fully connected layers at the end and all hidden layers use *rectifier units* (ReLU) activation function. VGG19 provides a flexible architecture for different tasks.

Similar to InceptionResNet, we used data augmentation in training. Figure 4 represents the architecture of the VGG19 model.

IV. Results and Evaluation Criteria

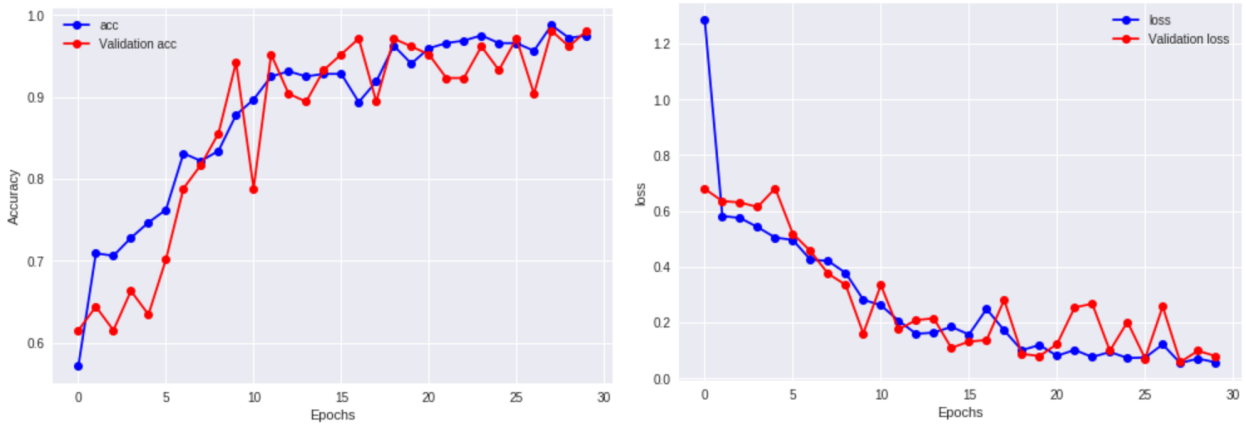
This section presents our results when applying the Inception-ResNet-V2 or the VGG19 model.

Figure 5 presents accuracy versus loss in the training and validation stages of the Inception-ResNet-V2 model. Figure 5a shows how accuracy improves on both training and validation with an increase in the number of epochs. Figure 5b illustrates how loss decreases for both training and validation with an increase in the number of epochs. Accuracy and loss scores are totally converged into each other. The test and retest results show that this model is working accurately.



(a) Accuracy scores (b) Loss scores

Fig. 5: Scores of accuracy and loss function for training and validation stages on Inception-ResNet-V2 model.



(a) Accuracy scores (b) Loss scores

Fig. 6: Scores of accuracy and loss function for training and validation stages on VGG19 model.

Figure 6 demonstrates accuracy versus loss in the training and validation stages of the VGG19 model. Figure 6b illustrates the loss of training and validation with increasing numbers of epochs. Figure 6a describes the accuracy trend on training and validation with increasing numbers of epochs. Similar to the Inception-ResNet-V2 model, the accuracy and loss functions are consistent for training and validation datasets in the VFGG19 model, but the test and retest results show that VGG19 could not overcome overfitting problem. It might be two reasons; lack of data for training, and inadequate depth for glaucoma detection.

All in all, Figs. 5 and 6 indicate that scores of both VGG19 and Inception-ResNet-V2 models converged into each other, but the results on the test and retest datasets show that the VGG19 model has an overfitting problem. Figures 5 and 6 show that the record of epochs is between 0 to 30.

Tables I and II show more detailed results on training, validation, testing, and re-testing for all datasets. We recorded the outcome of the models on epochs five to 50, for every five epochs.

As can be seen, the loss of training decreases consistently with an increase in the number of epochs, except for the 15th epoch in the InceptionResNet-V2 model (Table I). The best result is achieved on the 30th epoch, in that all normal images and 90% of glaucoma cases are identified correctly. Moreover, in this epoch, the system detected 93.3% of the normal eyes and 66.7% of glaucoma eyes correctly.

Table II shows the outcome of the VGG19 model. The accuracy of this model on both test and re-test datasets is not acceptable which indicates that this model suffers from an overfitting problem.

V. Conclusions

We developed a deep learning model for detection of glaucoma from retinal fundus images using InceptionResNetV2 glaucoma and compared it to VGG19, another widely used model.

We used transfer learning to overcome the overfitting problem caused by the limited number of input images. We used two independent datasets for training and re-testing of the model to assure generalizability of the proposed model.

We showed that while VGG19 is unable to provide a generalizable framework, InceptionResNet-V2 provides acceptable

TABLE I: Results of InceptionResNet-V2 model.
- "VAL" indicates validation, "ACC" is accuracy, "N" is normal, and "G" is glaucoma or suspected.
- Bold data belongs in the best epoch for the proposed method.

Results on UCLA dataset							Results on HRF dataset	
Epoch	Train_loss	Train_ACC (%)	VAL_loss	VAL_ACC (%)	Test_N ACC (%)	Test_G ACC (%)	Retest_N ACC (%)	Retest_G ACC (%)
5	0.40	81.88	0.61	71.15	70	100	86.67	60
10	0.26	90.00	0.31	86.54	80	90	86.67	40
15	0.16	93.13	0.26	92.31	80	100	66.67	40
20	0.12	96.88	0.17	92.37	80	90	93.33	20
25	0.11	96.25	0.28	91.35	70	100	73.33	40
30	**0.06**	**98.44**	**0.22**	**92.31**	**100**	**90**	**93.33**	**66.67**
35	0.05	97.81	0.23	89.42	70	100	53.33	53.33
40	0.03	99.06	0.19	92.31	100	90	93.33	40
45	0.02	99.37	0.16	92.31	90	100	66.67	66.67
50	0.05	98.75	0.07	96.15	90	100	86.67	53.33

TABLE II: Results of VGG19 model.
- "VAL" indicates validation, "ACC" is accuracy, "N" is normal, and "G" is glaucoma or suspected.

Results on UCLA dataset							Results on HRF dataset	
Epoch	Train_loss	Train_ACC (%)	VAL_loss	VAL_ACC (%)	Test_N ACC (%)	Test_G ACC (%)	Retest_N ACC (%)	Retest_G ACC (%)
5	0.39	82.19	0.34	84.62	100	0	100	0
10	0.19	91.25	0.19	93.27	100	0	100	0
15	0.20	93.12	0.20	90.38	100	0	100	0
20	0.08	96.88	0.13	94.23	100	0	100	0
25	0.08	96.88	0.14	95.19	100	0	100	0
30	0.03	98.13	0.42	89.42	0	100	40	90
35	0.05	98.44	0.07	96.15	100	0	100	0
40	0.02	99.37	0.12	96.15	100	0	100	0
45	0.06	97.81	0.20	91.35	100	0	100	0
50	0.01	99.69	0.01	99.04	100	0	100	0

accuracy for validation, test, and re-test datasets. The average specificity and sensitivity of InceptionResNet-V2 on test and re-test datasets were over 100%, 90.1%, 90.9% and 93.3% respectively.

The proposed framework could be used clinically and in a research setting for automated glaucoma diagnosis.

REFERENCES

[1] B. Al-Bander, B. M. Williams, W. Al-Nuaimy, M. A. Al-Taee, H. Pratt, and Y. Zheng, "Dense fully convolutional segmentation of the optic disc and cup in colour fundus for glaucoma diagnosis," Symmetry, 10(4):87, 2018.

[2] S. Kingman, "Glaucoma is second leading cause of blindness globally," Bulletin World Health Organization, 82:887–888, 2004.

[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: A systematic review and meta-analysis," Ophthalmology, 121(11):2081–2090, 2014.

[4] Vision problems in the U.S., 2010, see www.visionproblemsus.org

[5] Latest stats at a glance, 2015, see www.blindfoundation.org.nz

[6] Glaucoma New Zealand, see www.glaucoma.org.nz

[7] C. A. Johnson, "Detecting functional changes in the patient's vision: Visual field analysis," in Clinical Glaucoma Care, pp. 117–159, Springer, New York, 2014.

[8] C. A. Johnson, P. A. Sample, G. A. Cioffi, J. R. Liebmann, and R. N. Weinreb, "Structure and function evaluation (SAFE): I. criteria for glaucomatous visual field loss using standard automated perimetry (SAP) and short wavelength automated perimetry (SWAP) 1," American J. Ophthalmology, 134(2):177–185. 2002.

[9] J. Nayak, R. Acharya, P. S. Bhat, N. Shetty, and T. -C. Lim, "Automated diagnosis of glaucoma using digital fundus images," J. Medical Systems, 33(5):337–346, 2009.

[10] J. B. Jonas and A. Dichtl, "Evaluation of the retinal nerve fiber layer," Survey Ophthalmology, 40:369–78, 1996.

[11] N. Thakur and M. Juneja, "Survey on segmentation and classification approaches of optic cup and optic disc for diagnosis of glaucoma," Biomedical Signal Processing Control, 42:162–189, 2018.

[12] C. Bowd, R. N. Weinreb, and L. M. Zangwill, "Evaluating the optic disc and retinal nerve fiber layer in glaucoma. I: Clinical examination and photographic methods," Seminars Ophthalmology, 15:194–205, 2000.

[13] B. Al-Bander, W. Al-Nuaimy, M.A. Al-Taee, and Y. Zheng, "Automated glaucoma diagnosis using deep learning approach," in IEEE Int. Conf. Systems Signals Devices, pp. 207–210, 2017.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Advances Neural Information Processing Systems, pp. 1097–1105, 2012.

[15] H. Nomoto, C. Matsumoto, S. Takada, S. Hashimoto, E. Arimura, S. Okuyama, and Y. Shimomura, "Detectability of glaucomatous changes using SAP, FDT, flicker perimetry, and OCT," J. Glaucoma, 18: 165–71, 2009.

[16] M. C. Westcott, A. I. McNaught, D. P. Crabb, F. W. Fitzke, and R. A. Hitchings, "High spatial resolution automated perimetry in glaucoma," British J. Ophthalmology, 81:452–9, 1997.

[17] G. Lim, Y. Cheng, W. Hsu, and M. L. Lee, "Integrated optic disc and cup segmentation with deep learning," in IEEE Int. Conf. Tools Artificial Intelligence, pp. 162–169, 2015.

[18] 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 Int. Conf. Medical Image Computing Computer-Assisted Intervention, pp. 669–677, 2015.

[19] 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, 441:41–49, 2018

[20] X. Chen, Y. Xu, D.W.K. Wong, T. Y. Wong, and J. Liu, "Glaucoma detection based on deep convolutional neural network," IEEE Int. Conf. Engineering Medicine Biology Society, pp. 715–718, 2015.

[21] H. Wahab, S. R. Haider, S. Khitran, N. ul Huda, and M. U. Akram, "Bright region and vessel density based robust optic disc segmentation," in IEEE Int. Conf. Image Processing Theory Tools Applications, pp. 1–6, 2014.

[22] A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson. "Robust vessel segmentation in fundus images," Int. J. Biomedical Imaging, 20, 2013.

[23] Inception-ResNet-v2 Network, 2017, see www.mathworks.com, access date: 17 September, 2018.

[24] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. "Inception-v4, inception-resnet and the impact of residual connections on learning," in AAAI Conf. Artificial Intelligence, pp. 4278–4284, 2017.

[25] The latest news from Google AI in 2016, see ai.googleblog.com/2016/08/improving-inception-and-image.html, access date: 16 September, 2018.

[26] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and F.-F. Li, "ImageNet: A large-scale hierarchical image database," in IEEE Conf. Computer Vision Pattern Recognition, pp. 248–255, 2009.

[27] K. Simonyan and A. Zisserman. "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409. 1556, 2014.

[28] VGG19 Network, 2017, see www.mathworks.com, access date: 17 September, 2018.

[29] J. Y. Choi, T. K. Yoo, J. G. Seo, J. Kwak, T. T. Um, and T. H. Rim, "Multi-categorical deep learning neural network to classify retinal images: A pilot study employing small database," PloS one, 12(11): p.e0187336, 2017.