



## 저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학박사 학위논문

**Deep Learning Based Glaucoma  
Diagnosis Support System**

딥러닝을 이용한 녹내장 진단 보조  
시스템

2021 년 2 월

서울대학교 대학원

협동과정 바이오엔지니어링 전공

선석규

Deep Learning Based Glaucoma  
Diagnosis Support System

지도교수 김 희 찬

이 논문을 공학박사 학위논문으로 제출함  
2020 년 12 월

서울대학교 대학원  
협동과정 바이오엔지니어링 전공  
선석규

선석규의 공학박사 학위논문을 인준함  
2020 년 12 월

위 원 장 \_\_\_\_\_ 박 기 호 (인)

부위원장 \_\_\_\_\_ 김 희 찬 (인)

위 원 \_\_\_\_\_ 김 남 국 (인)

위 원 \_\_\_\_\_ 공 현 중 (인)

위 원 \_\_\_\_\_ 유 병 욱 (인)

**Ph. D. Dissertation**

**Deep Learning Based Glaucoma  
Diagnosis Support System**

**BY**

**SUKKYU SUN**

**FEBRUARY 2021**

**INTERDISCIPLINARY PROGRAM IN  
BIOENGINEERING  
THE GRADUATE SCHOOL  
SEOUL NATIONAL UNIVERSITY**

# Deep Learning Based Glaucoma Diagnosis Support System

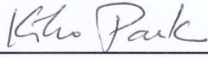
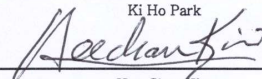
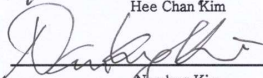
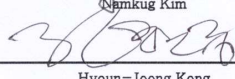
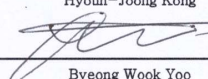
BY  
SUKKYU SUN

INTERDISCIPLINARY PROGRAM IN  
BIOENGINEERING  
THE GRADUATE SCHOOL  
SEOUL NATIONAL UNIVERSITY

THIS DISSERTATION IS APPROVED FOR  
THE DEGREE OF DOCTOR OF PHILOSOPHY

DECEMBER 2020

Approved by Thesis Committee:

Professor	 Ki Ho Park	Chairman
Professor	 Hee Chan Kim	Vice chairman
Professor	 Namkug Kim	Member
Professor	 Hyoun-Joong Kong	Member
Ph. D.	 Byeong Wook Yoo	Member

## **Abstract**

# **Deep Learning Based Glaucoma Diagnosis Support System**

Sukkyu Sun

Interdisciplinary Program in Bioengineering

The Graduate School

Seoul National University

This paper presents deep learning-based methods for improving glaucoma diagnosis support systems. Novel methods were applied to glaucoma clinical cases and the results were evaluated.

In the first study, a deep learning classifier for glaucoma diagnosis based on spectral-domain optical coherence tomography (SD-OCT) images was proposed and evaluated. Spectral-domain optical coherence tomography (SD-OCT) is commonly employed as an imaging modality for the evaluation of glaucomatous structural damage. The classification model was developed using convolutional neural network (CNN) as a base, and was trained with SD-OCT retinal nerve fiber layer (RNFL) and macular ganglion cell-inner plexiform layer (GCIPL) images. The proposed network architecture, termed Dual-Input Convolutional Neural Network (DICNN), showed great potential as an effective classification algorithm based on two input images. DICNN was trained with both RNFL and GCIPL thickness maps that enabled it to discriminate between normal and glaucomatous eyes. The performance of the proposed DICNN was evaluated with accuracy and area under the receiver operating characteristic curve (AUC), and was compared to other methods using these metrics. Compared to other methods, the proposed DICNN model

demonstrated high diagnostic ability for the discrimination of early-stage glaucoma patients in normal subjects. AUC, sensitivity and specificity was 0.869, 0.921, 0.756 respectively.

In the second study, a deep-learning method for increasing the resolution and improving the legibility of Optic-disc Photography(ODP) was proposed. ODP has been proven to be useful for optic nerve evaluation in glaucoma. But in clinical practice, limited patient cooperation, small pupil or media opacities can limit the performance of ODP. A model to enhance the resolution of ODP images, termed super-resolution, was developed using Super Resolution Generative Adversarial Network(SR-GAN). To train this model, high-resolution original ODP images were transformed into two counterparts: (1) down-scaled ‘low-resolution ODPs’, and (2) ‘compensated high-resolution ODPs’ with enhanced visibility of the optic disc margin and surrounding retinal vessels which were produced using a customized image post-processing algorithm. The SR-GAN was trained to learn and recognize the differences between these two counterparts. The performance of the network was evaluated using Peak Signal to Noise Ratio (PSNR), Structural Similarity (SSIM), and Mean Opinion Score (MOS). The proposed study demonstrated that deep learning can be applied to create a generative model that is capable of producing enhanced ophthalmic images with 4x resolution and with improved structural details. The proposed method can be used to enhance ODPs and thereby significantly increase the detection accuracy of optic disc pathology. The average PSNR, SSIM and MOS was 25.01, 0.75, 4.33 respectively

In the third study, a deep-learning model was used to classify suspected glaucoma and to predict subsequent glaucoma onset-year in glaucoma suspects using clinical data and retinal images (ODP & Red-free Fundus RNFL Photo). Clinical data contains useful information about glaucoma diagnosis and prediction. However, no study has been undertaken to investigate how combining different types of clinical information would be helpful for predicting the subsequent course of glaucoma in an individual patient. For this

study, image features extracted using Convolutional Auto Encoder (CAE) along with clinical features were used for glaucoma suspect classification and onset-year prediction. The performance of the proposed model was evaluated using accuracy and Mean Squared Error (MSE). Combining the CAE extracted image features and clinical features improved glaucoma suspect classification and on-set year prediction performance as compared to using the image features and patient features separately. The average MSE between onset-year and predicted onset year was 2.613

In this study, deep learning methodology was applied to clinical images related to glaucoma. DICNN with RNFL and GCIPL images were used for classification of glaucoma, SR-GAN with ODP images were used to increase detection accuracy of optic disc pathology, and CAE & machine learning algorithm with clinical data and retinal images was used for glaucoma suspect classification and onset-year prediction. The improved glaucoma diagnosis performance was validated using both technical and clinical parameters. The proposed methods as a whole can significantly improve outcomes of glaucoma patients by early detection, prediction and enhancing detection accuracy.

**Keywords:** Deep learning, Convolutional neural network, Super resolution, Glaucoma diagnosis, Glaucoma onset-year prediction

**Student Number: 2016-30277**



# Contents

Abstract .....	i
Contents.....	iv
List of Tables .....	vii
List of Figures .....	viii

## **Chapter 1 General Introduction..... 1**

1.1 Glaucoma .....	1
1.2 Deep Learning for Glaucoma Diagnosis.....	3
1.4 Thesis Objectives .....	3

## **Chapter 2 ..... 6**

### **Dual-Input Convolutional Neural Network for Glaucoma Diagnosis using Spectral-Domain Optical Coherence Tomography**

2.1 Introduction.....	6
2.1.1 Background.....	6
2.1.2 Related Work.....	7
2.2 Methods.....	8
2.2.1 Study Design.....	8
2.2.2 Dataset .....	9
2.2.3 Dual-Input Convolutional Neural Network (DICNN) .....	15
2.2.4 Training Environment .....	18

2.2.5	Statistical Analysis .....	19
2.3	Results .....	20
2.3.1	DICNN Performance .....	20
2.3.1	Grad-CAM for DICNN.....	34
2.4	Discussion .....	37
2.4.1	Research Significance.....	37
2.4.2	Limitations .....	40
2.5	Conclusion .....	42

## **Chapter 3 ..... 43**

### **Deep-learning-based enhanced optic-disc photography**

3.1	Introduction.....	43
3.1.1	Background .....	43
3.1.2	Needs .....	44
3.1.3	Related Work.....	45
3.2	Methods.....	46
3.2.1	Study Design.....	46
3.2.2	Dataset .....	46
	3.2.2.1 Details on Customized Image Post-Processing	
Algorithm	47	
3.2.3	SR-GAN Network.....	50
	3.2.3.1 Design of Generative Adversarial Network .....	50
	3.2.3.2 Loss Functions .....	55
3.2.4	Assessment of Clinical Implications of Enhanced ODPs ...	58
3.2.5	Statistical Analysis .....	60
3.2.6	Hardware Specifications & Software Specifications .....	60

3.3	Results .....	62
3.3.1	Training Loss of Modified SR-GAN .....	62
3.3.2	Performance of Final Network.....	66
3.3.3	Clinical Validation of Enhanced ODP by MOS comparison .....	77
3.3.4	Comparison of DH-Detection Accuracy .....	79
3.4	Discussion .....	80
3.4.1	Research Significance.....	80
3.4.2	Limitations .....	85
3.5	Conclusion .....	88

## **Chapter 4 ..... 89**

### **Deep Learning Based Prediction of Glaucoma Onset Using Retinal Image and Patient Data**

4.1	Introduction.....	89
4.1.1	Background .....	89
4.1.2	Related Work.....	90
4.2	Methods.....	90
4.2.1	Study Design.....	90
4.2.2	Dataset .....	91
4.2.3	Design of Overall System .....	94
4.2.4	Design of Convolutional Auto Encoder .....	95
4.2.5	Glaucoma Suspect Classification.....	97
4.2.6	Glaucoma Onset-Year Prediction.....	97
4.3	Result .....	99
4.3.1	Performance of Designed CAE.....	99
4.3.2	Performance of Designed Glaucoma Suspect Classification .....	101

4.3.3	Performance of Designed Glaucoma Onset-Year Prediction	105
4.4	Discussion .....	110
4.4.1	Research Significance .....	110
4.4.2	Limitations .....	110
4.5	Conclusion .....	111
<b>Chapter 5 Summary and Future Works .....</b>		<b>112</b>
5.1	Thesis Summary .....	112
5.2	Limitations and Future Works .....	113
Bibliography .....		115
Abstract in Korean .....		127
Acknowledgement .....		130

# List of Tables

<b>Table 2.1</b>	Comparison of Demographic and Clinical Characteristics Between Primary Open-angle Glaucoma and Normal Subjects.....	<b>21</b>
<b>Table 2.2</b>	The test accuracy and the area under receiver operating characteristic curve of machine learning algorithms distinguishing between normal and glaucoma subjects .....	<b>22</b>
<b>Table 2.3</b>	The final test accuracy of comparison model trained with 3classes .....	<b>23</b>
<b>Table 2.4</b>	The test accuracy and the area under receiver operating characteristic curve of machine learning algorithms distinguishing between normal, early and glaucoma subjects (one vs rest).....	<b>24</b>
<b>Table 3.1</b>	Learning rate vs PSNR and SSIM table. ....	<b>65</b>
<b>Table 3.2</b>	Comparison of the SSIM index values for the representative test images. ....	<b>69</b>
<b>Table 3.3</b>	Comparison of the PSNR values for the representative test image sets. ....	<b>70</b>
<b>Table 3.4</b>	Comparison of the MOS values for the representative test image sets. ....	<b>73</b>
<b>Table 3.5</b>	Demographic and Clinical Characteristics of Study Subjects (N = 98) .....	<b>77</b>
<b>Table 4.1</b>	Statistical characteristic of 210 total patient.....	<b>91</b>
<b>Table 4.2</b>	Glaucoma suspect classification result .....	<b>101</b>
<b>Table 4.3</b>	Feature importance of glaucoma suspect classification networks .....	<b>102</b>
<b>Table 4.4</b>	Glaucoma onset-year regression result.....	<b>105</b>
<b>Table 4.5</b>	Feature importance of glaucoma suspect classification networks .....	<b>105</b>
<b>Table 4.6</b>	Linear correlation glaucoma onset-year vs patient features	<b>106</b>

# List of Figures

<b>Figure 1.1</b>	Present method of glaucoma diagnosis .....	2
<b>Figure 2.1</b>	SD-OCT images .....	6
<b>Figure 2.2</b>	Representative Images RNFL and GCIPL .....	11
<b>Figure 2.3</b>	Gaussian noised added RNFL & GCIPL .....	14
<b>Figure 2.4</b>	Dual-Input Convolutional Neural Network (DICNN)'s architecture. ....	17
<b>Figure 2.5</b>	Overall SD-OCT Data Description.....	20
<b>Figure 2.6</b>	Test accuracy obtained by different optimizers .....	25
<b>Figure 2.7</b>	Receiver operating characteristic (ROC) curve of DICNN ...	27
<b>Figure 2.8</b>	Receiver operating characteristic (ROC) curve of DICNN(one vs rest) .....	28
<b>Figure 2.9</b>	Receiver operating characteristic (ROC) curve of DICNN(macro averaging).....	29
<b>Figure 2.10</b>	Test accuracy for DICNN (black square) and CNN with GCIPL (red dot) or RNFL (blue triangle), respectively .....	30
<b>Figure 2.11</b>	Representative cases of DICNN's misclassification between early-stage POAG patients (A-E) and normal subjects (F-J) .....	31
<b>Figure 2.12</b>	Calculating Grad-CAM for DICNN .....	34
<b>Figure 2.13</b>	Grad-CAM for test images .....	35
<b>Figure 2.14</b>	Test accuracy obtained by VGG16 and VGG19 as baseline model .....	40
<b>Figure 3.1</b>	Representative images of Optic Disc Hemorrhage(DH), Peripapillary chorioretinal atrophy(PPA), vessel alterations.....	43
<b>Figure 3.2</b>	Low Quality ODPs .....	44
<b>Figure 3.3</b>	Customized image post-processing algorithm for maximized visibility of hemorrhage .....	47
<b>Figure 3.4</b>	Principle of enhanced image formation via super-resolution generative adversarial network (SR-GAN) .....	51

<b>Figure 3.5</b>	Architecture of generator and discriminator network.....	<b>53</b>
<b>Figure 3.6</b>	Validation results for representative test image sets.....	<b>58</b>
<b>Figure 3.7</b>	Training curve for SR-GAN algorithm with learning rate ....	<b>63</b>
<b>Figure 3.8</b>	Training PSNR curve and validation PSNR and SSIM curve for SR-GAN algorithm with learning rate.....	<b>64</b>
<b>Figure 3.9</b>	Image mapped PSNR of generated by SR-GAN.....	<b>71</b>
<b>Figure 3.10</b>	Representative optic-disc photography (ODP) of eye with optic disc hemorrhage (DH).....	<b>72</b>
<b>Figure 3.11</b>	Validation Results for Control Group .....	<b>76</b>
<b>Figure 3.12</b>	Scatter plot of delta mean opinion score ( $\Delta$ MOS) against MOS of original optic-disc photography (ODP).....	<b>78</b>
<b>Figure 3.13</b>	Example of fundus photography transferred from another institution as printed document .....	<b>82</b>
<b>Figure 4.1</b>	t-SNE for Dataset Embedding .....	<b>92</b>
<b>Figure 4.2</b>	Overall systems of designed classification and regression network .....	<b>94</b>
<b>Figure 4.3</b>	Architecture of CAE network.....	<b>96</b>
<b>Figure 4.4</b>	Reconstructed image while training CAE .....	<b>99</b>
<b>Figure 4.5</b>	Training loss curve for CAE.....	<b>99</b>
<b>Figure 4.6</b>	ROC curve for glaucoma suspect classification networks. .	<b>103</b>
<b>Figure 4.7</b>	Regression result for glaucoma onset-year prediction.....	<b>107</b>
<b>Figure 4.8</b>	95% Confidence interval for regression result (each individual patient) .....	<b>108</b>
<b>Figure 4.9</b>	Onset-year prediction for normal patient.....	<b>109</b>

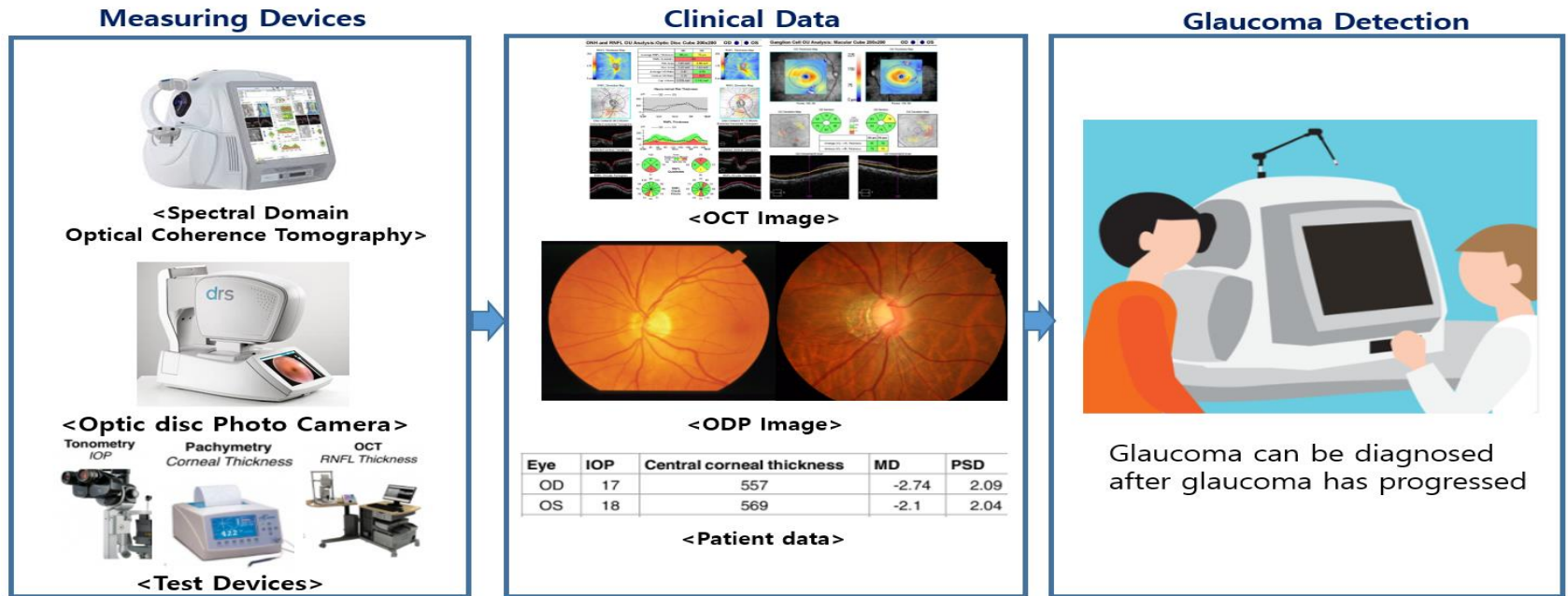
# Chapter 1 General Introduction

## 1.1 Glaucoma

Glaucoma, a neurodegenerative disease brought on by progressive retinal ganglion cell (RGC) loss, is associated with structural changes of the optic nerve head (ONH) and retinal nerve fiber layer (RNFL).<sup>[1, 2]</sup> As glaucoma progression runs its course, structural changes often are detected before functional losses become clearly apparent.[3] Therefore, early detection of glaucomatous structural damage is essential for proper management and maintenance of patients' vision-related quality of life.[4] The anterior chamber angle formed by the cornea and iris remains open, but the outflow resistance through the trabecular meshwork is increased. This causes pressure in the eye to gradually increase. The pressure induced injury is one of the main mechanisms of the optic nerve damage in glaucoma.

Glaucoma is a leading cause of irreversible blindness worldwide; however, the pathophysiology of glaucoma is poorly understood and the exact factors leading to its progression are unclear. An estimated 79.6 million persons are expected to have glaucoma worldwide by 2020 and 111 million by 2040[5]





**Figure 1.1 Present method of glaucoma diagnosis**

Current measuring device for glaucoma diagnosis(left), clinical data obtained by measuring devices(middle), process of glaucoma detection(right)

## **1.2 Deep Learning for Glaucoma Diagnosis**

The application of CNN to various glaucoma related clinical data, such as spectral-domain optical coherence tomography (SD-OCT) image optic disc photograph (ODP), patient data have been reported. For SD-OCT CNN based classification methods have been done using RNFL, ganglion cell-inner plexiform layer (GCIPL), super resolution applied to ODPs, machine learning based glaucoma classification on patient data have been reported.

## **1.3 Thesis Objectives**

The objective of this study is to support glaucoma diagnosis using deep learning on clinical data and to validate performance. Performances of designed network was validated with clinical and engineering parameters.

In chapter 2 was to evaluate with spectral-domain optical coherence tomography (SD-OCT). The previous studies have reported on SD-OCT using convolutional neural networks (CNN), using only scanned image or ensemble logit values when RNFL and GCIPL are feeding to CNN respectively, or depth wise stacked RNFL GCIPL thickness and deviation map and using as inputs of CNN[6]. However, did not yield diagnostic performance that was superior to

that of RNFL probability alone. In our study, the deep-learning algorithms analyzed by DICNN with combined RNFL and GCIPL features on thickness maps showed better diagnostic performance for early-stage glaucoma relative to the conventional deep-learning methods. Model was evaluated with accuracy, sensitivity, specificity and ROC.

In chapter 3, was to propose a deep-learning approach for increased resolution and improved legibility of Optic-disc Photography(ODP) by contrast, color, and brightness compensation. The previous studies have reported on Optic-disc super resolution using convolutional neural networks (CNN), solely focused only on resolution improvement. (up to x32[7]). Even when high-resolution ODPs can be obtained, red-colored blood vessels and red-orange-colored retina sometimes cause indistinct pathologies such as small-sized DH to be missed. Heuristic algorithms cannot enhance image universally. (contrast, color, and brightness). Model was evaluated with PSNR, SSIM, MOS.

In chapter 4, was to develop a deep-learning model using clinical data and retinal images (ODP & Red-free Fundus RNFL Photo) for classification glaucoma suspect and prediction of subsequent glaucoma onset-year in glaucoma suspects. The previous studies[8] have reported using glaucoma patient data is for diagnosis of glaucoma. No study has been undertaken to investigate whether and how combining different types of clinical information would be helpful for predicting the subsequent course of glaucoma in an

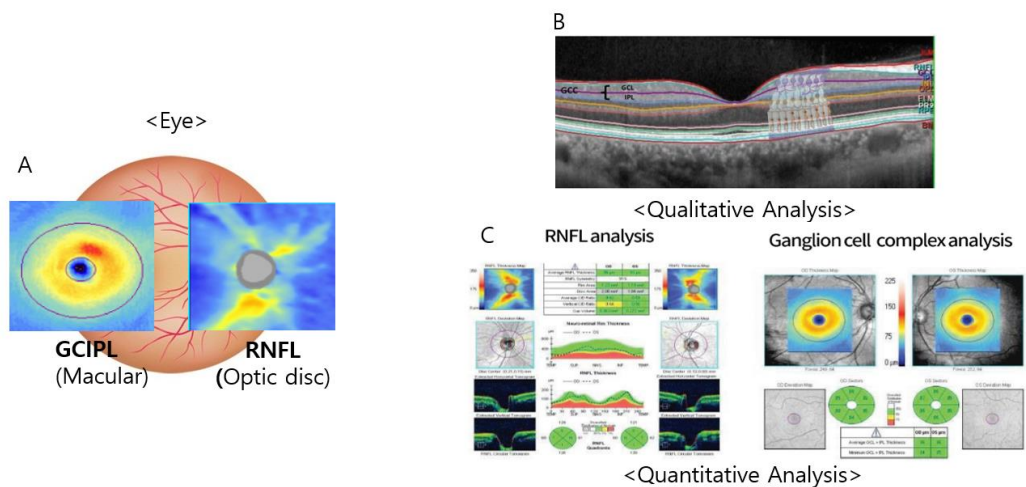
individual patient. Classification model was evaluated with accuracy, sensitivity, specificity and ROC. Regression model was evaluated with MSE.

# Chapter 2

## Dual-Input Convolutional Neural Network for Glaucoma Diagnosis using Spectral-Domain Optical Coherence Tomography

### 2.1 Introduction

#### 2.1.1 Background



**Figure 2.1 SD-OCT images**

Mapping SD-OCT color image with eye(A), scanned image of SD-OCT(B),  
raw data of SD-OCT(C)

Spectral-domain optical coherence tomography (SD-OCT) is commonly employed as an imaging modality for evaluation of glaucomatous structural damage.[9] Its clinical utility for glaucoma is primarily in the evaluation of RNFL parameters for provision of comprehensive assessments of RGC axons approaching the ONH.[10, 11] Furthermore, SD-OCT's enhanced performance allows for assessment of macular parameters in glaucoma evaluation, which utility is especially valuable, as the macula contains the highest RGC concentration in the retina;[12] thus too, RGC loss typically is more readily detectable in the macula.[13]

The most effective and widely used deep-learning algorithm for images and videos is convolutional neural network (CNN). CNN allows for both extraction of features from image data in the convolutional layers and pooling layers and classification based on the fully connected layers' features.[14] If multi-domain feature extraction can be combined with deep learning, sufficient information can be provided for the classification task, and better results are likely to be achieved.

### **2.1.2 Related Work**

Recently, the Dual-Input Convolutional Neural Network (DICNN) has shown great utility as well as potential as a deep-learning algorithm for effective classification based on two images used as inputs. Li et al.[15] showed

that a dual-input network structure using both electrocardiogram and phonocardiogram signals identified more underlying features inside the signals, thereby improving the performance for diagnosing coronary artery disease. Likewise, Choi et al.[16] developed a dual-input deep-learning model that interprets 2 different radiographs simultaneously for detection of supracondylar fracture, showing a diagnostic accuracy comparable to radiologists'. This approach is promising, as it mimics how a human clinician interprets images in making a diagnosis.

## **2.2 Methods**

### **2.2.1 Study Design**

In this paper, we propose this deep-learning algorithm as an effective and efficient means of diagnosing early glaucoma accurately based on two images. DICNN was trained with both RNFL and macular ganglion cell-inner plexiform layer (GCIPL) thickness maps that enabled it to discriminate between normal and glaucomatous eyes. DICNN was evaluated along with other methods for accuracy and area under the receiver operating characteristic curve (AUC).

### **2.2.2 Dataset**

Eyes were chosen from a database of subjects examined for glaucoma at the Glaucoma Clinic, Seoul National University Hospital, between January 2012 and January 2019. This study was approved by the Seoul National University Hospital Institutional Review Board and adhered to the tenets of the Declaration of Helsinki. Informed consent was waived due to the study's retrospective nature.

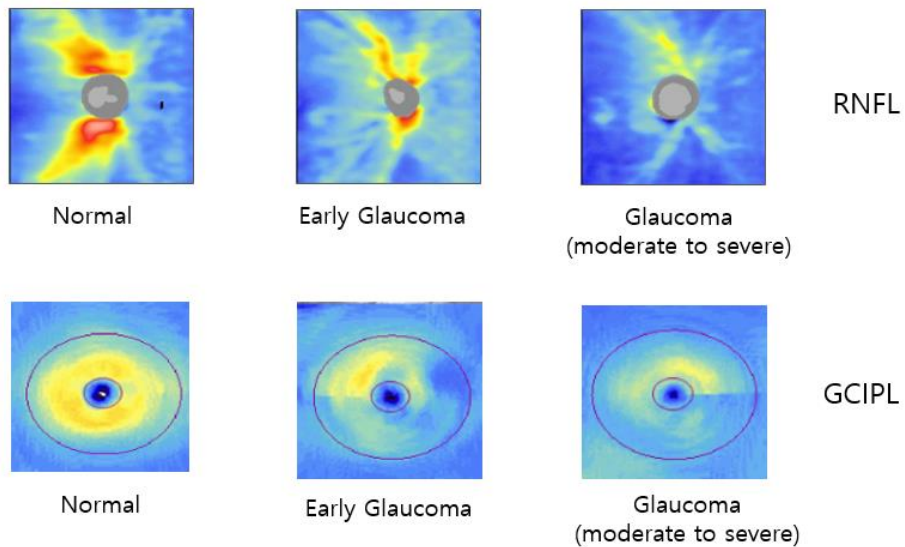
All of the study subjects underwent a complete ophthalmic examination that included best-corrected visual acuity, refraction, slit-lamp biomicroscopy, gonioscopy, Goldmann applanation tonometry (Haag-Streit, Koniz, Switzerland), dilated stereoscopic examination of optic disc, digital color stereo disc photography (SDP), red-free RNFL photography, Cirrus HD-OCT, and central 24-2 threshold testing of the Humphrey Visual Field (HVF) (HFA II; Humphrey Instruments Inc., Dublin, CA, USA). The patients additionally underwent, at the baseline examination, central corneal thickness measurement (CCT; Orbscan™ 73 II, Bausch & Lomb Surgical, Rochester, NY, USA) and axial-length measurement (IOL Master™ ver. 5, Carl-Zeiss Meditec, Dublin, CA, USA).

For inclusion in the study, glaucoma patients had to satisfy the definition of primary open-angle glaucoma (POAG), irrespective of untreated intraocular pressure (IOP) level: the presence of certain characteristic changes (i.e., localized or diffuse neuroretinal rim thinning/notching) on SDP; RNFL



defects in corresponding regions on red-free RNFL photographs with open anterior chamber angles. The POAG diagnosis was made regardless of glaucomatous visual field (VF) defect presence or absence.

Final inclusion or exclusion of patients was determined by an experienced ophthalmologist (YKK). Since our outcomes were measured at the ocular (eye) level, we tried to minimize problems that could arise from violating the assumption made by the majority of statistical tests that each data value is independent. In cases where both eyes proved eligible, therefore, one eye was selected randomly. Subsequently, the POAG patients were divided into 2 groups, based on the standard automated perimetry (HVF Analyzer™ 24-2 SITA-Standard strategy) data: early glaucoma (VF mean deviation [MD] > -6 dB) and moderate-to-severe glaucoma (VF MD < -6 dB).



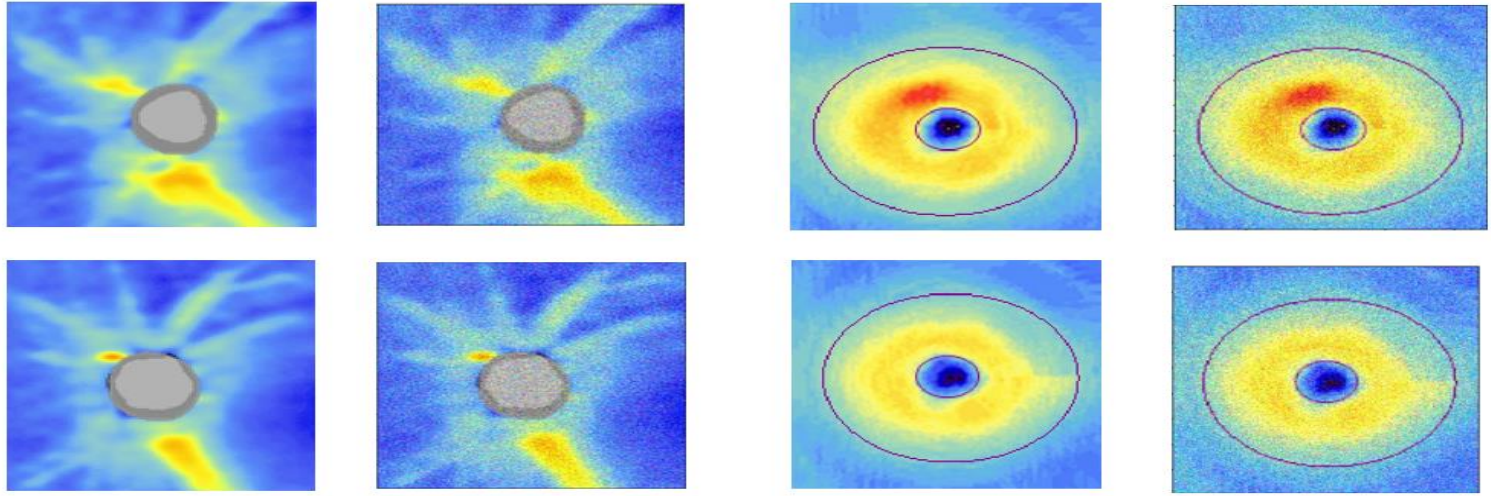
**Figure 2.2 Representative Images RNFL and GCIPL**

Normal, Early Glaucoma, Moderate to Severe Glaucoma, as glaucoma progresses thickness of optic nerve decreases (blue color)

The normal eyes in this study were from subjects without any ocular abnormalities in either eye, since patients with unilateral glaucoma constitute a group at higher risk for future development of glaucoma in the other, normal eye.[17] They met the following criteria: (1) baseline IOP less than or equal to 21 mmHg, with no IOP-elevation history; (2) no glaucomatous optic disc changes as detected on SDP; (3) no RNFL defects on red-free fundus images; 4) normal HVF results.

All scans were acquired according to the Optic Disc Cube 200×200 protocol of Cirrus HD-OCT™, which was designed for cube-scan positioning on the ONH, and which is the primarily utilized modality for glaucoma analysis. Once the subject was seated and properly aligned, the iris was brought into view by means of a mouse-driven alignment system, and the line-scanning ophthalmoscopic image was brought into focus with adjustment for refractive error. Then, the ONH was centered in the live image, and the centering (Z-offset) and enhancement (polarization) were optimized. The laser subsequently scanned a 6 mm×6 mm area, thus capturing a data cube consisting of 200 A-scans from 200 linear B-scans (40,000 points), all in approximately 1.5 seconds (at 27,000 A-scans/sec). The ONH parameters were measured automatically using a Carl Zeiss Meditec analysis algorithm that had been developed for Cirrus HD-OCT (version 6.0). For the macular cube, a 512×128 grid that consists of 128 horizontal B-scans each comprising 512 A-scans was used. Only high-quality scans (signal strength  $\geq 7$ , with absence of discontinuity or

misalignment, involuntary saccade, blinking artifacts, segmentation failure or artifacts) were used in the final analysis.



**Figure 2.3 Gaussian noised added RNFL & GCIPL**

For augmenting original data, Gaussian noise was added (Mean -0, Std-(10~30)), added noise changes from epoch to epoch.

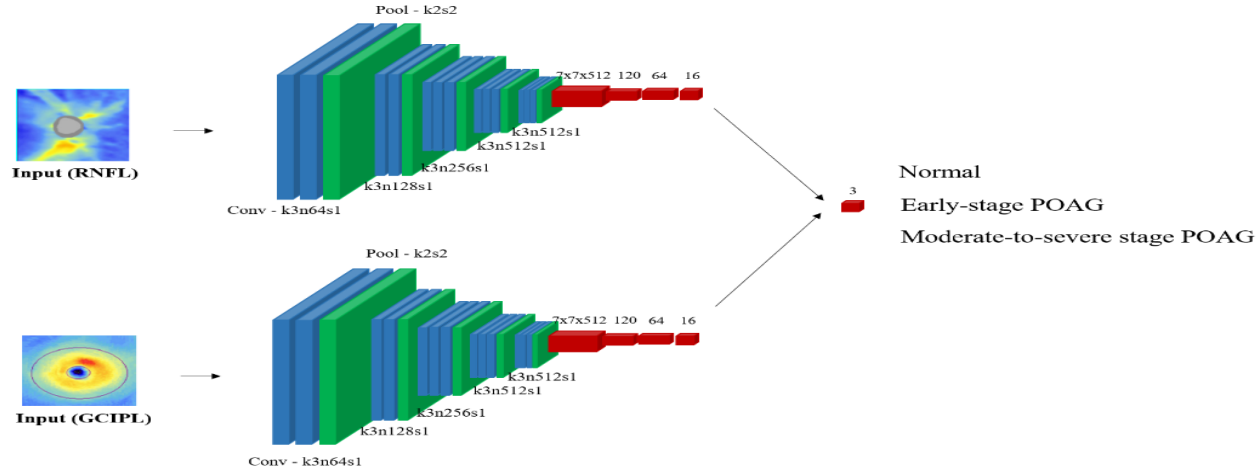
Right eye images are mirrored to left eye image. Images were augmented with Gaussian noise added average 0 standard deviation of 10~30, no geometric transformation techniques such as flip, rotation, or inversion applied.

### **2.2.3 Dual-Input Convolutional Neural Network (DICNN)**

Deep-learning-based features were extracted by training CNN on either peripapillary RNFL or macular GCIPL thickness map images. The images were extracted from the average RNFL thickness and RNFL quadrants, and then were used for analysis of glaucoma on the RNFL thickness map.[18-20] For medical images with only a small number of datasets, CNN analysis proceeds in two ways (learns RNFL and GCIPL separately), given both the computational expensiveness of the deep-learning algorithm and the likeliness of overfitting of extracted features. However, DICNN can prevent overfitting while utilizing the features of both RNFL and GCIPL simultaneously.

The RNFL thickness map provides a color-coded display of RNFL thickness within the 0-370  $\mu\text{m}$  range. In the present study, the value of each thickness map pixel was normalized to the RNFL thickness, and that value was converted to a value within the 0-255 range for better image representation. Images were resized to 224x224x3 for learning by DICNN. Since a large number of images must be available for better training of CNN, augmentation,

which entailed enlargement of image data by addition of Gaussian noise, was performed. Learning was set at 0.01 (gradient descent optimizer[21]). Since the last-fully-connected layer must have three classes (i.e., normal, early POAG, moderate-to-severe POAG), other fully connected layers had 120, 64, 16 and each 16 features from last-fully-connected layer (RNFL and GCIPL) was reduced to three nodes (three classes).



**Figure 2.4 Dual-Input Convolutional Neural Network (DICNN)’s architecture.** The kernel size (k), number of feature maps (n), and stride (s) correspondingly are indicated for each of the convolutional layers. The extracted features from the convolutional layers in each image (i.e., RNFL and GCIPL thickness maps) were flattened and fed to successive fully-connected layers consisting of 120, 64, 16 nodes. The features from each convolutional layer were fused at the final layers. Three-layer Multi-Layer Perception (MLP) was adopted as the final classification task for reduction to 3 nodes. RNFL, retinal nerve fiber layer; GCIPL, ganglion cell-inner plexiform layer; POAG, primary open-angle glaucoma



DICNN proceeded by extracting features selected for deep learning (Figure 2.4). The fully connected layer combines all of the features that were extracted from the convolutional and pooling layers and analyzes the relationships among features in the neural network architecture. DICNN was trained by backpropagating the output logit value simultaneously with the training of the weights of the VGG16s. The extracted features from the convolutional layers in each image (i.e., RNFL and GCIPL thickness maps) were flattened and fed to successive fully-connected layers consisting of 120, 64, 16 nodes.[22] The features from each convolutional layer were fused at the final layers. Three-layer Multi-Layer Perception (MLP) was adopted as the final classification task for reduction to 3 nodes.

## **2.2.4 Training Environment**

### **Hardware specifications**

CPU: Intel core i7-7700 3.60Hz

GPU: TITAN X (Pascal) 12GB

RAM: 16GB

### **Software specifications**

Deep learning libraries

Tensorflow – 1.14.0 with cuda 10.0 and cudnn 7.6.3

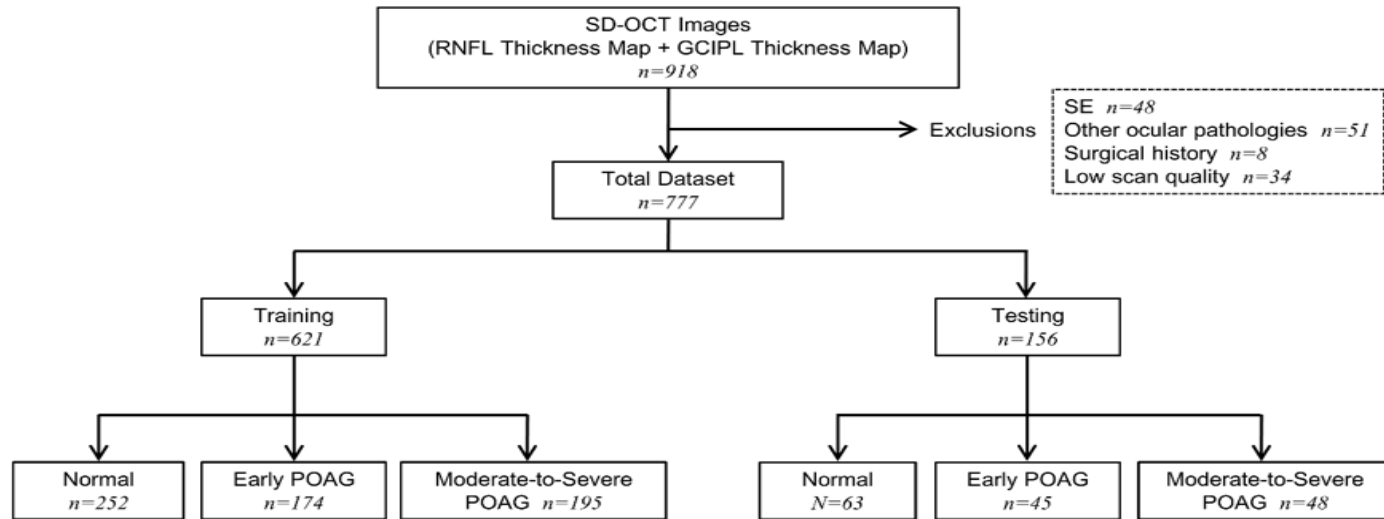
Python – version 3.6

### **2.2.5 Statistical Analysis**

The demographic data were compared among the 3 groups with the Bonferroni post-hoc test by one-way analysis of variance. The categorical data were analyzed by  $\chi^2$  test with Bonferroni correction. We used the AUC with 95% confidence intervals (CIs) in the evaluations of the algorithms' performances. 95% CIs were calculated by applying, out of each first and second fully connected layer of the VGG19, 20% of drop.[23] The sensitivity and specificity in the receiver operating characteristic (ROC) curves were obtained by thresholding the logit value coming out of the network output.[24] For comparison between the 2 groups, the logit values were normalized to make the sum of their respective logit values 1. All of the statistical analyses were performed with Matlab (The MathWorks Inc, MA, USA). pROC package of R software version 2.2.1 (R Project for Statistical Computing, Vienna, Austria) was used as a DeLong test for comparison of ROC curves.

## 2.3 Results

### 2.3.1 DICNN performance



**Figure 2.5 Overall SD-OCT Data Description**

Total 918 SD-OCT was collected, 171 patients was excluded with other ocular pathologies, surgical history and low scan quality.

**Table 2.1 Comparison of Demographic and Clinical Characteristics Between Primary Open-angle Glaucoma and Normal Subjects**

Variable	Normal ( <i>n</i> = 315)	Early POAG ( <i>n</i> = 219)	Moderate-to-severe POAG ( <i>n</i> = 243)	<i>P</i>	Post Hoc Test
Age (yrs)	57.5 ± 9.93	59.2 ± 8.56	59.4 ± 8.61	0.033*	0.112 <sup>†</sup> (N = E), 0.061 <sup>†</sup> (N = M), 1.000 <sup>‡</sup> (E = M)
Male, n (%)	158 (50.2)	121 (55.3)	130 (53.5)	0.768 <sup>‡</sup>	
Spherical equivalent (D)	-0.62 ± 2.00	-0.60 ± 1.77	-0.70 ± 1.88	0.808*	
IOP (mmHg)	12.7 ± 2.37	13.0 ± 2.13	13.1 ± 2.20	0.094*	
CCT (μm)	546.6 ± 35.9	542.7 ± 36.5	541.0 ± 39.3	0.185*	
Axial length (mm)	23.5 ± 0.52	23.4 ± 0.50	23.5 ± 0.48	0.489*	
VF MD (decibels)	-0.16 ± 1.17	-2.83 ± 1.71	-9.54 ± 2.50	< 0.001*	< 0.001 <sup>†</sup> (N > E), < 0.001 <sup>†</sup> (N > M), < 0.001 <sup>†</sup> (E > M)
VF PSD (decibels)	0.97 ± 1.03	3.91 ± 1.93	8.10 ± 2.46	< 0.001*	< 0.001 <sup>†</sup> (N > E), < 0.001 <sup>†</sup> (N > M), < 0.001 <sup>†</sup> (E > M)

Values are mean ± standard deviation. \*One-way analysis of variance, <sup>†</sup>Bonferroni post hoc test, <sup>‡</sup> chi-square test with Bonferroni correction.

POAG, primary open-angle glaucoma; D, diopters; IOP, intraocular pressure; CCT, central corneal thickness; VF, visual field; MD, mean deviation; PSD, pattern standard deviation. N, normal group; E, early glaucoma group; M, moderate-to-severe glaucoma group

**Table 2.2 The test accuracy and the area under receiver operating characteristic curve of machine learning algorithms distinguishing between normal and glaucoma subjects**

	<b>Normal vs. POAG</b>		<b>Normal vs. Early POAG</b>	
	Accuracy (%)	AUC	Accuracy (%)	AUC
DICNN (RNFL/GCIPL)	92.79	0.957	85.19	0.869
VGG16 (RNFL)	94.59	0.978	80.00	0.850
VGG16 (GCIPL)	88.29	0.921	72.90	0.767

POAG, primary open-angle glaucoma; AUC, area under the receiver operating characteristic curve; DICNN, dual-input convolutional neural network; RNFL, retinal nerve fiber layer; GCIPL, ganglion cell-inner plexiform layer; VGG, visual geometry group.

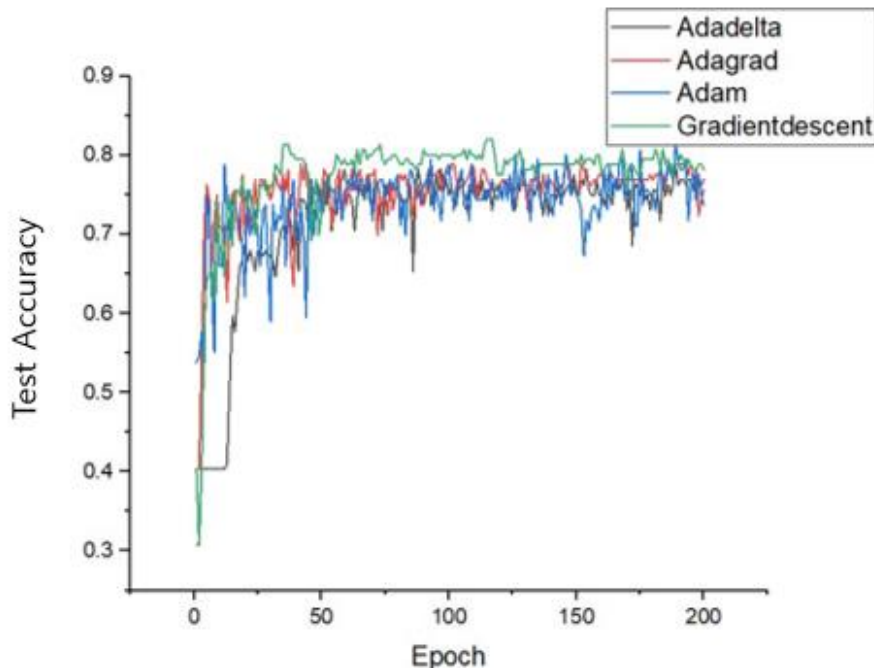
**Table 2.3 The final test accuracy of comparison model trained with 3classs**

	<b>InceptionV3[25] (RNFL)</b>	<b>ResNet50[26] (RNFL)</b>	<b>Inception ResNet- V2[27] (RNFL)</b>	<b>DenseNet121[28] (RNFL)</b>	<b>Xception[29] (RNFL)</b>	<b>DICNN (both)</b>	<b>VGG16 (GCIPL)</b>	<b>VGG16 (RNFL)</b>
Maximum Test Accuracy(%)	0.75	0.7756	0.7756	0.7756	0.7564	0.821	0.718	0.75

**Table 2.4 The test accuracy and the area under receiver operating characteristic curve of machine learning algorithms distinguishing between normal, early and glaucoma subjects (one vs rest)**

	Early				Normal			Severe		
	Test Accuracy(3class)	AUC	Specificity	Sensitivity	AUC	Specificity	Sensitivity	AUC	Specificity	Sensitivity
<b>DICNN (RNFL/GCIPL)</b>	0.821	0.85	0.898	0.689	0.92	0.833	0.92	0.94	0.967	0.813
<b>VGG16 (RNFL)</b>	0.750	0.79	0.861	0.533	0.92	0.769	0.905	0.95	0.931	0.750
<b>VGG16 (GCIPL)</b>	0.705	0.75	0.825	0.556	0.84	0.695	0.841	0.89	0.963	0.667

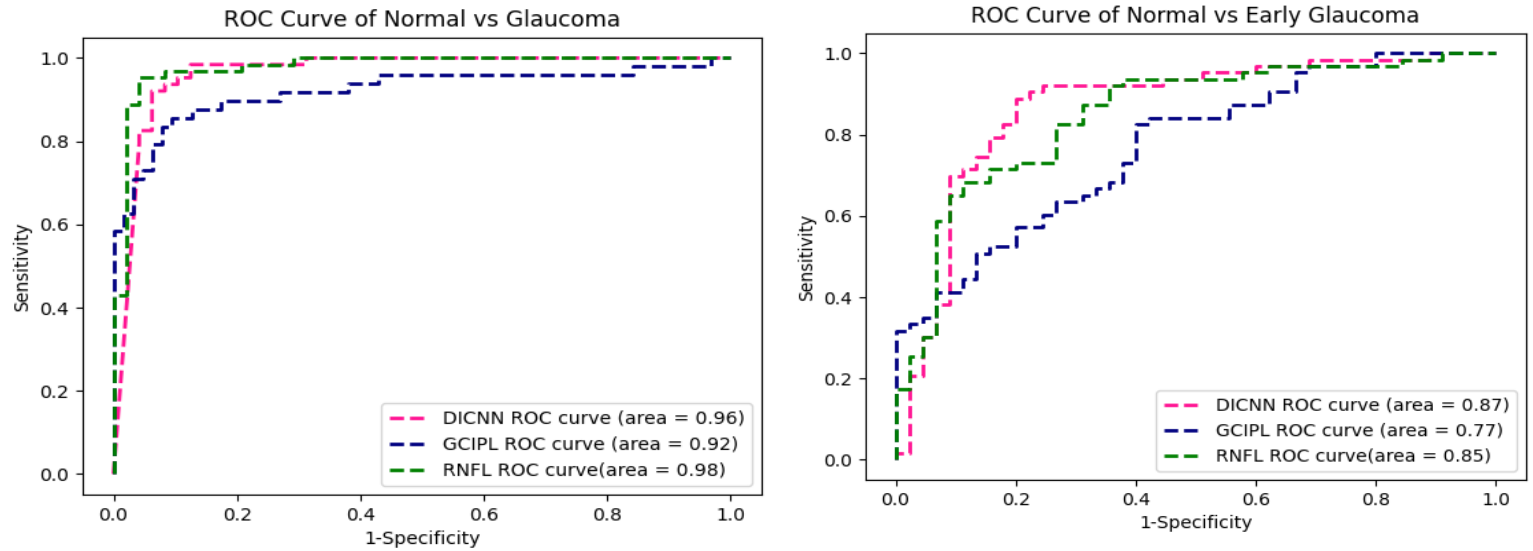
An overview of datasets is shown in Figure 2.5 The dataset used in this study (training and testing) included 777 eyes of 777 subjects (462 eyes of 462 subjects with POAG and 315 eyes of 315 normal subjects). Among the 462 POAG eyes, 219 were early-stage, and 243 were moderate-to-severe stage. For the training dataset, 621 image sets in total (252 normal, 174 early POAG, 195 moderate-to-severe POAG) were used. During the testing, 156 independent image sets (63 normal, 45 early glaucoma, 48 moderate-to-severe POAG) were assessed by each of the deep-learning algorithms. The subjects' demographic and ocular characteristics are provided Table 2.1.



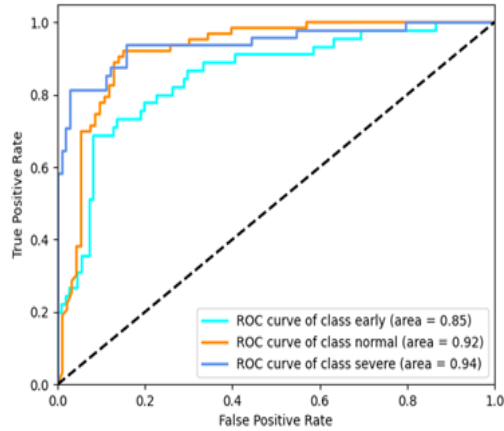
**Figure 2.6 Test accuracy obtained by different optimizers**



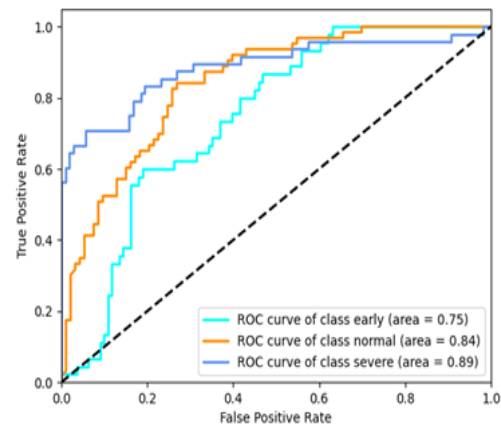
The validation accuracies obtained by other optimizers (i.e., Adam,[30] Adadelata,[31] and Adagrad[31]) are shown in Figure 2.6



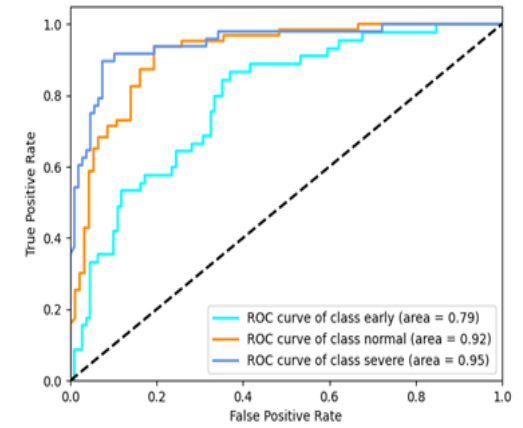
**Figure2.7 Receiver operating characteristic (ROC) curve of DICNN with both retinal nerve fiber layer (RNFL) and ganglion cell-inner plexiform layer (GCIPL), CNN with single image (RNFL or GCIPL: (left) distinguishing between normal subjects and glaucoma patients; (right) between normal subjects and early-stage glaucoma patients.**



A. DICNN

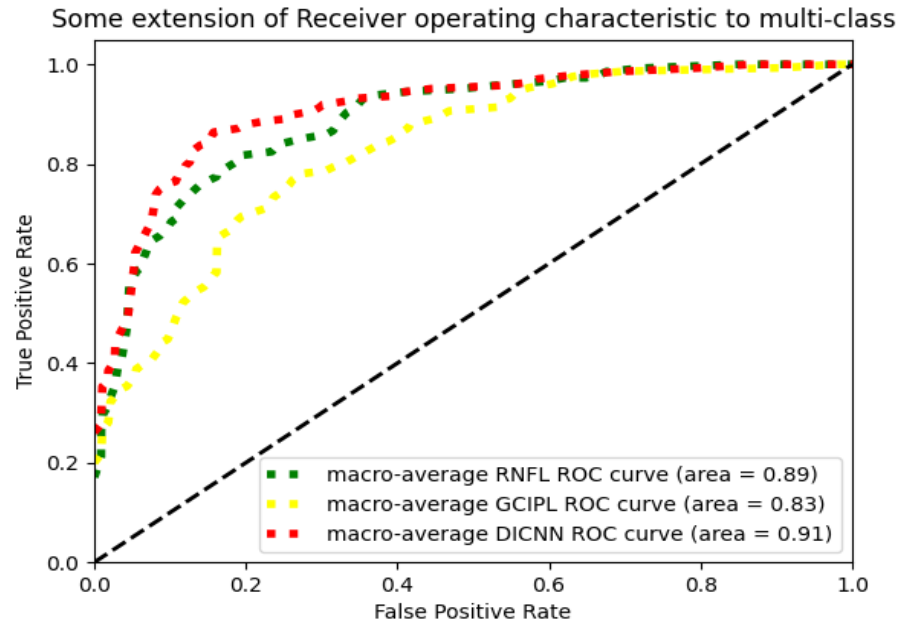


B. GCIPL

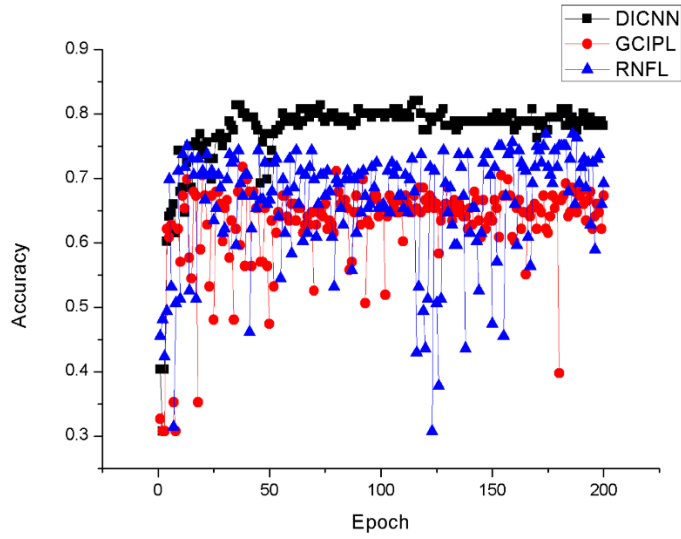


C. RNFL

**Figure 2.8 Receiver operating characteristic (ROC) curve of DICNN (one vs rest) and with both retinal nerve fiber layer (RNFL) and ganglion cell-inner plexiform layer (GCIPL), CNN with single image (RNFL or GCIPL): comparison with one vs rest.**

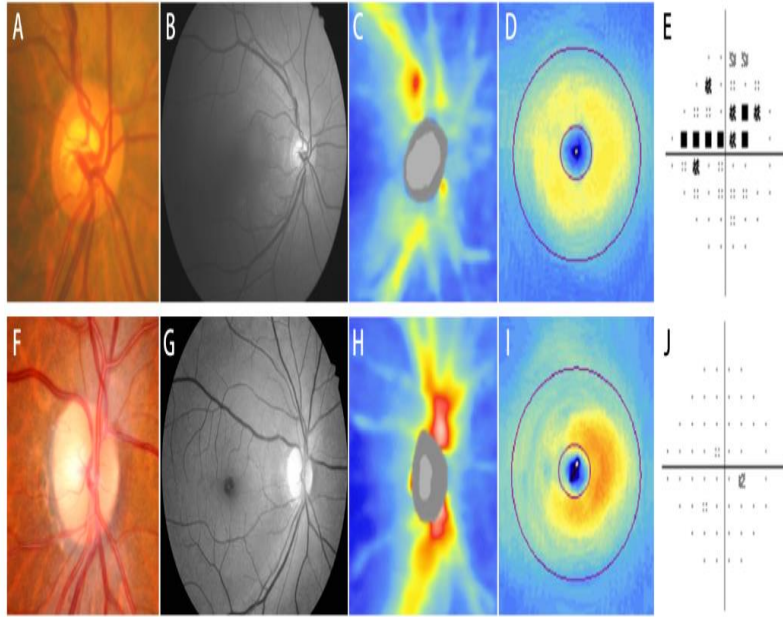


**Figure2.9 Receiver operating characteristic (ROC) curve of DICNN (macro averaging) and with both retinal nerve fiber layer (RNFL) and ganglion cell-inner plexiform layer (GCIPL), CNN with single image (RNFL or GCIPL): comparison with one vs rest (macro averaging each class).**



**Figure2.10 Test accuracy for DICNN (black square) and CNN with GC IPL (red dot) or RNFL (blue triangle), respectively.**

DICNN converges faster and overall test accuracy shifted to higher region.



**Figure 2.11 Representative cases of DICNN's misclassification between early-stage POAG patients (A-E) and normal subjects (F-J).** The first row shows the early-stage POAG findings misdiagnosed as normal images by DICNN. Inferotemporal optic disc rim narrowing (A) along with RNFL defect (B) with corresponding superior visual field defect (E) are shown. On that basis, a diagnosis of glaucoma was made. Note that in fact, the inferotemporal RNFL defect had a relatively long angular distance from the macula. The second row shows the normal subject misclassified as early-stage POAG by DICNN. The major temporal blood vessels were more temporally located, thereby causing shifting of peaks on the RNFL thickness plot.

The accuracy and AUC calculated for DICNN are shown in Table 2.2, which also indicates the results obtained by the trained VGG16 with RNFL or GCIPL,

respectively. The ROC curves for all of the analysis algorithms are plotted in Figure 2.7 The accuracy and AUC were 92.793% and 0.957 (95% CI, 0.943-0.966), the sensitivity was 0.896 (95% CI, 0.896-0.917), and the specificity was 0.952 (95% CI, 0.921-0.952) for DICNN's distinguishing between normal and glaucoma subjects.

DeLong's test for two correlated ROC of DICNN and RNFL was  $Z = -0.77793$ ,  $p\text{-value} = 0.4366$ , GCIPL and RNFL was  $Z = -2.4482$ ,  $p\text{-value} = 0.01436$ , DICNN and GCIPL was  $Z = 1.8963$ ,  $p\text{-value} = 0.05792$ . All correlated ROCs are not significantly different.

As for VGG16 with RNFL, the accuracy and AUC were 94.595% and 0.978 (95% CI, 0.954-0.978), the sensitivity was 0.938 (95% CI, 0.771-0.938), and the specificity was 0.952 (95% CI, 0.937-0.981); for VGG16 with GCIPL, the accuracy and AUC were 88.288% and 0.921 (95% CI, 0.861-0.921), the sensitivity was 0.833 (95% CI, 0.614-0.833), and the specificity was 0.921 (95% CI, 0.857-0.933).

For another comparison group in Table 2.3 the final 3calss test accuracy of comparison group, Inception V3 with RNFL was 0.75, ResNet50 with RNFL was 0.7756, Inception ResNet-V2 with RNFL was 0.7756, DensNet121 with RNFL was 0.7756, Xception with RNFL was 0.7564 DICNN with both image was 0.821, DICNN outperformed comparison group.

For DICNN's distinguishing between normal and early-stage glaucoma subjects, the accuracy and AUC were 85.185% and 0.869 (95% CI, 0.825-0.879), the sensitivity was 0.921 (95% CI, 0.813-0.905), and the specificity was 0.756 (95% CI, 0.610-0.790); for VGG16 with RNFL, the accuracy and AUC were 80.000% and 0.850 (95% CI, 0.739-0.850), the sensitivity was 0.905 (95% CI, 0.877-0.952), and the specificity was 0.644 (95% CI, 0.356-.644); for VGG16 with GCIPL, the accuracy and AUC were 72.900% and 0.767 (95% CI, 0.721-0.784), the sensitivity was 0.839 (95% CI, 0.742-0.790), and the specificity was 0.578 (95% CI, 0.556-0.639). DeLong's test for two correlated ROC of DICNN and RNFL was  $Z = 0.5141$ ,  $p\text{-value} = 0.6072$ , GCIPL and RNFL was  $Z = -1.5526$ ,  $p\text{-value} = 0.1205$ , DICNN and GCIPL was  $Z = 2.424$ ,  $p\text{-value} = 0.01535$ . All correlated ROCs are not significantly different except DICNN and GCIPL.

Evaluating DICNN's performance with one vs rest, Table 2.4, with class of early, AUC was 0.85, specificity 0.898, sensitivity 0.689 with class of normal AUC was 0.92. specificity 0.833, sensitivity 0.92. with class of severe AUC was 0.94, specificity 0.967, sensitivity 0.813. DICNN outperformed VGG16 with RNFL, VGG16 with GCIPL.



### 2.3.2 Grad-CAM for DICNN

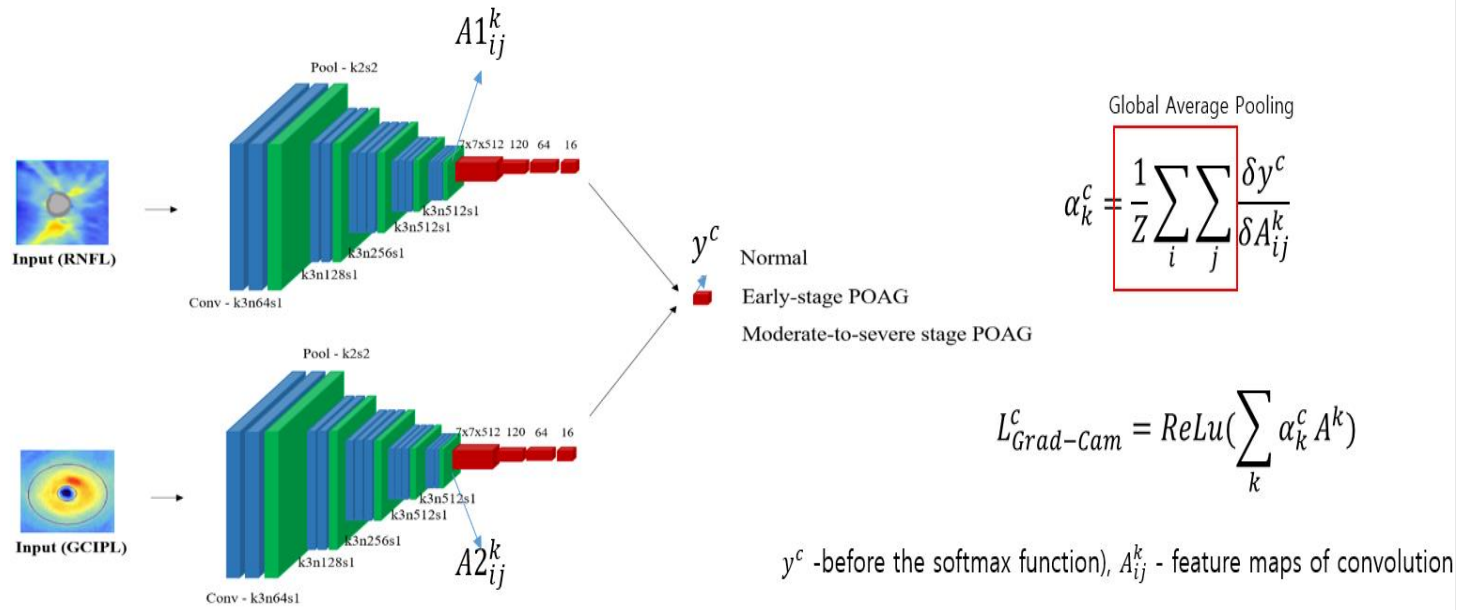
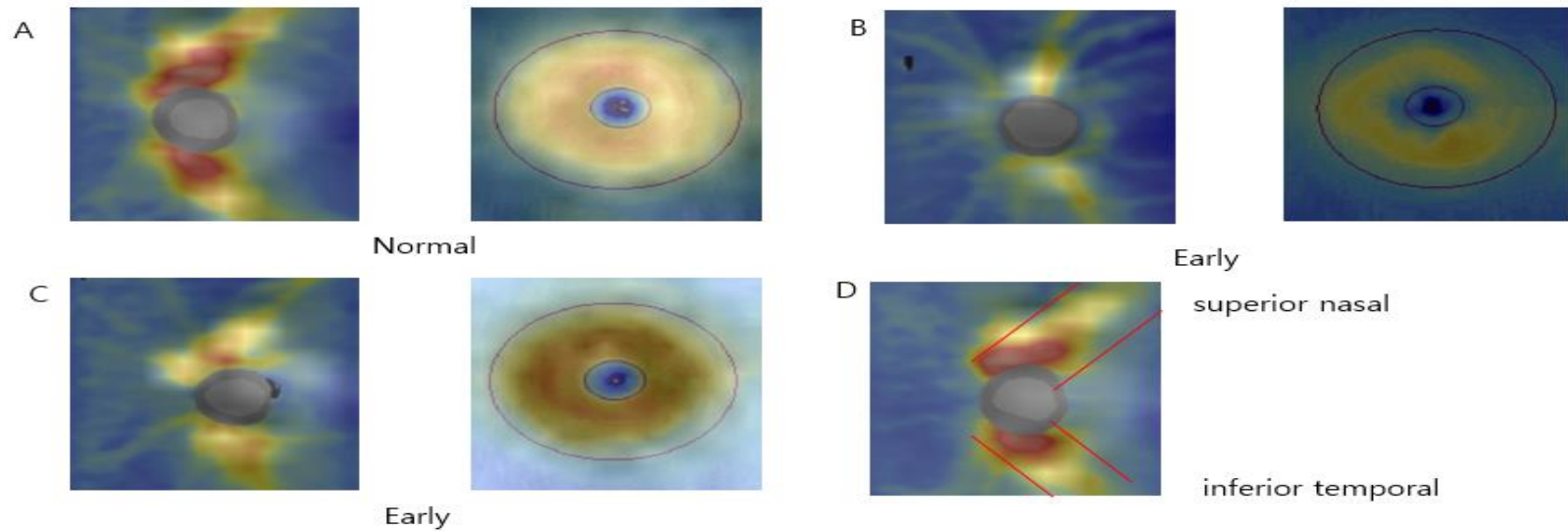


Figure2.12 Calculating Grad-CAM for DICNN

Grad-CAM was calculated with feature maps of last convolution layer( $A1_{ij}^k$ ) with output value  $y^c$  representing classification score of 3class classification



**Figure2.13 Grad-CAM for test images.**

Attention is focused on the superior nasal and inferior-temporal region(D)

Gradient Class Activation Map(Grad-CAM)[32] was calculated gradient between output logit before softmax( $y^c$ ) and feature map of last convolution  $A1_{ij}^k$  (RNFL), and  $y^c A1_{ij}^k$ (GCIPL). Figure 2.13 shows that attention region is located at superior nasal and inferior temporal.

## **2.4 Discussion**

### **2.4.1 Research Significance**

In the present study, we determined that DICNN combined with features of RNFL and GCIPL thickness was effective in distinguishing glaucomatous from normal SD-OCT images. DICNN was developed by combining the features of both RNFL and GCIPL those extracted from last-fully-connected layer from VGG16. Notably, DICNN differentiated between early-stage glaucoma and normal subjects more accurately than did the other methods.

VGG16 with RNFL or GCIPL, respectively, showed similar results for differentiation of glaucoma patients and normal subjects to a certain extent, though the data amount and glaucoma-diagnostic criteria differed from the previous research.[18-20, 33] Muhammad et al.[34] utilized a hybrid deep-learning method (HDLM) according to a single wide-field OCT protocol. Their combination of RNFL and GCIPL probability images, however, did not yield diagnostic performance that was superior to that of RNFL probability alone. They suggested that a novel method of combining RNFL with GCIPL information could be a feasible means of enhancing diagnostic performance. In our study, the deep-learning algorithms analyzed by DICNN with combined RNFL and GCIPL features on thickness maps showed better diagnostic

performance for early-stage glaucoma relative to the conventional deep-learning methods.

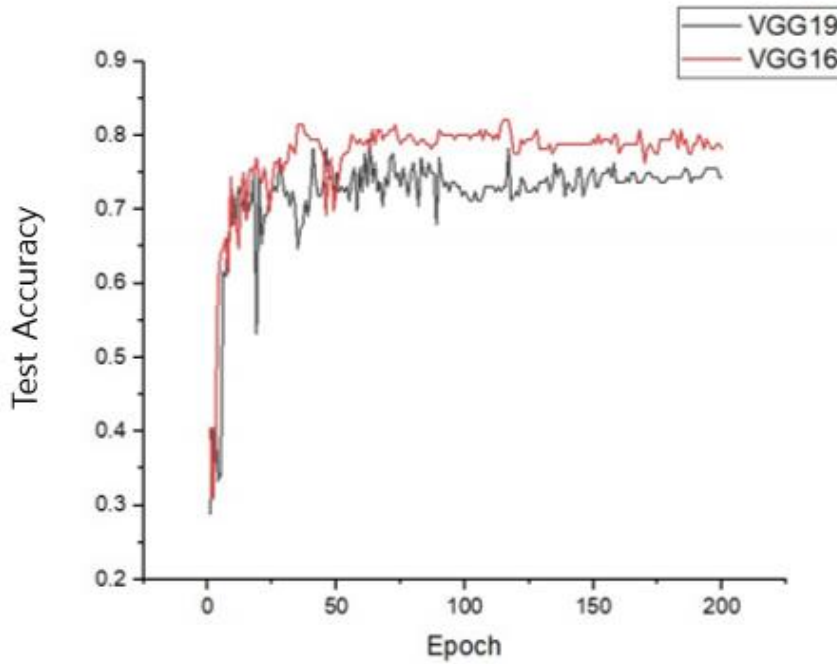
Early detection of glaucoma is essential, certainly, but early-onset glaucoma is much more challenging to diagnose than is advanced glaucoma, since early structural damage sometimes is revealed in one parameter but not another. Wang et al.[35] reported that their RNFL thickness evaluation missed 47 (77%) of the 61 eyes showing abnormal thinning on macular scans. The typical analyses of peripapillary RNFL thickness alone can overlook glaucomatous macular damage.[13, 36] On the other hand, it is known that macular GCIPL parameters are less sensitive in cases where the temporal margin of the defect is located far from the fovea.[37] Therefore, in terms of detecting subtle early-glaucomatous damage, more combined information is needed. In fact, based on our results, the features extracted by both RNFL and GCIPL boosted the classification results for early-glaucomatous eyes by an average MD -2.83 dB

The higher accuracy along with the larger AUC of the network trained by DICNN compared with VGG16 (Figure 2.7,8,9) was perhaps owed to the DICNN's having more diverse features for analysis of early-glaucoma patients and normal subjects. The RNFL and GCIPL thickness maps present 'related but different' clinical information. DICNN integrated variables from the RNFL and GCIPL thickness maps for training in order to identify more underlying features inside the images. Since the final output values of the fused nodes were trained

through backpropagation with softmax cross-entropy loss, the entire network could share variables in both the RNFL and GCIPL thickness maps. The fused features in the last 16 fully connected layers (fully connected layer of 120, 64, 16) of DICNN had the best accuracy and AUC, as shown in Figure 2.10. If there were too many last-fully-connected layers (above 16), DICNN would not converge. On the other hand, with too few fully connected layers, the trained network would show poor classification accuracy. With DICNN, the network converges more reliably than with VGG16 trained on either RNFL or GCIPL (Figure 2.10).

Our study has some advantages relative to the previous work. First, we used a class-balanced dataset for the early-glaucoma, moderate-to-severe glaucoma and normal groups. In that way, we were able to minimize the problems incurred from biased training and estimation.[34] Second, we utilized both RNFL and GCIPL information for training of the deep-learning model. As structural damage sometimes is revealed in only one parameter, not also in the other, it is important that in glaucoma diagnostics, information is gathered for both the RNFL and GCIPL.[38] Likewise, a deep-learning model that is trained with both RNFL and GCIPL images would be expected to offer better diagnostic performance.

## 2.4.2 Limitations



**Figure 2.14 Test accuracy obtained by VGG16 and VGG19 as baseline model**

Our study has several limitations. First, the training results were based on a relatively small amount of data. In the current study, we increased the size of the dataset by augmentation (i.e., by adding Gaussian noise), thereby improving the network's diagnostic accuracy for glaucoma.[39] However, geometric transformation techniques such as flip, rotation, or inversion were not applied. Clinically, the specific location of optic neuropathy is an important basis for glaucoma diagnosis, since glaucomatous optic neuropathy generally begins in the inferotemporal disc region and then progresses to the superotemporal and other regions.[40] We therefore applied only the Gaussian

noise, which preserves the spatial dimensions of the images.[41] Further studies will validate this algorithm using a larger dataset. Second, the performance of our deep-learning algorithm is dependent on the quality of the training images. Exclusion of low-quality images from the training set could limit the usefulness of the algorithm in actual clinical settings. Third, we used VGG16 as the baseline model in order to fuse the fully connected layer, since it showed higher accuracy than VGG19's (Figure 2.14). It has been reported that deeper networks usually have better classification performance. The reasons for the different results in our study are not clear, but the number of images is one of the possible ones, since the improved performance of deeper models might not be evident in smaller datasets[42] and FC layer has a lot of parameters compared to convolution, so in the future work it is necessary to use Siamese Neural Network training[43] or training whole layer with concatenating the output of convolution layer[44]. The performance differences of different neural network architectures in larger datasets should be evaluated in further studies. Fourth, it remained unclear how the deep-learning model discriminated cases as either glaucoma or normal. There have been many attempts to visualize exactly what a deep-learning network learns, such as by creating heat maps,[45, 46] which process will certainly be applied in a future clinical study.



## 2.5 Conclusion

In conclusion, we constructed a deep-learning model incorporating DICNN as trained by both RNFL and GCIPL thickness map data, and compared and validated the high diagnostic ability for discrimination of early-stage glaucoma patients from normal subjects. It might be possible to further improve the model's performance by increasing the size as well as diversification of the training set.

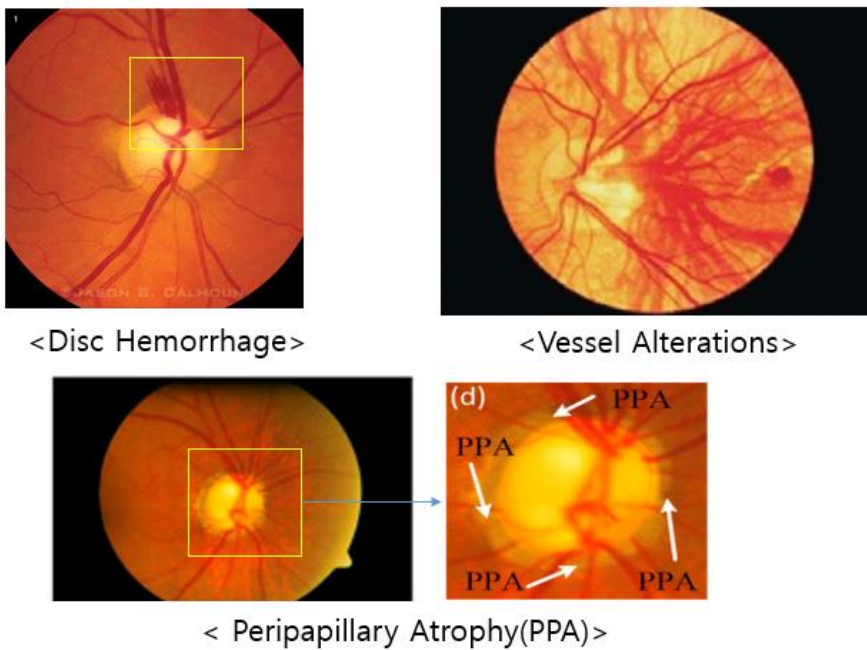
\* Large sections of this chapter were published previously in *British Journal of Ophthalmology*. Sun, S., Ha, A., Kim, Y. K., Yoo, B. W., Kim, H. C., & Park, K. H. (2020). Dual-input convolutional neural network for glaucoma diagnosis using spectral-domain optical coherence tomography. *British Journal of Ophthalmology*.)

## Chapter 3

# Deep-learning-based enhanced optic-disc photography

### 3.1 Introduction

#### 3.1.1 Background



**Figure 3.1 Representative images of Optic Disc Hemorrhage(DH), Peripapillary chorioretinal atrophy(PPA), vessel alterations**

Optic nerve head (ONH) examination is essential to glaucoma diagnosis and progression assessment[2, 47]. Optic-disc photography (ODP) has been proven

to be very effective for documentation of optic nerve appearance, as it allows for more detailed scrutinization and subsequent comparison for determination of progressive change [48-50]. Furthermore, ODP enables clinicians to qualitatively assess ONH structures such as detailed neuroretinal rim contours, presence of optic disc hemorrhage (DH), parapapillary chorioretinal atrophy (PPA) or vessel alterations, which is not possible in optical coherence tomography (OCT) [51].

### 3.1.2 Needs



**Figure3.2 Low Quality ODPs**

Out of focus(left), low brightness(middle), high brightness(right)

In real clinical practice, limited patient cooperation, small pupils, or media opacities can limit the performance of ODP[52]. As a result, ODPs can have several limitations, such as insufficient resolution, low color contrast, and inconsistency of image quality (especially in cases of media opacity due to cataracts). Even when high-resolution ODPs can be obtained, red-colored blood vessels and red-orange-colored retina sometimes cause indistinct pathologies

such as small-sized DH to be missed. ODP-quality improvement techniques that can obviate the limitations of the current imaging acquisition devices are essential, especially when considering the indispensability of ONH structural evaluation in glaucoma treatment.

### **3.1.3 Related Work**

The popularity of deep-learning algorithms offering modeling of high-level abstractions in data by means of multiple processing layers has exploded in recent years as powerful graphics processing units (GPUs) have become available. The very intricate process of high-resolution image estimation from a low-resolution counterpart is known as super-resolution (SR) [53, 54]. For image SR, generative adversarial network (GAN), which is a deep neural net architecture comprising two nets one pitted against the other (hence “adversarial”), has shown great utility and potential[55].

## **3.2 Methods**

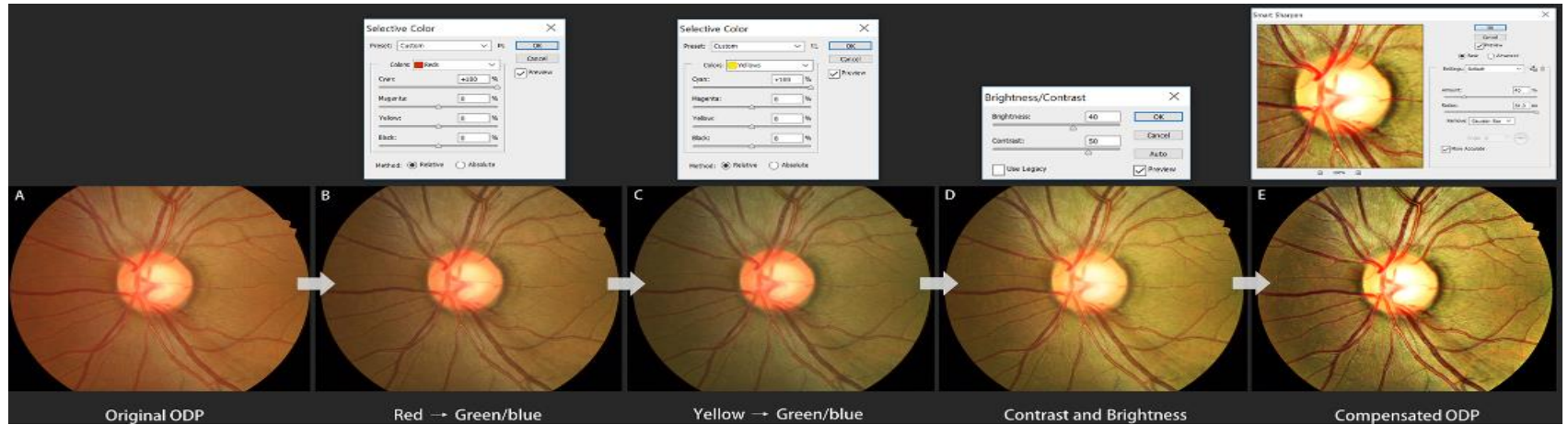
### **3.2.1 Study Design**

In this paper, we propose a modified super-resolution generative adversarial network (SR-GAN) that is capable not only of up-scaling but also of improving ODPs' details as well as the visibility of the optic disc margin and surrounding retinal vessels in computing 'enhanced' ODPs. In the present study, we performed a quantitative evaluation to assess enhanced ODPs' clinical utility

### **3.2.2 Dataset**

ODPs were obtained post-pupil-dilation using a digital fundus camera system (CF-60UVi/D60; Canon, Inc., Tokyo, Japan). The images were saved in the 384 x 384-pixel digital imaging and communications in medicine format and stored in the picture archiving communication system (PACS) of Seoul National University Hospital.

### 3.2.2.1 Details on Customized Image Post-Processing Algorithm



**Figure 3.3 Customized image post-processing algorithm for maximized visibility of hemorrhage.** (A) Original ODP. (B) With Selective color tool, red color was replaced by green/blue and (C) yellow was replaced by green/blue. (D) With the Contrast/Brightness tool, the contrast level was improved, and with the Smarten sharpen tool, the degree and range of the sharpness was increased. (E) Finally, a compensated high-resolution ODP could be obtained.

The purpose of this type of processing is to generate an improved ODP image in terms of both color and spatial contrast. The processing entails the following steps: contrast optimization, edge enhancement, spatial and frequency filtering, image combining, and noise reduction. Detailed manual adjustment has to be applied differently according to each ODP's image quality. In the present study, all of the image post-processing was performed using a commercial image-processing tool (Adobe Photoshop CS3, version 10.0.1) by a single glaucoma-image-processing specialist (YKK). In detail, the histogram data of the downloaded high-resolution original ODPs (384 x 384 pixels) were evaluated to determine whether an image was over- or underexposed, flat (i.e., of little contrast), and the tonal range in which image adjustment was required. Then, using the Curves tool (specifically by clicking on the image Levels curve and dragging on it), the tonal ranges of an image were adjusted to improve its details and fine structures. Next, the visibility of retinal vessels or DH was enhanced by adjustment of the contrast and brightness between the blood vessels and background fundus. With the Selective color tool, red and yellow colors were completely replaced by green/blue (i.e., the blood vessel color was changed from red to bright red, and conversely, the background retinal color was changed from red-orange to light brown in order to maximize the visibility of hemorrhage). Then, with the Contrast/Brightness tool, the contrast and level were improved, and with the Smarten sharpen tool, the degree and range of the

sharpness were increased (Figure 3.3). Finally, the ‘compensated high-resolution ODP’ could be obtained.

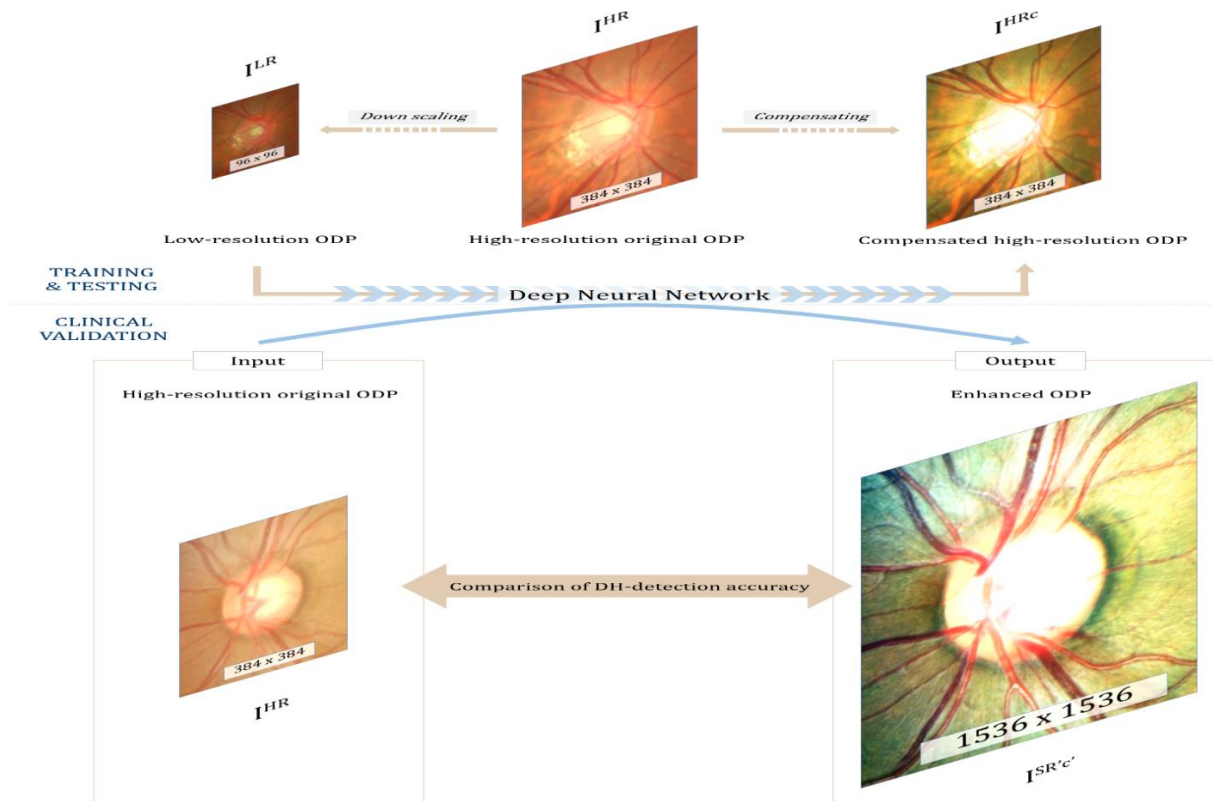
This study was approved by the Seoul National University Hospital Institutional Review Board (1805-027-944) and faithfully adhered to the tenets of the Declaration of Helsinki. All of the subjects provided their written informed consent. Eyes were chosen from subjects examined for glaucoma at the Glaucoma Clinic, Seoul National University Hospital, between January and December 2018. All of the relevant data are in the manuscript and its Supporting Information files.



### 3.2.3 SR-GAN Network

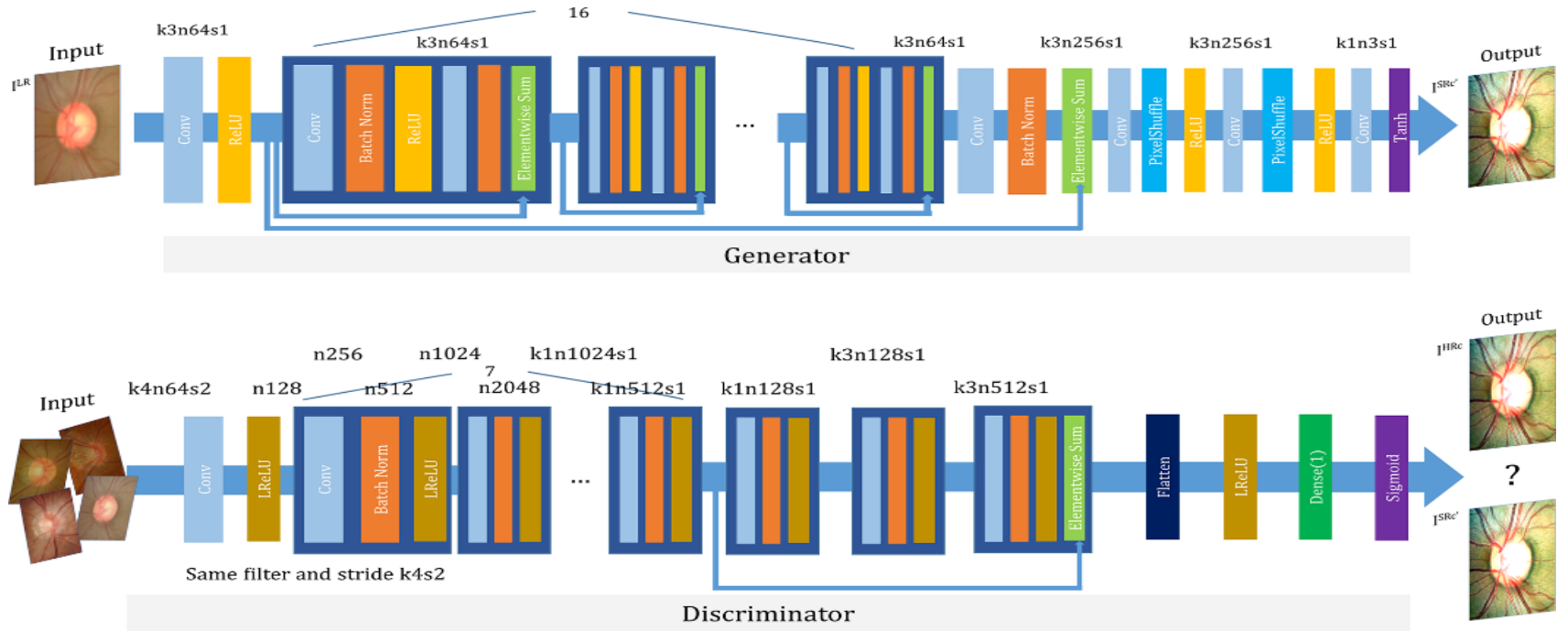
#### 3.2.3.1 Design of Generative Adversarial Network

An  $I^{SR'c'}$  is a super-resolved, compensated image, and an  $I^{HR}$  is the high-resolution original image. The final goal of this study was to obtain  $I^{SR'c'}$  (1536 x 1536 pixels) from an  $I^{HR}$  (384 x 384 pixels) for clinical validation (see Figure 3.4). For that purpose, the differences between the  $I^{HRc}$  (384 x 384 pixels) and  $I^{LR}$  (96 x 96 pixels) were learned directly through a modified SR-GAN. An  $I^{LR}$ , which is the low-resolution version of the  $I^{HR}$ , was obtained by 1/4 resizing of the  $I^{HR}$  using bicubic interpolation (down-scaled width and height:  $1/4W \times 1/4H \times C$ ) [56], the  $I^{HRc}$  being the high-resolution image manually customized by a post-processing algorithm. The SR-GAN consists of a GAN and a pre-trained VGG19 (Visual Geometry Group) network[57]



**Figure 3.4 Principle of enhanced image formation via super-resolution generative adversarial network (SR-GAN).**

A modified SR-GAN was used to learn the differences between the low-resolution optic-disc photography (ODP) and the manually compensated high-resolution ODP. By inputting the high-resolution original ODP into the algorithm, an X4 up-scaled and overall contrast-, color- and brightness-transformed ‘enhanced ODP’ could be obtained.



The GAN includes an additional discriminator for evaluation of the generator's reliability [58]. The discriminator makes a judgement on whether a randomly inputted image is a guess of the generator or a high-resolution measurement. For optimized discriminator judgement, an adversarial loss is created that iteratively optimizes the discriminator for enhanced decision-making capability. Also, the adversarial loss, together with the content loss, are used to optimize the generator in pushing it in the direction in which more perceptually realistic outputs can be generated to further fool the discriminator [55]. By this process of adversarial training, the quality of images from the generator can be improved. The training is terminated once the generator produces results that the discriminator cannot distinguish from the high-resolution images[59]. The generator and discriminator network architecture with the corresponding kernel size (k), number of feature maps (n) and stride (s) is shown in Figure 3.5 We applied Tensorlayer SubpixelConv2d as a PixelShuffle [60].

Each original ODP ( $I^{HR}$ ) was transformed into two counterparts: (1) down-scaled 'low-resolution ODPs ( $I^{LR}$ , 96 x 96 pixels)' and (2) 'compensated high-resolution ODPs ( $I^{HRc}$ , 384 x 384 pixels)' produced via enhancement of the visibility of the optic disc margin and surrounding retinal vessels using a customized image post-processing algorithm. Then, the differences between the two were directly learned through the modified SR-GAN. Finally, by inputting of the high-resolution original ODPs ( $I^{HR}$ , 384 x 384 pixels) into the trained SR-

GAN, 4-times-up-scaled and overall contrast-, color-, and brightness-transformed ‘enhanced ODPs ( $I^{SR'}$ , 1536 x 1536 pixels)’ could be obtained.

### 3.2.3.2 Loss Functions

Our ultimate goal was to train a generating function  $G$  by training a generator network as a feed-forward CNN  $G_{\theta_G}$  parametrized by  $\theta_G$ . Here,  $\theta_G$  denotes the weight and bias of the designed network, and is obtained by optimizing loss function  $l^{SR}$ . The sum of loss functions,  $l^{SR}$ , is obtained. For training of image  $I_n^{Obtained}$  with corresponding  $I_n^{Target}$   $n = 1, 2, 3 \dots N$ , the following equation is solved:

$$\widehat{\theta}_G = \underset{\theta_G}{argmin} \frac{1}{N} \sum_{n=1}^N l^{SR}(G_{\theta_G}(I_n^{Obtained}), I_n^{Target}) \quad (1)$$

First, the pixel-wise Mean Squared Error (MSE) loss was calculated as follows:

$$L_{MSE} = \frac{1}{r^2WH} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (I_{x,y}^{Target} - G_{\theta_G}(I^{Obtained})_{x,y})^2 \quad (2)$$

where  $W_{i,j}$ ,  $H_{i,j}$  are the width and height, respectively, of the feature map.

MSE loss, widely utilized for image SR, calculates the squared difference in

pixels between the obtained and target images during the training process. However, MSE optimization results in blurring of the edges of the generated image.

Therefore, we also adopted VGG loss as defined with the pre-trained (trained with ImageNet) VGG19 Network. VGG loss was calculated as follows:

$$L_{VGG/i,j} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{Obtained})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{Target}))_{x,y})^2 \quad (3)$$

where  $\phi_{i,j}$  indicates the feature maps of the pre-trained VGG19 Network after the  $j^{\text{th}}$  convolution and before the  $i^{\text{th}}$  maxpooling layer. In this study, we employed feature maps of conv4\_3 ( $j = 4, i = 3, W_{i,j} = 28, H_{i,j} = 28$ ). VGG loss was used to calculate the squared difference between the feature maps of the target and generated images via SR-GAN. By using both MSE and VGG loss, the overall resolution and style of the generated image could be improved [61].

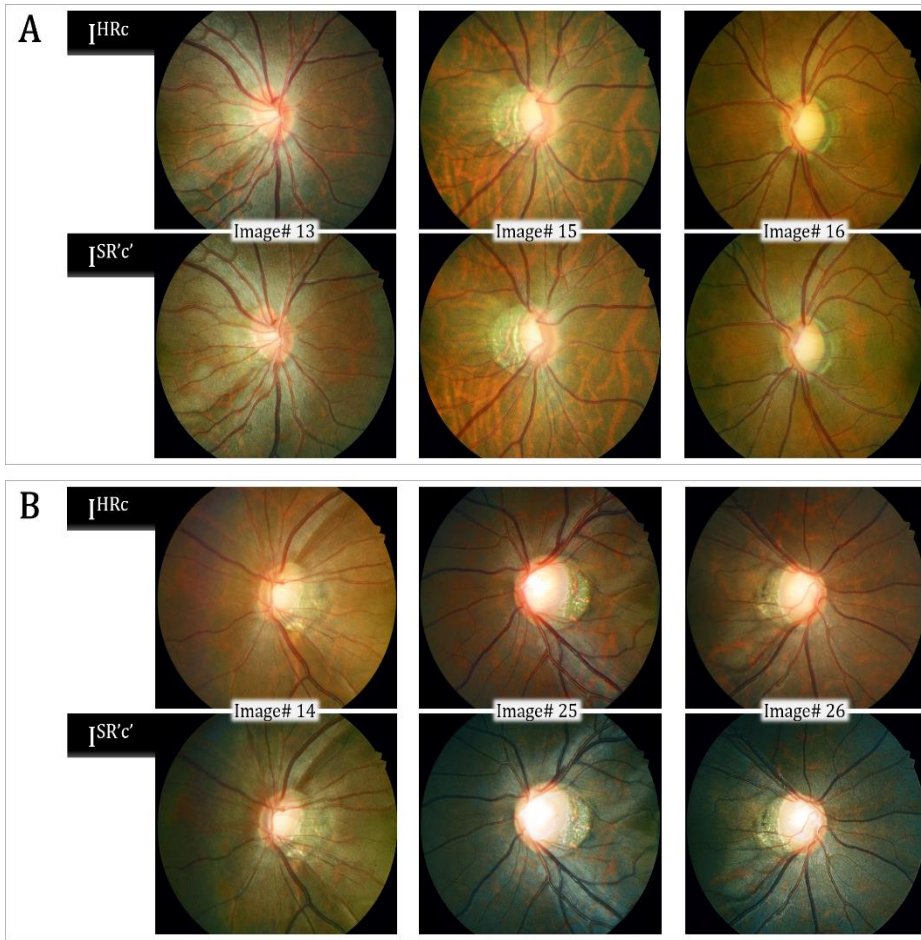
Finally, the adversarial loss function computes the Sigmoid Cross Entropy (SCE) loss by calculating the difference between the output logits of the generated image ( $G_{\theta_G}(I^{LR}))$  and target image to fool the discriminator, as follows:

$$L_{Gen} = \sum_{n=1}^N -\log D_{\theta_D}(G_{\theta_G}(I^{LR})) \quad (4)$$

where  $D_{\theta_D}(G_{\theta_G}(I^{LR}))$  is the probability of  $(G_{\theta_G}(I^{LR}))$  to be considered as a target image. The generator tries to fool the discriminator by generating higher-quality images. The final goal of adversarial loss is the minimization of  $-\log D_{\theta_D}(G_{\theta_G}(I^{LR}))$ .

### 3.2.4 Assessment of clinical implications of enhanced ODPs





**Figure 3.6 Validation results for representative test image sets.** The Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) values were higher in the upper 3 sets, and the obtained optic-disc photography (ODPs) were perceptually similar to the targeted ODPs (A). In the lower image sets, the change in background color caused relatively lower PSNR and SSIM values, even though those images were perceptually convincing (B).

For the test, 50 high-resolution original ODPs and 50 paired SR-GAN-enhanced ODPs in two respective datasets were used. Three glaucoma

specialists (AH, JL and KHP) independently evaluated the original ODPs of the test datasets and confirmed a total of 23 DHs in 23 original ODPs. Then, 12 general ophthalmologists were asked (1) to assess ODP image quality in 5 grades (excellent, good, fair, poor or bad), and (2) to note, for each of the original ODPs and enhanced ODPs separately at 1-month intervals, any abnormal findings including DH. In the process of the image quality grading, ‘excellent’ was defined as a clearly identified optic disc margin and distinct major vessel structures, while ‘bad’ was defined as unidentifiability of the optic disc margin. The ranges from good to fair quality and from fair to poor quality were determined subjectively by each ophthalmologist. The 5 grades were numbered between 1 (‘poor’) and 5 (‘excellent’); then, we performed a mean opinion score (MOS) test to compare the qualitative assessments in and among the image groups[62]. Figure3.6 each compare an original ODP image with its enhanced version.

### 3.2.5 Statistical Analysis

All of the values are presented as means  $\pm$  standard deviation. Paired  $t$  tests were used to determine the MOS differences between the two image types. The Mann-Whitney test was applied for comparison of the nonparametric data. The categorical data were analyzed by  $\chi^2$  test, and a statistical analysis was performed using the SPSS statistical package (SPSS 22.0; Chicago, IL, USA.). A 2-sided  $P$ -value  $< 0.05$  was considered to be statistically significant.

### 3.2.6. Hardware Specifications & Software Specifications

#### Hardware specifications

CPU: Intel core i7-7700 3.60Hz x 8

GPU: TITAN X (Pascal) 12GB

RAM: 16GB

#### Software specifications

Deep-learning libraries:

Tensorflow – 1.14.0 with cuda 10.0 and cudnn 7.6.3

Tensorflow Tensorlayer – 2.1.1

Python libraries (version - 3.6)

Numpy – 1.16.4 for model loading and array processing

Scipy – 1.1.0 for image loading, resizing, saving

Scikit-image – 0.15.0 for image transformation (augmentation)

Matplotlib – 3.1.1 for plotting image

Easydict – 1.9 for dictionary values as attributes

Os – for filename load

Pickle – for vgg19 model loading

Random – for train data shuffle

Time – for calculating time for data loading, training, testing time

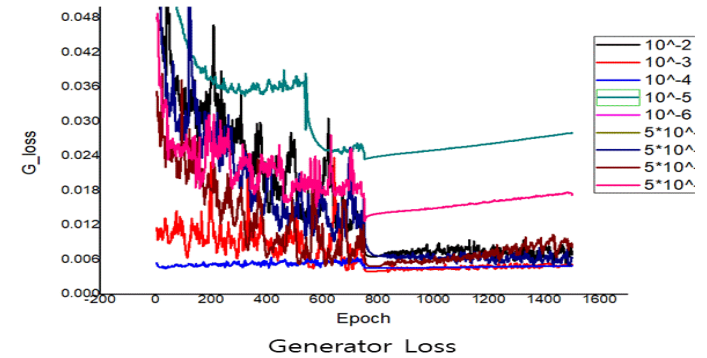
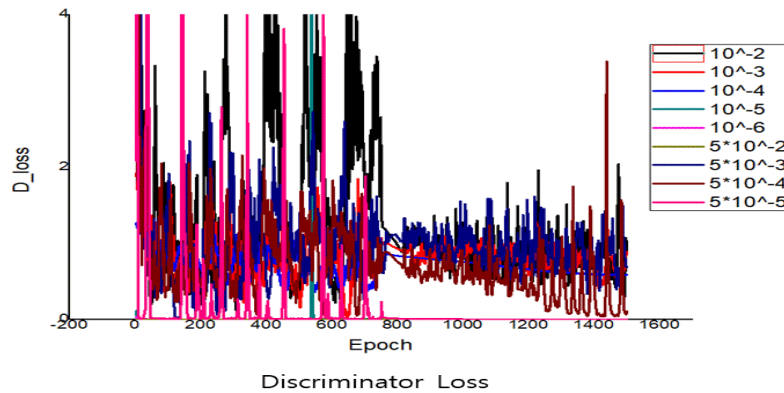
Optimizer:

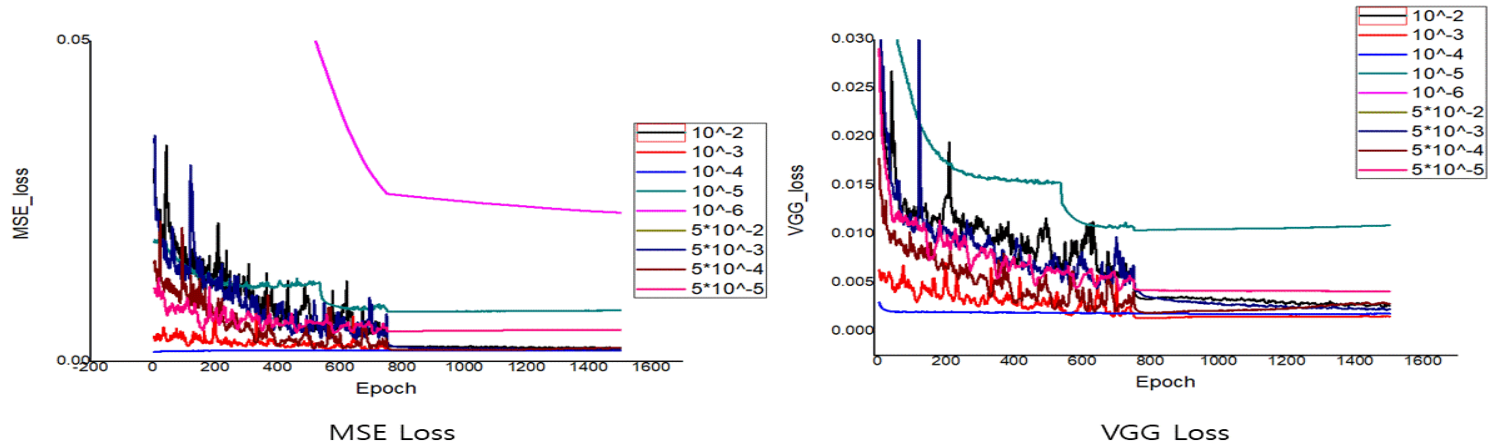
Generator: AdamOptimizer – learning rate of  $5 \times 10^{-3}$ , beta1 of 0.9

Discriminator: AdamOptimizer – learning rate of  $5 \times 10^{-3}$ , beta1 of 0.9

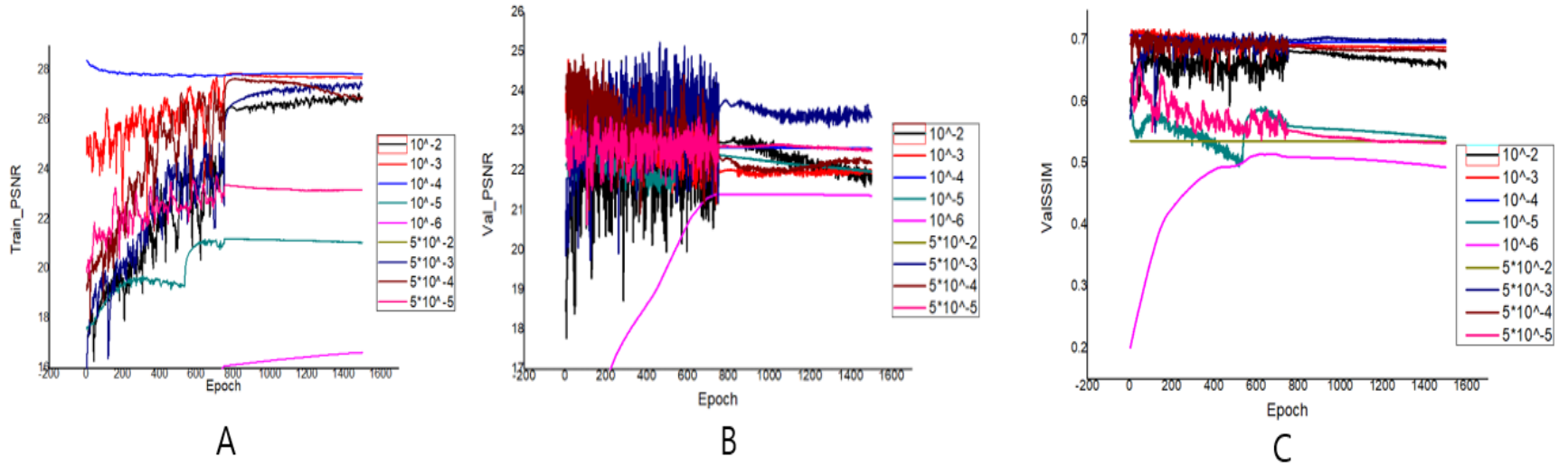
## 3.3 Results

### 3.3.1 Training Loss of Modified SR-GAN





**Figure 3.7 Training curve for SR-GAN algorithm with learning rate.** The colored line shows the loss of the discriminator, generator, VGG, MSE loss over the training course, while each line represents the learning rate of model. As can be seen, the lowest loss was at learning rate of  $5 \times 10^{-3}$ .



**Figure 3.8 Training PSNR curve and validation PSNR and SSIM curve for SR-GAN algorithm with learning rate.** The colored line shows the training PSNR, validation PSNR and SSIM over the training course, while each line represents the learning rate of model. As can be seen, the highest PSNR and SSIM was at learning rate of  $5 \times 10^{-3}$ .

Learning Rate	$10^{-2}$	$10^{-3}$	$10^{-4}$	$10^{-5}$	$10^{-6}$	$5 \times 10^{-2}$	$5 \times 10^{-3}$	$5 \times 10^{-4}$	$5 \times 10^{-5}$
Max Test									
SSIM(Epoch>850)	0.6839	0.6929	0.6975	0.5587	0.5104	0.5369	0.7063	0.6885	0.5488

---

Learning Rate	$10^{-2}$	$10^{-3}$	$10^{-4}$	$10^{-5}$	$10^{-6}$	$5 \times 10^{-2}$	$5 \times 10^{-3}$	$5 \times 10^{-4}$	$5 \times 10^{-5}$
Max Test									
PSNR(Epoch>850)	0.6839	0.6929	0.6975	0.5587	0.5104	0.5369	0.7063	0.6885	0.5488

**Table 3.1 Learning rate vs PSNR and SSIM table**

Learning rate of  $5 \times 10^{-3}$ , PSNR and SSIM was the highest.



Figure 3.7 depicts the generator and discriminator loss over the course of the training set epochs. After 800 epochs, the loss of the discriminator decreased and that of the generator increased. Up to the 800<sup>th</sup> epoch, the discriminator loss decreased and the generator loss increased for every 200 epochs, which is ideal for adversarial loss. After every 200 epochs, the generator loss abruptly decreased and the discriminator loss peaked. This is a characteristic of the Adam optimizer[30]. The value of the Adam optimizer soars when gradients are smaller and the whole denominator is smaller. Between the 800<sup>th</sup> and 1400<sup>th</sup> epochs, the loss values were stabilized. Accordingly, we trained our model to the 1500<sup>th</sup> epoch.

The SR-GAN model was optimized with test PSNR value Adam optimizer learning rate of (  $10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 5 \times 10^{-2}, 5 \times 10^{-3}, 5 \times 10^{-4}, 5 \times 10^{-5}$  ) Table 3.1. The model with the highest PSNR value was selected among the models after epoch 850, where the loss converged. At learning rate of  $5 \times 10^{-3}$ , The PSNR and SSIM value was the highest.

### 3.3.2 Performance of Final Network

We validated the enhanced ODPs according to Structural Similarity (SSIM) and Peak Signal-to-Noise Ratio (PSNR) and compared them with other machine-learning or state-of-the-art deep-learning methods[63]. SSIM is used to

calculate the similarity between two images based on three measurements:  
luminance, contrast and structure.

$$\mathbf{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (5)$$

Here,  $\mu_x$  and  $\mu_y$  are the averages of  $x$  and  $y$ ,  $\sigma_x^2$ ,  $\sigma_y^2$  are variances of  $x$  and  $y$ , and  $\sigma_{xy}$  is the covariance of  $x$  and  $y$ .  $c_1 = (k_1L)^2$ ,  $c_2 = (k_2L)^2$ , where  $L$  is the dynamic range of pixel values,  $k_1 = 0.001$  and  $k_2 = 0.003$ .

The optimization target of SR-GAN algorithms commonly is MSE minimization between the obtained and the targeted image. This is convenient, as minimizing the MSE also maximizes the PSNR, which is a measure commonly used to evaluate and compare SR algorithms.

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - k(i, j)]^2 \quad (6)$$

$$\text{PSNR} = 10 \log_{10} \left( \frac{MAX_I}{MSE} \right)^2 \quad (7)$$

Here,  $I(i, j)$ ,  $k(i, j)$  describe the original image and the target image, respectively.  $MAX_I$  is the maximum pixel value of the image, and in the present case, the  $MAX_I$  value was 255. Also, the PSNR is calculated as the ratio between the maximum signal power and the noise power.

<b>SSIM</b>	<b># 13</b>	<b># 15</b>	<b># 16</b>	<b># 14</b>	<b># 25</b>	<b># 26</b>	<b>mean</b>
Bicubic	0.89	0.90	0.92	0.92	0.90	0.91	0.91
SRRF	0.90	0.90	0.93	0.91	0.90	0.90	0.91
NBSRF	0.92	0.93	0.95	0.93	0.92	0.93	0.93
SRFBN	0.92	0.93	0.95	0.93	0.92	0.93	0.93
SRResNet	0.90	0.91	0.93	0.91	0.90	0.91	0.91
Modified SR-GAN	0.91	0.97	0.95	0.78	0.43	0.43	0.75

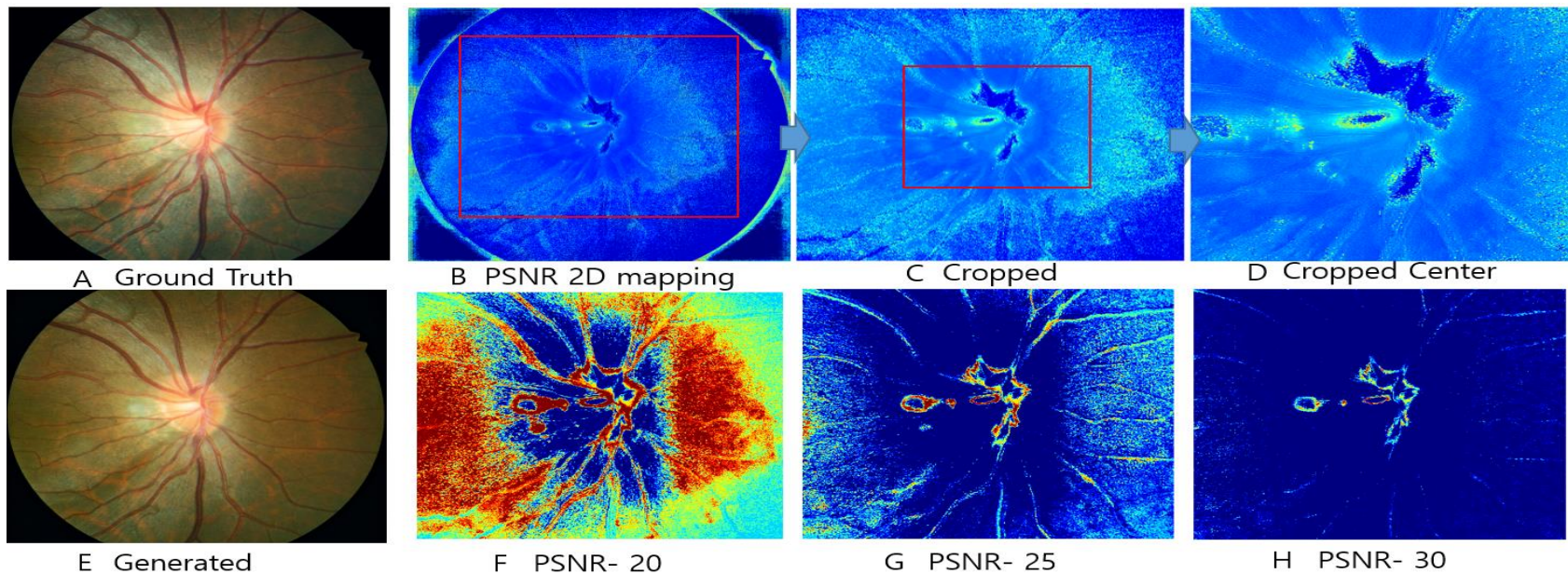
**Table3.2 Comparison of the SSIM index values for the representative test images.**

SSIM, structural similarity; SRRF, Super-Resolution Forests; NBSRF, Naive Bayes Super-Resolution Forest; SRFBN, Feedback Network for Image Super-Resolution; SRResNet, Super-Resolution Residual Network; SR-GAN, Super-Resolution Generative Adversarial Network

<b>PSNR (dB)</b>	<b># 13</b>	<b># 15</b>	<b># 16</b>	<b># 14</b>	<b># 25</b>	<b># 26</b>	<b>mean</b>
Bicubic	36.17	36.13	38.06	37.60	36.46	37.47	36.98
SRRF	33.20	32.22	33.49	34.41	34.77	34.97	33.84
NBSRF	38.36	38.72	40.81	39.61	37.70	39.02	39.04
SRFBN	38.72	38.98	41.33	39.99	37.90	39.34	39.38
SRResNet	36.98	37.13	38.72	88.20	36.79	37.80	37.60
Modified SR-GAN	27.95	29.81	28.86	20.05	21.69	21.71	25.01

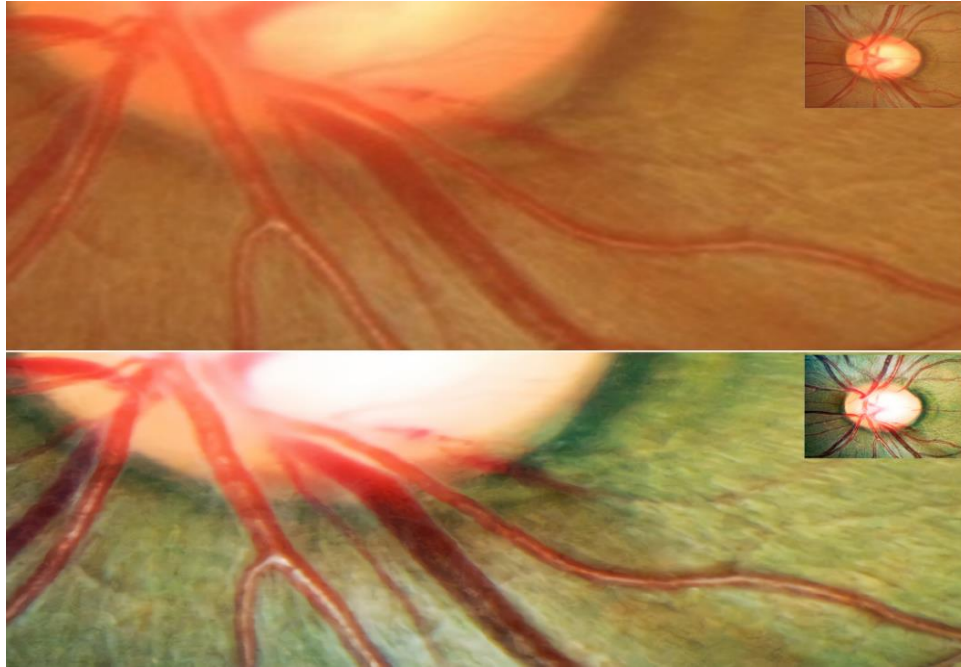
**Table3.3 Comparison of the PSNR values for the representative test image sets.**

PSNR, peak signal-to-noise ratio; SRRF, Super-Resolution Forests; NBSRF, Naive Bayes Super-Resolution Forest; SRFBN, Feedback Network for Image Super-Resolution; SRResNet, Super-Resolution Residual Network; SR-GAN, Super-Resolution Generative Adversarial Network



**Figure 3.9 Image mapped PSNR of generated by SR-GAN.**

The ground truth and generated image (A, E), The PSNR between ground truth and generated image, was mapped as 2D image (B), partially cropped (C, D), thresholding PSNR and interpolated value above 20(F), 25(G), 30(H). The high PSNR region distributed in the central part and around the blood vessels.



**Figure 3.10 Representative optic-disc photography (ODP) of eye with optic disc hemorrhage (DH).** (A) Magnified image of inferotemporal area in original high-resolution ODP, (B) Magnified image of inferotemporal area in deep-learning-based enhanced ODP. The enhanced ODP improved the color and spatial contrast between the DH and the background retinal color.

<b>MOS</b>	<b># 13</b>	<b># 15</b>	<b># 16</b>	<b># 14</b>	<b># 25</b>	<b># 26</b>	<b>mean</b>
Bicubic	2.87	2.75	2.95	3.05	2.84	3.08	2.92
SRRF	2.96	3.12	3.56	3.00	2.88	2.94	3.08
NBSRF	2.98	3.15	3.64	3.25	2.86	2.92	3.13
SRFBN	3.87	4.05	4.12	3.91	3.75	3.80	4.06
SRResNet	3.73	3.96	4.02	3.88	3.74	3.61	3.82
Modified SR-GAN	4.25	4.46	4.52	4.11	4.28	4.33	4.33

**Table3.4 Comparison of the MOS values for the representative test image sets.**

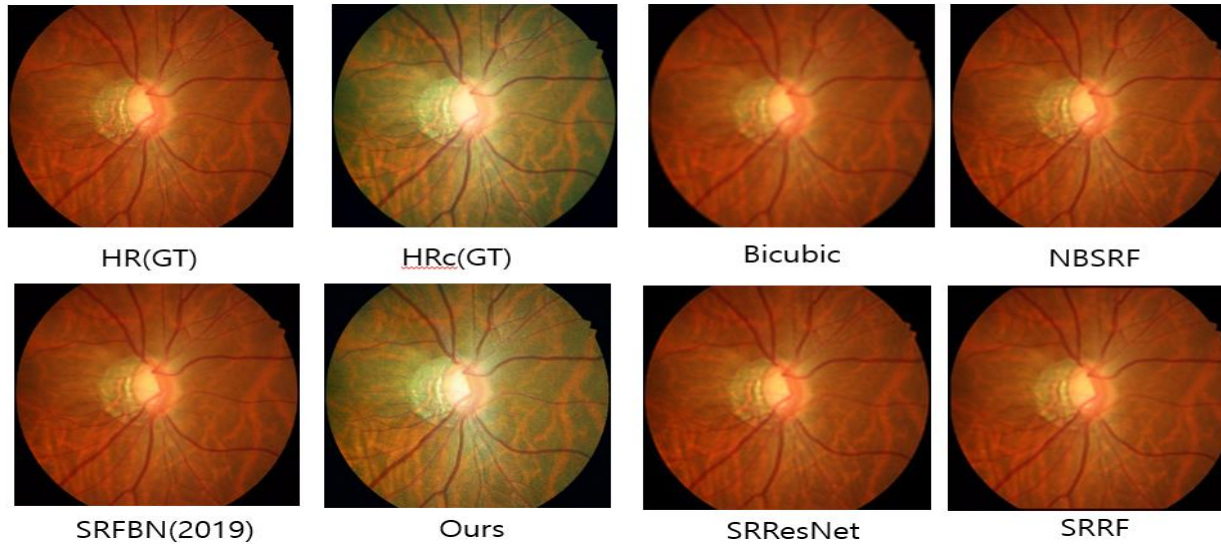
MOS, mean opinion score; SRRF, Super-Resolution Forests; NBSRF, Naive Bayes Super-Resolution Forest; SRFBN, Feedback Network for Image Super-Resolution; SRResNet, Super-Resolution Residual Network; SR-GAN, Super-Resolution Generative Adversarial Network



The Tables 3.2,3 provide validation results for the representative six test image sets (all 384 x 384-pixel size). Both the mean SSIM and PSNR values were lower by our SR-GAN compared with other methods including Bicubic [64], NBSRF (Naive Bayes Super-Resolution Forest)[65], SR-RF (Super-Resolution Forests)[66], SRResNet (Super Resolution Residual Network)[55], and SRFBN (Feedback Network for Image Super-Resolution)[67]. Since our modified SR-GAN was designed to generate images to improve not only the resolution but also the color and spatial contrast, some of the enhanced images had greatly different color composition compared with the reference images (Figure 3.6). By our modified SR-GAN method, the SSIM and PSNR values were higher in images #13, #15, and #16 than in images #14, #25, and #26. In images #13, #15, and #16, the obtained ODPs were similar to the targeted ODPs. In images #14, #25, and #26, however, the overall background color of the ODPs was transformed from red-orange to green. This might have caused the relatively lower SSIM and PSNR values, even though they were perceptually convincing images (Figure 3.6). The MOS test showed significant gains in perceptual quality using our modified SR-GAN compared with the Bicubic, NBSRF, SR-RF, SRResNet, and SRFBN methods (Table 3.4).

The Figure 3.9, represents the PSNR between ground truth and generated image, was mapped as 2D image. It indicates that indirectly check which part of the generated image is considered important. PSNR value was thresholded

with above 20, 25, 30 respectively. The high PSNR region distributed in the central part (neural rim) and around the blood vessels.



**Figure 3.11 Validation Results for Control Group**

SRRF, Super-Resolution Forests; NBSRF, Naive Bayes Super-Resolution Forest; SRFBN, Feedback Network for Image Super-Resolution; SRResNet, Super-Resolution Residual Network; SR-GAN, Super-Resolution Generative Adversarial Network. SRFBN showed best perceptual image quality.

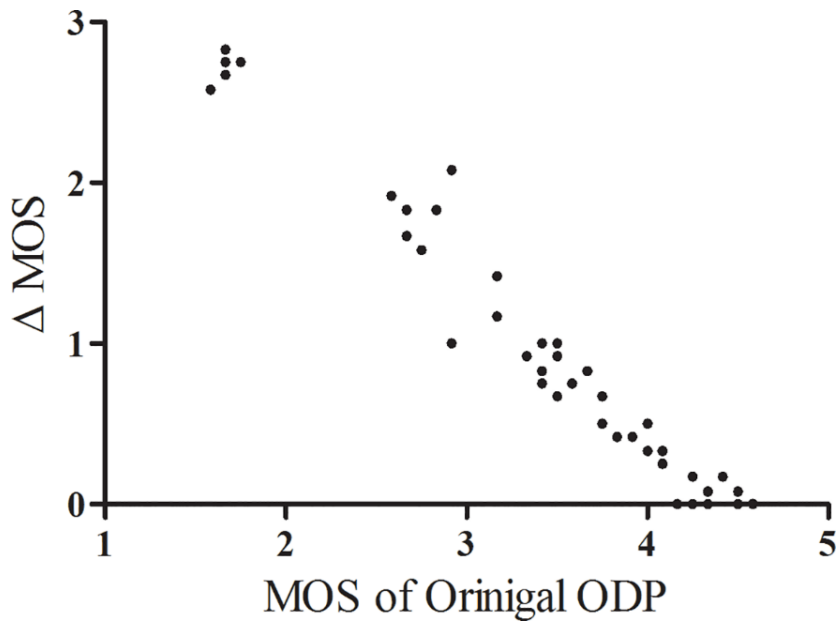
### 3.3.3 Clinical Validation of Enhanced ODP by MOS Comparison

Characteristics	Values
Age (yrs)	57.5 ± 11.9
Male, n (%)	46 (46.9)
Spherical equivalent (D)	-1.4 ± 3.0
IOP (mmHg)	14.7 ± 3.4
CCT (μm)	541.6 ± 32.4
Axial length (mm)	24.5 ± 1.75
VF MD (decibels)	-4.5 ± 8.6

**Table 3.5 Demographic and Clinical Characteristics of Study Subjects (N = 98)**

Values are mean ± standard deviation.

D, diopters; IOP, intraocular pressure; CCT, central corneal thickness; VF, visual field; MD, mean deviation.



**Figure 3.12 Scatter plot of delta mean opinion score ( $\Delta$  MOS) against MOS of original optic-disc photography (ODP).** The  $\Delta$  MOS was calculated as the difference between the original and enhanced ODP scores. Note that the lower the original ODP image quality score, the larger the  $\Delta$  MOS.

A total of 1200 responses comprising 12 ophthalmologists' image quality assessments of 50 original and 50 enhanced ODPs were analyzed. The subjects' demographic and ocular characteristics are provided in Table 3.5. The image quality grades were numbered between 1 and 5 (higher scores indicating better quality), and all of the 50 enhanced ODPs were graded as either 'excellent' or 'good.' The MOS for the enhanced ODPs was significantly higher than that for the original ODPs ( $4.36 \pm 0.38$  vs.  $3.51 \pm 0.88$ ,  $P < 0.001$ ). The lower the

original ODP image quality score, the larger the difference between the original and enhanced ODPs' score (Figure 3.12).

### **3.3.4 Comparison of DH-Detection Accuracy**

The 12 ophthalmologists' assessments of the 50 original ODPs and 50 enhanced ODPs were analyzed. The overall DH-detection accuracy was 76.3% with the original ODPs and 90.7% with the enhanced ODPs ( $P < 0.001$ ). Among the misdiagnosed DHs, the rate of false-positive detection was 6.2 and 2.7% in the original and enhanced ODPs, respectively ( $P = 0.003$ ). The rate of false-negative detection was 17.5 and 6.7% in the original and enhanced ODPs, respectively, and the difference was statistically significant ( $P < 0.001$ ).

The group with the low original image quality (mean score  $< 3.0$ ) showed a much improved DH-detection rate with the enhanced ODPs. The DH-detection accuracy differences between the original and enhanced ODPs were  $29.5 \pm 17.6$  and  $9.0 \pm 12.8\%$  in the low- and high-original-image-quality groups, respectively ( $P < 0.001$ ).

## 3.4 Discussion

### 3.4.1 Research Significance

We have presented herein a novel deep-learning approach to ODP enhancement that is capable of (1) 4-times up-scaling and (2) enhancement of anatomical details by means of contrast, color, and brightness improvement. We found that the resultant enhanced ODPs significantly improved general ophthalmologists' accuracy of DH detection in glaucoma patients. The core novelty of our method lies in its clinical robustness in constructing image datasets. By applying a customized manual image post-processing algorithm to the training dataset, our network could improve both resolution and visibility of anatomical details, which compares favorably with other deep-learning approaches that focus only on resolution enhancement.

Recently, general image enhancement has achieved state-of-the-art performance, especially with the development of deep-learning techniques[68-70]. Dai et al. proposed a two-stage denoising method including fourth-order partial differential equations (PDEs) and a relaxed median filter for retinal image enhancement[71]. Bandara and Giragama applied a spatially adaptive contrast-enhancement technique for enhancement of fundus images[71, 72]. However, some retinal pathologies (e.g., hemorrhages, microaneurysms, and drusen) are mostly only a few pixels wide, causing them to be easily confused

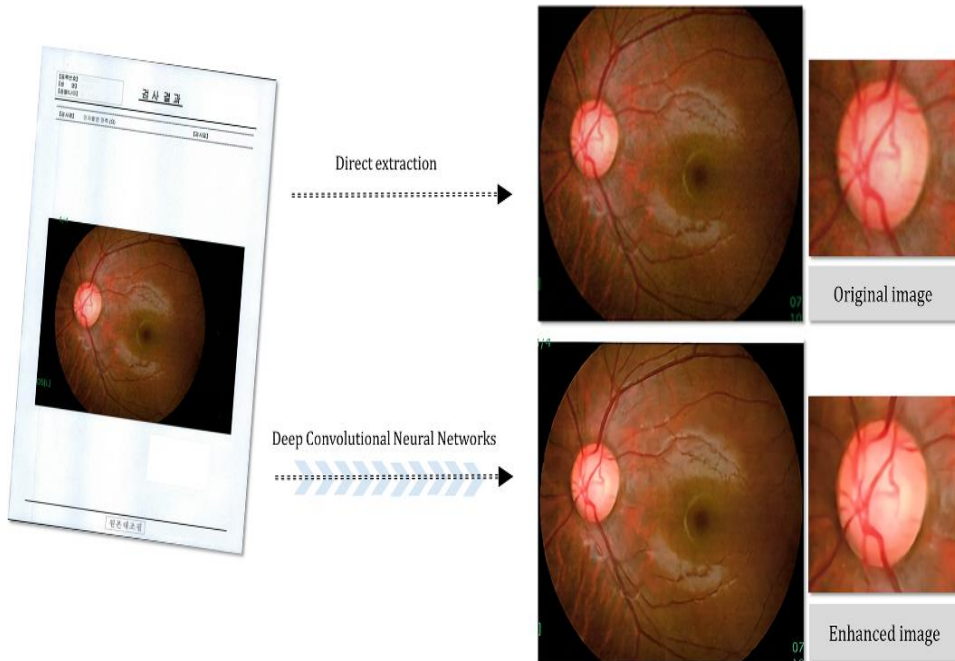
with artifacts of noise. Thus, a fundus image enhancement method must be able to both suppress the undesired low-quality factors and preserve the pathological characteristics simultaneously, which requirement general enhancement techniques cannot satisfy [73]. Zhao et al. applied adversarial loss to blurry retinal images [74]. However, despite its computational efficiency, this method focuses only on generating photo-realistic images, ignoring lesions significant to clinical applications. Thus, in our study, we focused on designing an effective deep-learning model for robust images suitable to the diagnosis of ophthalmic pathologies.

In this study, 74% of the original ODPs were evaluated as ‘better than fair’ image quality sufficient for detection and diagnosis of pathologic change. However, on those original ODPs, a large percentage (142/600, 23.7%) of images was misdiagnosed by the ophthalmologists. This might have been owed to the fact that, even with high-resolution, good-quality ODPs, there is often limited detectability of small and indistinct pathologies due to insufficient time, fatigue, and/or lack of experience [52]. We demonstrated that by use of enhanced ODPs, detection accuracy for ONH pathology can be greatly improved.

We expect that this deep-learning approach for enhanced ODP will see wide application for accurate evaluation of ophthalmic pathologies and precision assessment of disease progression. Enhanced ODPs, for example, are expected to more clearly show vessel alterations or PPA changes associated



with glaucomatous damage. Enhanced imaging enables ophthalmologists to zoom in on a suspect area and examine it in greater detail, without pixel loss. By application of this method to fundus photography, minute retinal hemorrhage or subtle enlargement of retinal nerve fiber layer defect can be more accurately and consistently detected.



**Figure 3.13 Example of fundus photography transferred from another institution as printed document.** (A) Original document, (B) directly extracted fundus image, (C) magnified image of optic disc in original fundus photography, (D) deep-learning-based enhanced fundus photography, (E) magnified image of optic disc in enhanced fundus photography. The enhanced fundus photography improved the structural details of the optic disc, neuroretinal rim margin and vessel contours.

The clinical utility of deep-learning based image enhancement is particularly high in cases of low-quality images having low resolution and/or low contrast. In eyes with cataract or corneal opacity, enhanced ODPs can be used to improve overall image quality and the accuracy of glaucoma and retinal disease diagnostics. Moreover, images showing contrast loss due to poor focus, eye movement or insufficient illumination can be up-scaled. Regarding the examination results printed on paper in low resolution and transferred from other institutions, applying our deep-learning method for enhanced ODPs makes possible not only magnification of such images but also improvement of their structural details, which in turn allows for more meticulous evaluation (Figure 3.13). Additionally, there is growing interest in the value of using telemedicine for detection, following, and treatment of ophthalmic diseases [75, 76]. In cases of tele-ophthalmology requiring transmission of acquired low-resolution results [77, 78], transformation to enhanced images certainly can help to overcome hardware limitations, thereby enabling ophthalmologists to more closely analyze suspicious regions.

It has been reported that using GAN for super-resolution imaging can incur image artifacts in fine details[55]. Thus, two glaucoma specialists (AH and YKK) checked each image for the presence of artifacts, and could confirm that there were no such cases in our dataset. The underlying reason for this difference in artifact occurrence rate is not yet clear. We speculate that previous

GANs are more vulnerable to artifacts because they utilize a variety of images as a training dataset, as compared with ours, which consists only of ODPs. However, since artifacts in medical applications may affect diagnosis or management of patients, caution needs to be exercised in any attempts to utilize our network for other image types.

Single-image SR via deep learning recently has attracted significant research attention. In the present study, a modified SR-GAN consisting of a GAN and a pre-trained VGG19 network was adopted for image training. A network trained for image classification (like VGG) stores, in its feature maps, detailed information on the appearance of common objects, thus enabling an up-scaled image to be made up, to the extent possible, of objects resembling real-world ones. A GAN also has additional merits including non-dependence on prior-knowledge, the lack of any need to design hand-engineered features, and high effectiveness in capturing image structures. Such underlying advantages render SR-GAN a robust platform allowing for multiple applications to be followed once well-trained SR artificial intelligence is established. In future studies, we will explore this modified SR-GAN's results for different datasets.

It is known that deep learning generally requires a large dataset for the training phase [61]. In the current study however, we demonstrated the ability to obtain clinically meaningful results with only 48 pairs of datasets. We carefully modeled both the image degradation process for generation of low-

resolution ODPs and the image customization process for production of compensated high-resolution ODPs; in this way, we eliminated the need for complicated alignment of high- and low-resolution pairs. These steps simplified data processing and improved the modified GAN's robustness. We believe that this example-based method using standardized low- and high-resolution image pairs can maximize the time efficiency of the training process.

### **3.4.2 Limitations**

The present study's findings must be interpreted in light of its limitations. First, numerical evaluation of enhanced image quality was unsuitable for some of our dataset[79]. Different metrics, such as PSNR, SSIM, and multi-scale SSIM, are widely used for quantitative assessment of image restoration quality [80]. These metrics measure reconstructed image quality with respect to the reference or ground-truth image. Some of our enhanced ODPs had greatly different color composition compared with the reference image, due to the fact that we had used compensated ODPs with color-contrast customization in the training process [81]. With such alterations in color composition, direct comparison by numerical evaluation with other deep-learning methods that focus only on resolution improvement would be inappropriate[81]. Furthermore, none of these metrics are known to be well matched with human

visual responses to image quality[61]. For these reasons, we focused on the clinical implications of the use of enhanced ODPs for diagnosis of optic disc pathology. Further numerical evaluation of enhanced ophthalmic image quality with reasonable metrics should be carried out in future studies. Second, we applied a customized image post-processing algorithm for optimization of both the color and spatial contrast of each ODP. This detailed manual adjustment was applied differently according to each ODP's image quality. Although this variability in the image-processing procedure may incur reproducibility issues, we believe that the core novelty of our method lies in the customization of image-processing procedures. In real-world clinical practice, ODPs can have several different limitations other than insufficient resolution, such as low color and spatial contrast. Based on the customized image compensation process to optimize the visibility of ophthalmic pathology, we enabled GAN to generate enhanced ODPs with both higher resolution and improved anatomical details. However, different image post-processing methods or training strategies might manifest different results. Third, the clinical implications of enhanced ODP were not evaluated for other optic-disc characteristics such as neuroretinal rim contours. This was due to the fact that the image-compensation process of the present study was mainly focused on the enhancement of the visibility of the optic disc margin and surrounding retinal vessels, not on the rim contours. Therefore, further research is certainly needed to determine the usefulness of deep-learning-based enhanced ODP in glaucoma diagnostics. Fourth, our

meaningful training results were based on relatively little data. Further studies will validate this algorithm using a larger dataset.

## 3.5 Conclusion

The current study demonstrated that deep learning can be applied to create an algorithm that is capable of producing enhanced ophthalmic images that are 4-times up-scaled and improved in their structural details. The enhanced ODPs thereby obtained significantly increased the detection accuracy of optic disc pathology. Further studies exploring the usefulness of this algorithm's deployment in different clinical settings are warranted.

\* Large sections of this chapter were published previously in *PloS one*, (Ha, A., Sun, S., Kim, Y. K., Lee, J., Jeoung, J. W., Kim, H. C., & Park, K. H. (2020). Deep-learning-based enhanced optic-disc photography. *PloS one*, 15(10), e0239913.)

## **Chapter 4**

### **Deep Learning Based Prediction of Glaucoma**

### **Onset Using Retinal Image and Patient Data**

#### **4.1 Introduction**

##### **4.1.1 Background**

In treating glaucoma, timely detection of disease progression is crucial to preserve patients' vision. Structural and functional clinical modalities such as optical coherence tomography (OCT) and visual field (VF) test are utilized for monitoring glaucoma progression[82, 83]. However, due to the variability of these modalities, it can take in many cases several years for confirmation of disease progression. To mitigate this problem, guided progression analysis (GPA) is performed with event-based analysis or trend-based analysis. GPA has been available for VF assessment as well as for OCT scanning of the macular ganglion cell-inner plexiform layer (mGCIPL) or retinal nerve fiber layer (RNFL). However, it helps clinicians only to monitor disease progression, not to predict progression.



Convolutional auto encoder(CAE)[84] is a type of unsupervised learning method trained without a label. CAE has been used to reconstruct image or removing noise of image. For other purposes, CAE can extract the features of images. The decision tree based ensemble[85] has recently become popular among data scientists due to high efficiency and effectiveness of solving classification and regression problem. XGboost[86] and Random forests[87] are one of the most powerful decision tree based ensemble method.

#### **4.1.2 Related Work**

No study has been undertaken to investigate whether combining different types of clinical information would be helpful in predicting the subsequent course of glaucoma in individual patients.

## **4.2 Methods**

### **4.2.1 Study Design**

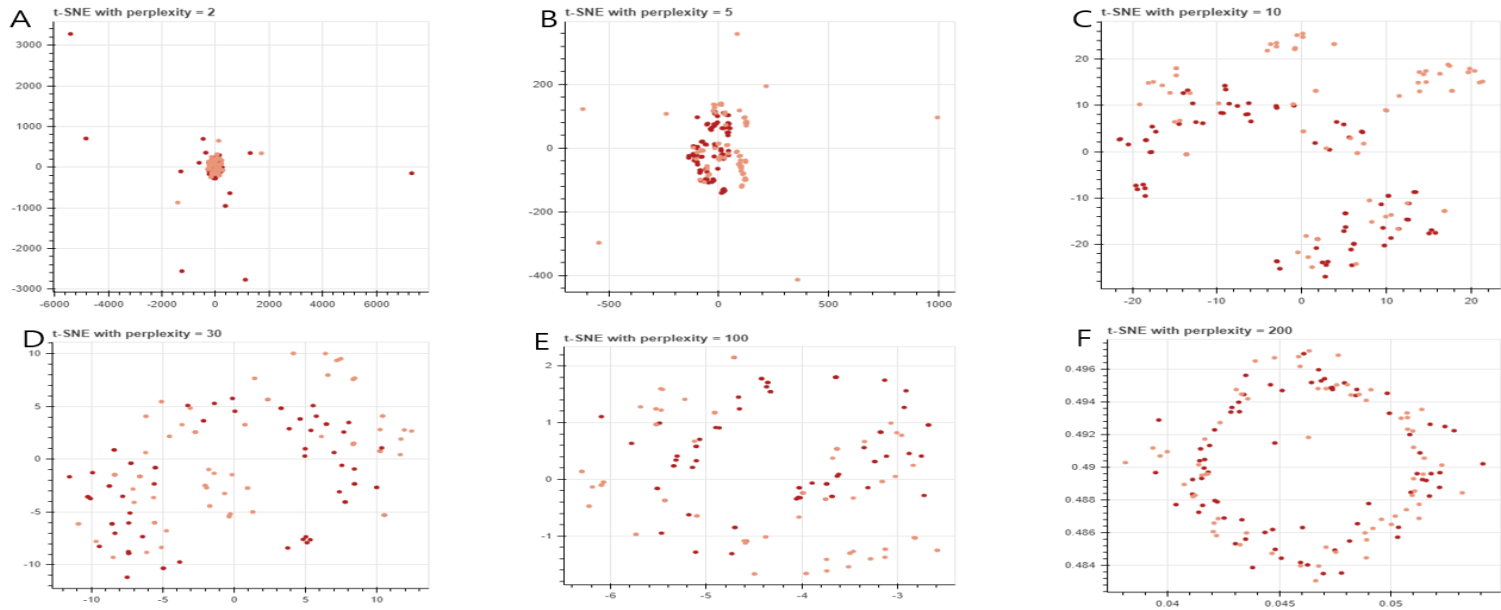
The purpose of this study was to develop a machine-learning model using clinical data for prediction of subsequent glaucoma progression in glaucoma suspects.

#### 4.2.2 Dataset

Sex	Onset-year	Age	Laterality	Family History	Diabetes Mellitus	CCT	RNFL Thickness
M=121	1~7 = 15	Mean =	Right = 104	O =36	O = 27	Mean =	Mean = 90.99
F=99	Normal = 105	55.848	Left = 106	X = 174	X=183	537.04	Std =3.83
		Std = 9.45				Std =29.71	
IOP	AXL	SE	DBP	SBP	Height	Weight	BMI
Mean =	Mean =	Mean = -1.05	Mean =	Mean=116.69	Mean = 1.63	Mean = 63.02	Mean =
14.79	23.90	Std=1.59	69.18	Std = 12.97	Std = 0.09	Std=9.32	23.612
Std =2.89	Std =1.23		Std = 6.18				Std = 3.18

**Table4.1 Statistical characteristic of 210 total patient**

CCT (Central Corneal Thickness), IOP (Intraocular pressure), AXL (axial length), SE (Spherical Equivalent), DBP (Diastolic Blood Pressure), SBP (Systolic Blood Pressure)



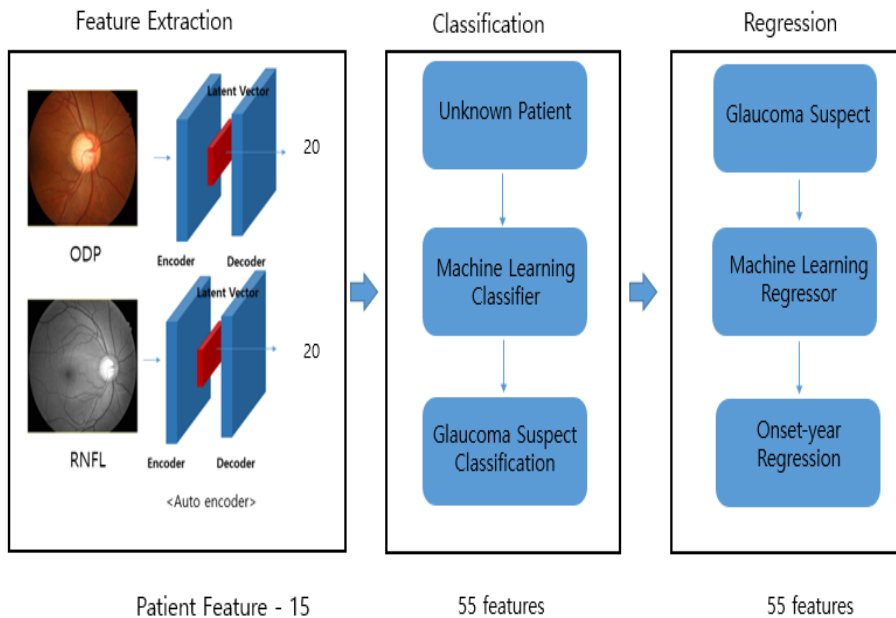
**Figure4.1 t-SNE for Dataset Embedding**

The distance between embedded data was expressed as stochastic probability depends on perplexity, perplexity 2, 5, 10, 30, 100, 200(A, B, C, D, E,F)

Eyes were chosen from subjects examined for glaucoma at the Glaucoma Clinic, Seoul National University Hospital. Retinal photos were originally 1872 x 901-pixel and resized and saved as 448x 448. Red-free retinal nerve fiber layer (RNFL) images and optic-disc photographs (ODPs) were obtained for all the patients. The patient data included the onset of glaucoma and the onset year. There are 15 features such as sex, age, glaucoma laterality, glaucoma family history, diabetes mellitus, central corneal thickness (CCT), intraocular pressure (IOP), axial length (AXL), spherical equivalent (SE), diastolic blood pressure (DBP), and systolic blood pressure (SBP). All the patients were followed up for up to seven years for the occurrence of glaucoma; those who did not develop glaucoma after seven years were classified as normal groups. The demographic features of the subjects are provided in Table4.1

Data visualization was performed with t-SNE[88] with 55 features of the data and glaucoma onset year, perplexity of 2, 5, 10, 30, 100, 200

### 4.2.3 Design of Overall System



**Figure4.2 Overall systems of designed classification and regression network**

Features from ODP and RNFL image was extracted by convolutional auto encoder(left), extracted features were fed into machine learning classifier for classifying glaucoma suspect(middle) and for prediction glaucoma onset-year(right)

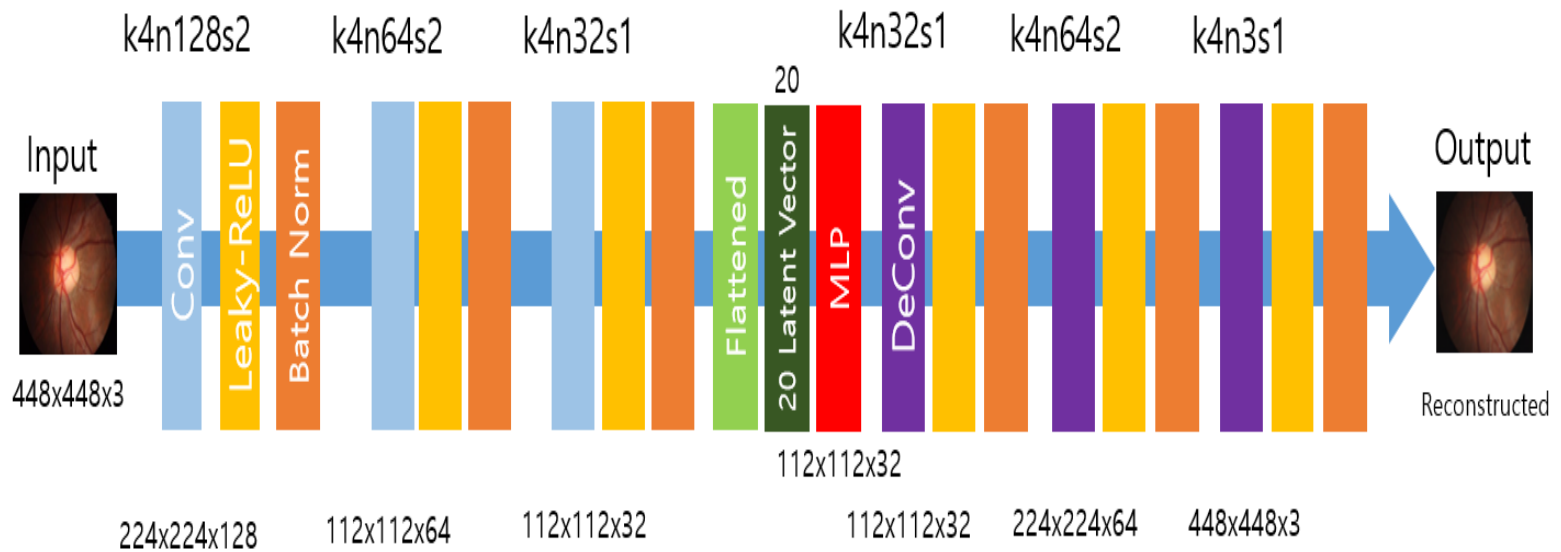
The overall system consisted of three steps as shown in Figure 4.2. In the first step, the features of the two images (ODP and RNFL) were extracted with CAE, and in the next step, the extracted features and demographic features of the patient were used to classify whether the patient was suspected of glaucoma. In the final stage, the onset year is predicted for the patients with suspected glaucoma using a regressor.

#### 4.2.4 Design of Convolutional Auto Encoder(CAE)

In general, CAE is used to restore an image entered as an input. CAE is a well-known deep-learning unsupervised feature extractor. Because of the lack of data to approach the classic deep-learning method, unsupervised learning method was adopted. The ODPs and RNFL images were fed into the CAE program for feature extraction. Mean squared error (MSE) was used for the loss function, and the CAE program was trained to minimize the MSE between the input and output images.

$$L_{MSE} = \frac{1}{r^2WH} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (I_{x,y}^{Original} - I_{x,y}^{Reconstructed})^2$$

As an input of CAE, the original ODPs and RNFL images were resized to  $448 \times 448 \times 3$ . By the process of convolution (filter sizes: 4, Strides: 2), the size of the feature map was reduced by 1/2. After 3-stacked convolution layer, the size of the feature map was reduced to  $112 \times 112 \times 3$ . The feature map was resized to 20 nodes using a multilayer perceptron (MLP) with 20 latent vectors as features. During three deconvolutions, 20 latent vectors were reconstructed to the size of the original image ( $448 \times 448 \times 3$ ). Leaky-ReLU was adopted for the activation function, and batch norm was adopted with a batch size of eight.



**Figure4.3 Architecture of CAE network.**

The corresponding kernel size (k), number of feature maps (n) and stride (s) are indicated for each convolutional layer.  $448 \times 448 \times 3$  reduced to 20 vector after convolution (sky blue) and reconstructed to  $448 \times 448 \times 3$  after deconvolution(purple).

#### **4.2.5 Glaucoma Suspect Classification**

40 features for the two input images (ODP and RNFL) were extracted from the latent vector of CAE, and 15 features were extracted from the patient information. A total of 55 features were used to classify the glaucoma suspects. Data from 210 patients (Training: 140; normal-67, glaucoma-73; Test: 70; normal-38, glaucoma-32) were used.

Classification was performed using Random Forest, XGboost, AdaBoost, Gradient Boosting, and SVM(RBF) classifiers with four cases: both images and demographic features (55 features), ODP and demographic features (35 features), RNFL with demographic features (35 features), and only demographic features (15 features). A five-fold cross validation with random and grid search was used for Hyperparameter optimization. To calculate the confidence interval, bootstrapping method was used.

#### **4.2.6 Glaucoma Onset-Year Prediction**

The features for the two input images (ODP and RNFL) were extracted from the latent vector of CAE, and 15 features were extracted from the patient information. A total of 55 features were used to classify the prediction for the glaucoma onset year. (Training 73 (Glaucoma), Test 34 (Normal-2, Glaucoma-32)) was used. At the glaucoma suspect classification stage, two normal patients were classified as glaucoma suspects.



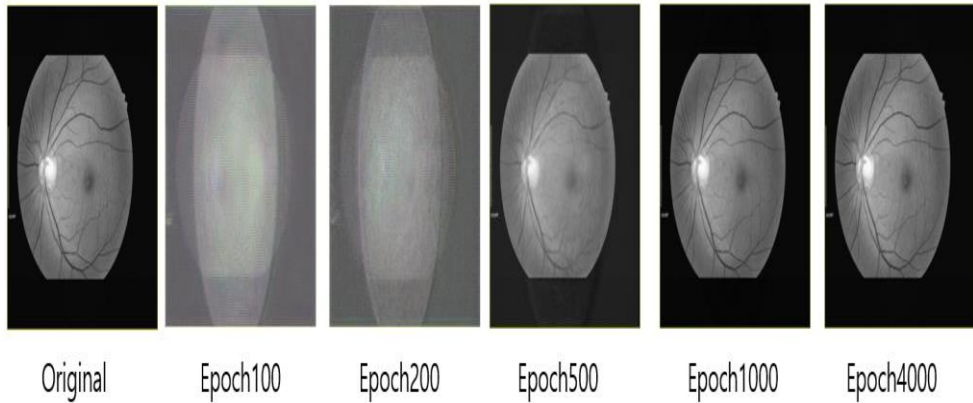
Regression was performed using Random Forest, XGboost, AdaBoost, Gradient Boosting, and SVM(RBF) classifiers with four cases: both images and demographic features (55 features), ODP and demographic features (35 features), RNFL with demographic features (35 features), and only demographic features (15 features). A five-fold cross validation with random and grid search was used for Hyperparameter optimization. The Hyperparameter was optimized using root mean squared log error (RMSLE). To calculate the confidence interval, bootstrapping method was used.

$$\text{RMSLE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(p_i + 1) - \log(a_i + 1))^2}$$

(p: prediction, a: ground truth)

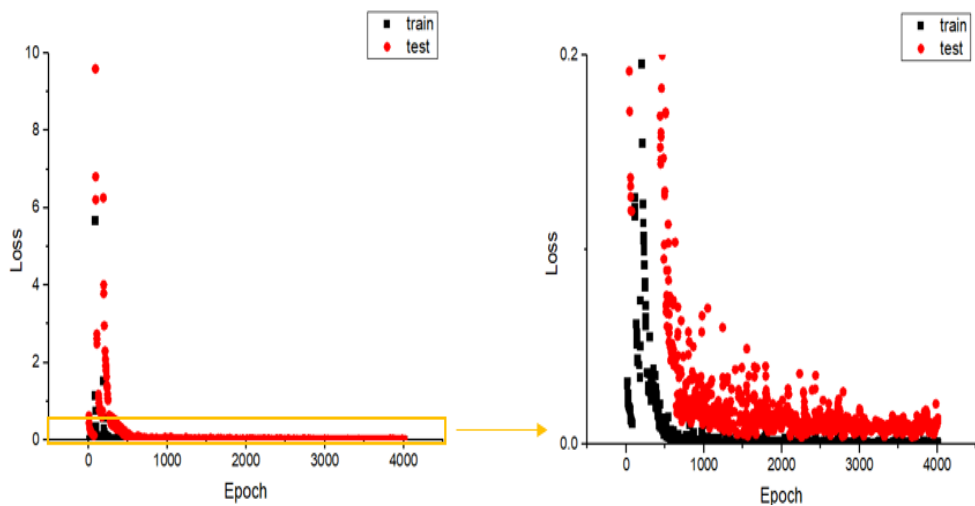
## 4.3 Result

### 4.3.1 Performance of Designed CAE



**Figure4.4 Reconstructed image while training CAE**

As epoch repeated, sharper images are obtained, from epoch 1000 or higher, model reconstruct visually similar to the original image



**Figure4.5 Training loss curve for CAE**

Training MSE loss (black dot) of model converges after 500 epochs, the test loss (red dot) is also converging but unstable

Figure 4.4 shows the loss of the designed CAE training set epochs. From the epoch 1000 started to restore the original image well, the loss graph dropped low and was saturated after epoch 2000 both in training and in test. Mean squared error was used as the loss function of CAE. Here,  $I(i, j)$ ,  $k(i, j)$  represent the original image and the target image, respectively. CAE was trained until 4000 epochs with Adam optimizer[89] – learning rate of  $10e-4$ , beta1 of 0.9.

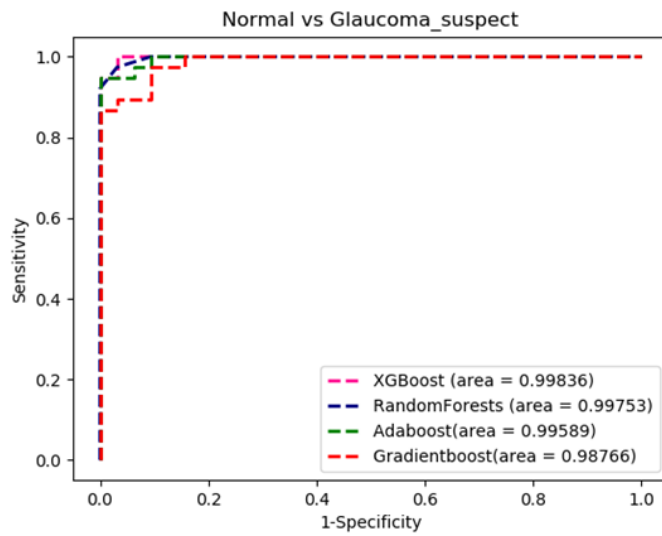
### 4.3.2 Performance of Designed Glaucoma Suspect Classification

	Random Forest (both image & Patient Feature)	XGBOOST (both image & Patient Feature)	AdaBoost (both image & Patient Feature)	Gradient Boosting (Patient Feature)
<b>Training Accuracy(%)</b>	96.429	94.286	95	93.57
<b>Test Accuracy(%)</b>	95.801(84.607-95.714)	97.14(88.107-98.571)	97.14(87.5-1)	95.71(86.036-97.143)
<b>AUC</b>	0.998(0.967-0.998)	0.998(0.973-1.)	0.996(0.970-0.998)	0.988(0.956-0.997)
<b>Sensitivity</b>	0.938(0.875-1)	1(0.906-1.)	1(0.906-1.)	1(1-1)
<b>Specificity</b>	0.974(0.716-974.)	0.947.(0.868-0.994)	0.947.(0.822-1.)	0.579(0.348-0.579.)

**Table4.2 Glaucoma suspect classification result**

	<b>IOP</b>	<b>RNFL Thickness</b>	<b>D_7</b>	<b>D_14</b>	<b>R_2</b>	<b>Age</b>
<b>Classification Importance (Normalized)</b>	0.123	0.1	0.077	0.077	0.077	0.077

**Table4.3 Feature importance of glaucoma suspect classification networks**



**Figure4.6 ROC curve for glaucoma suspect classification networks**

XGboost(pink dash), Random Forests(blue dash), Adaboost(green dash),  
Gradient boost(orange dash)

The accuracy and AUC calculated for designed network are shown in Table 4.2 Which is also indicated the results obtained by machine learning algorithms. In addition, the accuracy and AUC of XGBOOST with both red-free RNFL and ODP is highest. The accuracy of training phase was 94.286 and in test phase accuracy was 97.14 with (95% CI 88.107-98.571), AUC 0.998(0.973-1.)

, Sensitivity 1(0.906-1.), 0.947. (0.868-0.994). With Random forest training accuracy was 96.429, Test accuracy was 95.801(84.607-95.714), AUC 0.998(0.967-0.998) Sensitivity 0.938(0.875-1), Specificity 0.974(0.716-974.). With Adaboost training accuracy was 95, test accuracy was 97.14(87.5-1), 0.996(0.970-0.998), Sensitivity 1(0.906-1.), Specificity, 0.947. (0.822-1.). Gradient Boosting algorithm performed best when using only patient feature in training phase the accuracy was 93.57 test accuracy 95.71(86.036-97.143) AUC 0.988(0.956-0.997) Sensitivity 1(1-1). Specificity 0.579(0.348-0.579.)

The feature importance of best model XGBOOST was extracted by Scikit-learn library. IOP, RNFL Thickness, ODP feature 7, 14, red-free RNFL feature 2 and age in order.

### 4.3.3 Performance of Designed Glaucoma Onset-Year Prediction

<b>MSE</b>	<b>XGboost (both image)</b>	<b>Random Forest (Disc photo)</b>	<b>Adaboost (W/O image)</b>	<b>Gradient Boost (W/O image)</b>
<b>Train</b>	1.08	1.33	1.79	1.53
<b>Test</b>	2.613 (2.316- 3.880)	3.0588(2.4647- 5.822)	3.353(2.635- 5.660)	3.353(2.818-4.398)

**Table4.4 Glaucoma onset-year regression result.** MSE was calculated between onset-year and ground truth

The MSE values calculated for the designed network are listed in Table4.4. This table also presents the results obtained by the machine learning algorithms. The MSE of XGBOOST was the lowest with both red-free RNFL and ODP.

	<b>Age</b>	<b>BMI</b>	<b>DBP</b>	<b>R_2</b>	<b>D_3</b>	<b>CCT</b>
<b>Regression</b>	0.085	0.064	0.049	0.047	0.047	0.04
<b>Importance (Normalized)</b>						

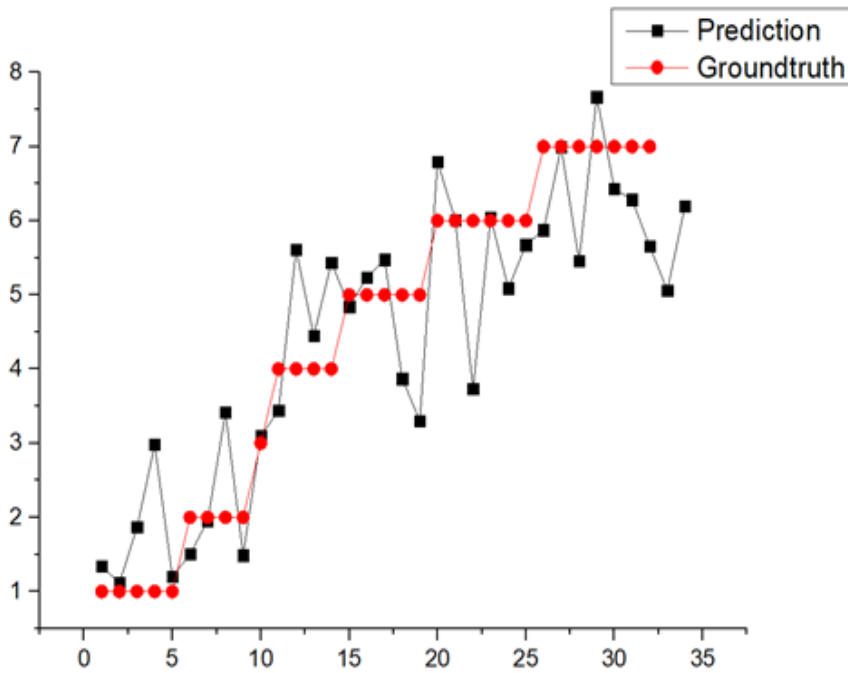
**Table4.5 Feature importance of glaucoma suspect classification networks**



	RNFL Thickness	IOP	Age	DBP	R_8	Family History
Linear correlation	0.69	-0.68	0.66	0.43	0.42	-0.38

**Table4.6 Linear correlation glaucoma onset-year vs patient features**

The important features of the designed network are shown in Table4.5 The lowest MSE (XGBOOST) values were observed for age, BMI, DBP, red-free RNFL feature 2, ODP feature 3, and CCT. The glaucoma onset year and patient features had a high linear correlation with RNFL thickness, IOP, age, DBP, RNFL feature 8, and family history.

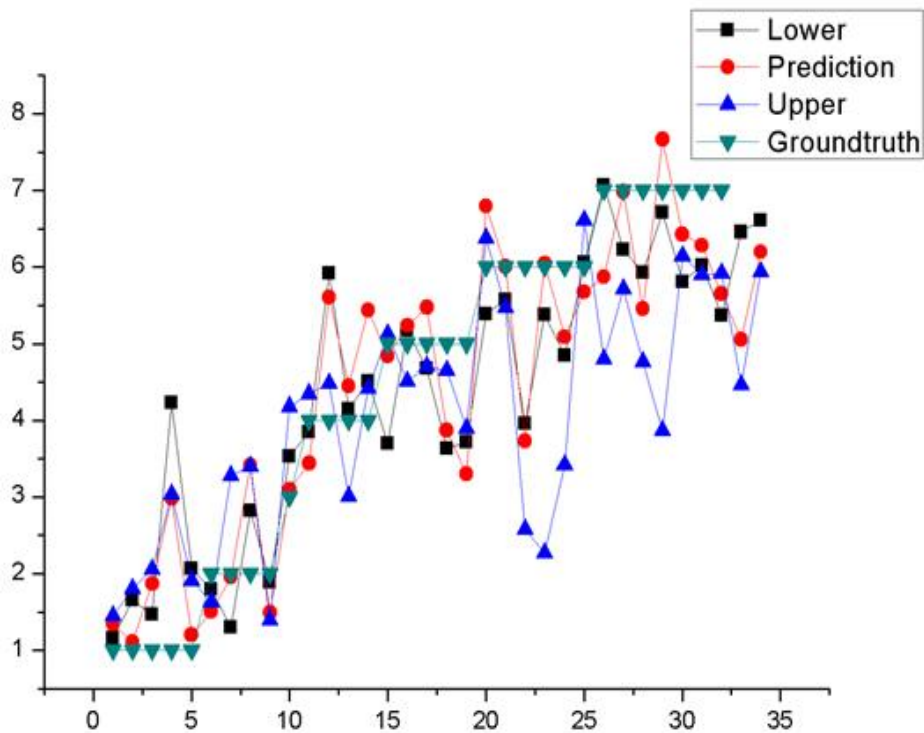


**Figure4.7 Regression result for glaucoma onset-year prediction**

x-axis=patient number, y-axis= onset-year, black line represents predicted onset year value of best model (XGboost with Red-free RNFL, ODP, patient feature), red line represents ground truth value of onset-year (follow up data)

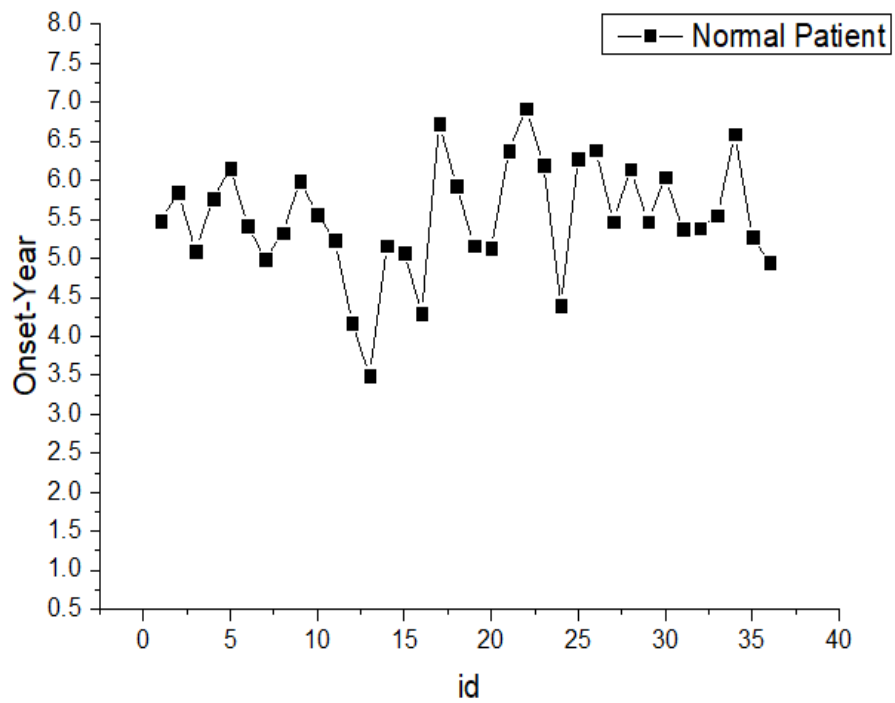
Figure.4.7 shows the onset year as per the actual prediction. Except for a few patients, the trend is consistent. Because of the misclassified patients in classification system, regression was done with the normal group (two normal patients). However, the regression value was predicted high (above five years) and classifier was designed to high false positive. The 95% confidence interval was calculated by bootstrapping method with 50 iterations. Figure 4.8 shows

the upper and lower limits of the 95% confidence interval. The Figure 4.9 shows the regression result for 38 normal patients. Except in the case of a few patients, the onset years of normal patients were predicted to be higher than five years, and the classifier was designed to high false positive.



**Figure4.8 95% Confidence interval for regression result (each individual patient)**

x-axis – patient number, y-axis – onset year, actual prediction value of best model(red line), ground truth(green line), 95% CI of lower prediction value(black line) , 95% CI of upper prediction value(blue line)



**Figure4.9 Onset-year prediction for normal patient**

x-axis – patient number, y-axis – onset year, predicted on-set year value when features of normal patient fed into final model

## **4.1 Discussion**

### **4.4.1 Research Significance**

In the present study, we observed that the designed system with combined use of the features of ODP, red-free RNFL extracted by CAE, and patient features was effective in classifying glaucoma suspects and prediction of glaucoma onset year. The combined use of both image features and patient features showed better performance in glaucoma suspect classification and prediction of glaucoma onset year than the use of image features and patient features individually.

### **4.4.2 Limitations**

Further study is necessary to improve the accuracy of prediction of glaucoma onset year and to validate the results in a larger sample. The availability of such data set is scarce, and it takes a long time to collect the data. Additional validation is needed if there is no ground truth, and hence it is unclear whether a patient will develop glaucoma or not. It is also unclear how the image features extracted in the unsupervised learning method will affect the classification and regression.

## 4.5 Conclusion

We expect that this machine learning approach for the prediction of glaucoma onset year will lead to a wide application for the prediction of glaucoma in very early glaucoma patients. It is possible to establish a follow-up period and provide evidence for starting medication.

As it is possible to have a relatively accurate prediction of glaucoma onset year using only simple patient information, it could be helpful while using it in underdeveloped areas.

# Chapter 5 Summary and Future Works

## 5.1 Thesis Summary

The goal of this study was to support glaucoma diagnosis using deep learning on clinical data and to validate its performance. The performance of the designed network was validated with clinical and engineering parameters.

In the first study, we developed deep-learning algorithms analyzed by DICNN with combined RNFL and GCIPL features on thickness maps; they showed better diagnostic performance for early stage glaucoma than the conventional deep-learning methods. The model was evaluated for accuracy, sensitivity, specificity, and ROC.

In the second study, we developed a deep-learning approach for increased resolution and improved legibility of ODPs by a contrast, color, and brightness compensation model. The model was evaluated using PSNR, SSIM, and MOS. The enhanced ODPs significantly increased the detection accuracy of optic disc pathology.

In the third study, we developed a deep-learning model using clinical data and retinal images (ODP & Red-free Fundus RNFL images) for classification of glaucoma suspects and prediction of subsequent glaucoma onset year in glaucoma suspects. The use of combined image features and patient features showed better performance in glaucoma suspect classification and prediction

of glaucoma onset year than the use of image features and patient features individually.

In conclusion, this study proved that deep learning on clinical data can be effective in improving glaucoma diagnosis, and also showed the possibility of replacing previous models or proposing new methods for glaucoma diagnosis by presenting new concepts and verifying the performance.

## **5.2 Limitations and Future works**

The application of the proposed method for glaucoma diagnosis in clinical practice has several limitations. In the first study, the Grad-CAM works inconsistently, and it remains unclear how the deep-learning model discriminates the cases as either glaucoma or normal. Additional validation is needed to confirm the actual performance. For generalization, it is necessary to apply it to OCT images of other hospitals. In the second study, MOS was measured solely on hemorrhage ODP and normal images, and not on peripapillary atrophy (PPA) or vessel alteration images. The Image compensation process in the present study was focused mainly on enhancement of the visibility of the optic disc margin and surrounding retinal vessels, and not on the rim contours. It is necessary to collect additional data with PPA and vessel alteration. In the third study, further work is necessary to improve the accuracy of prediction of glaucoma onset year and to validate the results on a



larger sample. The availability of such data set is scarce, and it takes a long time to collect the data. Additional validation is needed, but there is no ground truth, and hence it is unclear whether a patient will develop glaucoma or not. Therefore, additional follow-up data are needed for further study.

## Bibliography

- [1] H. A. Quigley, "Neuronal death in glaucoma," *Progress in retinal and eye research*, vol. 18, no. 1, pp. 39–57, 1999.
- [2] H. A. Quigley, J. Katz, R. J. Derick, D. Gilbert, and A. Sommer, "An evaluation of optic disc and nerve fiber layer examinations in monitoring progression of early glaucoma damage," *Ophthalmology*, vol. 99, no. 1, pp. 19–28, 1992.
- [3] R. S. Anderson, "The psychophysics of glaucoma: improving the structure/function relationship," *Progress in retinal and eye research*, vol. 25, no. 1, pp. 79–97, 2006.
- [4] F. Ekici *et al.*, "Relationships between measures of the ability to perform vision-related activities, vision-related quality of life, and clinical findings in patients with glaucoma," *JAMA ophthalmology*, vol. 133, no. 12, pp. 1377–1385, 2015.
- [5] 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*, vol. 121, no. 11, pp. 2081–2090, 2014.
- [6] K. E. Kim, J. M. Kim, J. E. Song, C. Kee, J. C. Han, and S. H. Hyun, "Development and Validation of a Deep Learning System for Diagnosing Glaucoma Using Optical Coherence Tomography," *Journal of Clinical Medicine*, vol. 9, no. 7, p. 2167, 2020.
- [7] D. Mahapatra and B. Bozorgtabar, "Progressive generative adversarial networks for medical image super resolution," *arXiv preprint arXiv:1902.02144*, 2019.

- [8] S. J. Kim, K. J. Cho, and S. Oh, "Development of machine learning models for diagnosis of glaucoma," *PLoS One*, vol. 12, no. 5, p. e0177726, 2017.
- [9] A. J. Tatham and F. A. Medeiros, "Detecting structural progression in glaucoma with optical coherence tomography," *Ophthalmology*, vol. 124, no. 12, pp. S57–S65, 2017.
- [10] J.-C. Mwanza, J. D. Oakley, D. L. Budenz, D. R. Anderson, and C. O. C. T. N. D. S. Group, "Ability of cirrus HD-OCT optic nerve head parameters to discriminate normal from glaucomatous eyes," *Ophthalmology*, vol. 118, no. 2, pp. 241–248. e1, 2011.
- [11] K. R. Sung, J. S. Kim, G. Wollstein, L. Folio, M. S. Kook, and J. S. Schuman, "Imaging of the retinal nerve fibre layer with spectral domain optical coherence tomography for glaucoma diagnosis," *British Journal of Ophthalmology*, vol. 95, no. 7, pp. 909–914, 2011.
- [12] C. A. Curcio and K. A. Allen, "Topography of ganglion cells in human retina," *Journal of comparative Neurology*, vol. 300, no. 1, pp. 5–25, 1990.
- [13] Y. K. Kim, A. Ha, K. I. Na, H. J. Kim, J. W. Jeoung, and K. H. Park, "Temporal relation between macular ganglion cell-inner plexiform layer loss and peripapillary retinal nerve fiber layer loss in glaucoma," *Ophthalmology*, vol. 124, no. 7, pp. 1056–1064, 2017.
- [14] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural computation*, vol. 29, no. 9, pp. 2352–2449, 2017.
- [15] H. Li *et al.*, "Dual-Input Neural Network Integrating Feature Extraction and Deep Learning for Coronary Artery Disease Detection Using Electrocardiogram and

- Phonocardiogram," *IEEE Access*, vol. 7, pp. 146457–146469, 2019.
- [16] J. W. Choi *et al.*, "Using a Dual-Input Convolutional Neural Network for Automated Detection of Pediatric Supracondylar Fracture on Conventional Radiography," *Investigative radiology*, vol. 55, no. 2, pp. 101–110, 2020.
  - [17] G. Wollstein, D. F. Garway-Heath, D. Poinoosawmy, and R. A. Hitchings, "Glaucomatous optic disc changes in the contralateral eye of unilateral normal pressure glaucoma patients," *Ophthalmology*, vol. 107, no. 12, pp. 2267–2271, 2000.
  - [18] Z. Burgansky-Eliash *et al.*, "Optical coherence tomography machine learning classifiers for glaucoma detection: a preliminary study," *Investigative ophthalmology & visual science*, vol. 46, no. 11, pp. 4147–4152, 2005.
  - [19] M.-L. Huang and H.-Y. Chen, "Development and comparison of automated classifiers for glaucoma diagnosis using Stratus optical coherence tomography," *Investigative ophthalmology & visual science*, vol. 46, no. 11, pp. 4121–4129, 2005.
  - [20] 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.
  - [21] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.
  - [22] T. Garipov, D. Podoprikin, A. Novikov, and D. Vetrov, "Ultimate tensorization: compressing convolutional and fc layers alike," *arXiv preprint arXiv:1611.03214*, 2016.

- [23] L. Zhu and N. Laptev, "Deep and confident prediction for time series at uber," in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2017: IEEE, pp. 103–110.
- [24] K. Woods and K. W. Bowyer, "Generating ROC curves for artificial neural networks," *IEEE Transactions on medical imaging*, vol. 16, no. 3, pp. 329–337, 1997.
- [25] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.
- [26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [27] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2017, vol. 31, no. 1.
- [28] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [29] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.
- [30] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

- [31] F. Fazayeli, "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization," 2014.
- [32] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.
- [33] F. A. Medeiros, L. M. Zangwill, C. Bowd, R. M. Vessani, R. Susanna Jr, and R. N. Weinreb, "Evaluation of retinal nerve fiber layer, optic nerve head, and macular thickness measurements for glaucoma detection using optical coherence tomography," *American journal of ophthalmology*, vol. 139, no. 1, pp. 44–55, 2005.
- [34] H. Muhammad *et al.*, "Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects," *Journal of glaucoma*, vol. 26, no. 12, p. 1086, 2017.
- [35] D. L. Wang *et al.*, "Central glaucomatous damage of the macula can be overlooked by conventional OCT retinal nerve fiber layer thickness analyses," *Translational vision science & technology*, vol. 4, no. 6, pp. 4–4, 2015.
- [36] H. N. Marshall *et al.*, "Macular ganglion cell-inner plexiform layer loss precedes peripapillary retinal nerve fiber layer loss in glaucoma with lower intraocular pressure," *Ophthalmology*, vol. 126, no. 8, pp. 1119–1130, 2019.
- [37] M. J. Kim, J. W. Jeoung, K. H. Park, Y. J. Choi, and D. M. Kim, "Topographic profiles of retinal nerve fiber layer defects affect the diagnostic performance of macular scans in preperimetric glaucoma,"

- Investigative Ophthalmology & Visual Science*, vol. 55, no. 4, pp. 2079–2087, 2014.
- [38] A. Ha and K. H. Park, "Optical coherence tomography for the diagnosis and monitoring of glaucoma," *The Asia-Pacific Journal of Ophthalmology*, vol. 8, no. 2, pp. 135–145, 2019.
  - [39] S. Koppers, C. Haarburger, and D. Merhof, "Diffusion MRI signal augmentation: from single shell to multi shell with deep learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016: Springer, pp. 61–70.
  - [40] J. B. Jonas, M. C. Fernández, and J. Stürmer, "Pattern of glaucomatous neuroretinal rim loss," *Ophthalmology*, vol. 100, no. 1, pp. 63–68, 1993.
  - [41] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, 2019.
  - [42] A. Schindler, T. Lidy, and A. Rauber, "Comparing Shallow versus Deep Neural Network Architectures for Automatic Music Genre Classification," in *FMT*, 2016, pp. 17–21.
  - [43] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *ICML deep learning workshop*, 2015, vol. 2: Lille.
  - [44] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.
  - [45] L. Li, M. Xu, X. Wang, L. Jiang, and H. Liu, "Attention Based Glaucoma Detection: A Large-scale Database and CNN Model," in *Proceedings of the IEEE*

*Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10571–10580.

- [46] A. A. Jammal *et al.*, "Detecting Retinal Nerve Fibre Layer Segmentation Errors on Spectral Domain–Optical Coherence Tomography with a Deep Learning Algorithm," *Scientific reports*, vol. 9, no. 1, pp. 1–9, 2019.
- [47] T. G. Zeyen and J. Caprioli, "Progression of disc and field damage in early glaucoma," *Archives of Ophthalmology*, vol. 111, no. 1, pp. 62–65, 1993.
- [48] J. Caprioli, B. Prum, and T. Zeyen, "Comparison of methods to evaluate the optic nerve head and nerve fiber layer for glaucomatous change," *American journal of ophthalmology*, vol. 121, no. 6, pp. 659–667, 1996.
- [49] J. M. Tielsch, J. Katz, H. A. Quigley, N. R. Miller, and A. Sommer, "Intraobserver and interobserver agreement in measurement of optic disc characteristics," *Ophthalmology*, vol. 95, no. 3, pp. 350–356, 1988.
- [50] D. L. Budenz *et al.*, "Detection and prognostic significance of optic disc hemorrhages during the Ocular Hypertension Treatment Study," (in eng), *Ophthalmology*, vol. 113, no. 12, pp. 2137–43, Dec 2006, doi: 10.1016/j.ophtha.2006.06.022.
- [51] D. S. Greenfield and R. N. Weinreb, "Role of optic nerve imaging in glaucoma clinical practice and clinical trials," *American journal of ophthalmology*, vol. 145, no. 4, pp. 598–603. e1, 2008.
- [52] J. S. Myers, S. J. Fudemberg, and D. Lee, "Evolution of optic nerve photography for glaucoma screening: a review," *Clinical & Experimental Ophthalmology*, vol. 46, no. 2, pp. 169–176, 2018, doi: doi:10.1111/ceo.13138.



- [53] K. Nasrollahi and T. B. Moeslund, "Super-resolution: a comprehensive survey," *Machine vision and applications*, vol. 25, no. 6, pp. 1423–1468, 2014.
- [54] A. Ha, Y. K. Kim, and K. H. Park, "Conversion of Single Optic Disc Photography into 3-Dimensional Image," *Ophthalmology*, vol. 125, no. 12, p. 1873, 2018.
- [55] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," *arXiv preprint*, 2016.
- [56] R. E. Carlson and F. N. J. S. j. o. n. a. Fritsch, "Monotone piecewise bicubic interpolation," vol. 22, no. 2, pp. 386–400, 1985.
- [57] K. Simonyan and A. J. a. p. a. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014.
- [58] I. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.
- [59] D. Mahapatra and B. Bozorgtabar, "Retinal vasculature segmentation using local saliency maps and generative adversarial networks for image super resolution," *arXiv preprint arXiv:1710.04783*, 2017.
- [60] W. Shi *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1874–1883.
- [61] B. Lv, Y. Liu, S. Zhang, H. Zeng, and G. Zhu, "Super Resolution with Generative Adversarial Networks," 2018.
- [62] R. C. Streijl, S. Winkler, and D. S. Hands, "Mean opinion score (MOS) revisited: methods and

- applications, limitations and alternatives," *Multimedia Systems*, vol. 22, no. 2, pp. 213–227, 2016.
- [63] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. J. L. Simoncelli, "Image quality assessment: from error visibility to structural similarity," vol. 13, no. 4, pp. 600–612, 2004.
  - [64] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE transactions on acoustics, speech, and signal processing*, vol. 29, no. 6, pp. 1153–1160, 1981.
  - [65] J. Salvador and E. Perez–Pellitero, "Naive bayes super–resolution forest," in *Proceedings of the IEEE International conference on computer vision*, 2015, pp. 325–333.
  - [66] S. Schuler, C. Leistner, and H. Bischof, "Fast and accurate image upscaling with super–resolution forests," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3791–3799.
  - [67] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback Network for Image Super–Resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3867–3876.
  - [68] X. Yang, K. Xu, Y. Song, Q. Zhang, X. Wei, and R. W. Lau, "Image correction via deep reciprocating HDR transformation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1798–1807.
  - [69] R. Wang, Q. Zhang, C.–W. Fu, X. Shen, W.–S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6849–6857.

- [70] W. Ren *et al.*, "Low-light image enhancement via a deep hybrid network," *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4364–4375, 2019.
- [71] P. Dai, H. Sheng, J. Zhang, L. Li, J. Wu, and M. Fan, "Retinal fundus image enhancement using the normalized convolution and noise removing," *International journal of biomedical imaging*, vol. 2016, 2016.
- [72] A. Bandara and P. Giragama, "A retinal image enhancement technique for blood vessel segmentation algorithm," in *2017 IEEE international conference on industrial and information systems (ICIIS)*, 2017: IEEE, pp. 1–5.
- [73] Z. Shen, H. Fu, J. Shen, and L. Shao, "Understanding and Correcting Low-quality Retinal Fundus Images for Clinical Analysis," *arXiv preprint arXiv:2005.05594*, 2020.
- [74] H. Zhao, B. Yang, L. Cao, and H. Li, "Data-driven enhancement of blurry retinal images via generative adversarial networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019: Springer, pp. 75–83.
- [75] L. A. Hark *et al.*, "Philadelphia Telemedicine Glaucoma Detection and Follow-up Study: Methods and Screening Results," (in eng), *Am J Ophthalmol*, vol. 181, pp. 114–124, Sep 2017, doi: 10.1016/j.ajo.2017.06.024.
- [76] J. L. Odden *et al.*, "Telemedicine in long-term care of glaucoma patients," (in eng), *Journal of telemedicine and telecare*, p. 1357633x18797175, Sep 12 2018, doi: 10.1177/1357633x18797175.
- [77] I. BenZion and E. M. Helveston, "Use of telemedicine to assist ophthalmologists in developing countries for

- the diagnosis and management of four categories of ophthalmic pathology," *Clinical ophthalmology (Auckland, NZ)*, vol. 1, no. 4, p. 489, 2007.
- [78] D. Kiage, I. N. Kherani, S. Gichuhi, K. F. Damji, and M. Nyenze, "The Muranga Teleophthalmology Study: Comparison of Virtual (Teleglaucoma) with in-person clinical assessment to diagnose glaucoma," *Middle East African journal of ophthalmology*, vol. 20, no. 2, p. 150, 2013.
  - [79] T. Vu, T. M. Luu, and C. D. Yoo, "Perception-enhanced image super-resolution via relativistic generative adversarial networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 0–0.
  - [80] B. B. Parfenenkov and M. A. Panachev, "Comparison of Some Image Quality Approaches," in *A/ST (Supplement)*, 2014, pp. 48–53.
  - [81] K. Ma, T. Zhao, K. Zeng, and Z. J. I. T. o. I. P. Wang, "Objective quality assessment for color-to-gray image conversion," vol. 24, no. 12, pp. 4673–4685, 2015.
  - [82] A. T. Nguyen, D. S. Greenfield, A. S. Bhakta, J. Lee, and W. J. Feuer, "Detecting glaucoma progression using guided progression analysis with OCT and visual field assessment in eyes classified by international classification of disease severity codes," *Ophthalmology Glaucoma*, vol. 2, no. 1, pp. 36–46, 2019.
  - [83] H. W. Bae, K. H. Lee, N. Lee, S. Hong, G. J. Seong, and C. Y. Kim, "Visual fields and OCT role in diagnosis of glaucoma," *Optometry and Vision Science*, vol. 91, no. 11, pp. 1312–1319, 2014.
  - [84] M. Chen, X. Shi, Y. Zhang, D. Wu, and M. Guizani, "Deep features learning for medical image analysis

- with convolutional autoencoder neural network," *IEEE Transactions on Big Data*, 2017.
- [85] P. Sollich and A. Krogh, "Learning with ensembles: How overfitting can be useful," *Advances in neural information processing systems*, vol. 8, pp. 190–196, 1995.
  - [86] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.
  - [87] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
  - [88] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.
  - [89] S. Bock, J. Goppold, and M. Weiß, "An improvement of the convergence proof of the ADAM-Optimizer," *arXiv preprint arXiv:1804.10587*, 2018.

## 초 록

본 논문에서는 딥 러닝 기반의 진단 보조 시스템을 제안하였다. 새로운 방법이 녹내장 데이터에 적용되었고 결과를 평가하였다.

첫번째 연구에서는 스펙트럼영역 빛간섭단층촬영기(SD-OCT)를 딥 러닝 분류 기를 이용해 분석하였다. 스펙트럼영역 빛간섭단층촬영기는 녹내장으로 인한 구조적 손상을 평가하기 위해 사용하는 장비이다. 분류 알고리즘은 합성곱 신경망을 이용해 개발되었으며, 스펙트럼영역 빛간섭단층촬영기의 망막신경섬유층(RNFL)과 황반부 신경절세포내망상층 (GCIPL) 사진을 이용해 학습했다. 제안한 방법은 두개의 이미지를 입력으로 받는 이중입력합성곱신경망(DICNN)이며, 딥 러닝 분류에서 효과적인 것으로 알려져 있다. 이중입력합성곱신경망은 망막신경섬유층 과 신경절세포층 의 두께 지도를 이용하여 학습했으며, 학습된 네트워크는 녹내장과 정상 군을 구분한다. 이중입력합성곱신경망은 정확도와 수신기동작특성곡선하면적(AUC)으로 평가 되었다. 망막신경섬유층과 신경절세포층 두께 지도로 학습된 설계한 딥 러닝 모델을 조기 녹내장과 정상 군을 분류하는 성능을 평가하고 비교하였다. 성능평가 결과 이중입력합성곱신경망은 조기 녹내장을 분류하는데 0.869의 수신기동작특성곡선의 넓이와 0.921의 민감도, 0.756의 특이도를

보였다.

두번째 연구에서는 딥 러닝을 이용해 시신경유두사진의 해상도와 대비, 색감, 밝기를 보정하는 방법을 제안하였다. 시신경유두사진은 녹내장을 진단하는데 있어 효과적인 것으로 알려져 있다. 하지만, 녹내장의 진단에서 환자의 나, 작은 동공, 매체 불투명성 등으로 인해 평가가 어려운 경우가 있다. 초 해상도와 보정 알고리즘은 초 해상도 적대적생성신경망을 통해 개발되었다. 원본 고해상도의 시신경 유두 사진은 저해상도 사진으로 축소되고, 보정된 고해상도 시신경유두사진으로 보정되며, 보정된 사진은 시신경여백의 가시성과 근처 혈관을 잘 보이도록 후처리 알고리즘을 이용한다. 저해상도이미지를 보정된 고해상도이미지로 복원하는 과정을 초해상도적대적신경망을 통해 학습한다. 설계한 네트워크는 신호 대 잡음 비(PSNR)과 구조적유사성(SSIM), 평균평가점(MOS)를 이용해 평가 되었다. 현재의 연구는 딥 러닝이 안과 이미지를 4배 해상도와 구조적인 세부 항목이 잘 보이도록 개선할 수 있다는 것을 보여주었다. 향상된 시신경유두 사진은 시신경의 병리학적인 특성의 진단 정확도를 명확히 향상시킨다. 성능평가결과 평균 PSNR은 25.01 SSIM은 0.75 MOS는 4.33으로 나타났다.

세번째 연구에서는 환자 정보와 안과 영상(시신경유두 사진과 붉은색이 없는 망막신경섬유층 사진)을 이용해 녹내장 의심 환자를 분별하고 녹내장 의심 환자의 발병 연수를 예측하는 딥 러닝 모델을 개발하였다. 임상 데이터들은 녹내장을 진단하거나

예측하는데 유용한 정보들을 가지고 있다. 하지만, 어떻게 다양한 유형의 임상정보들을 조합하는 것이 각각의 환자들에 대해 잠재적인 녹내장을 예측하는데 어떤 영향을 주는지에 대한 연구가 진행 된 적이 없다. 녹내장 의 심자 분류와 발병 년 수 예측은 합성곱 자동 인코더(CAE)를 비 지도적 특성 추출 기로 사용하고, 기계학습 분류 기와 회귀기를 통해 진행하였다. 설계한 모델은 정확도와 평균제곱오차(MSE)를 통해 평가 되었으며, 이미지 특징과 환자 특징은 조합했을 때 녹내장 의심 환자 분류와 발병 년 수 예측의 성능이 이미지 특징과 환자 특징을 각각 썼을 때보다 성능이 좋았다. 정답과의 MSE는 2.613으로 나타났다.

본 연구에서는 딥 러닝을 이용해 녹내장 관련 임상 데이터 중 망막신경섬유층, 신경절세포층 사진을 녹내장 진단에 이용되었고, 시신경유두 사진은 시신경의 병리학적인 진단 정확도를 높였고, 환자 정보는 보다 정확한 녹내장 의심 환자 분류와 발병 년 수 예측에 이용되었다. 향상된 녹내장 진단 성능은 기술적이고 임상적인 지표들을 통해 검증되었다.

**주요어:** 딥러닝, 합성곱 신경망, 초해상도, 녹내장 진단, 녹내장 예측

**학 번:** 2016-30277



## Acknowledgement

박사학위 과정 동안 감사했던 모든 분들께 이 글을 바칩니다. 가장 먼저 지도 교수님 이신 김희찬 교수님께 존경과 감사를 전합니다. 박사 과정 학생으로 다양한 연구를 할 수 있도록 지원해주시고, 연구 자세와 방향에 대해 가르쳐 주셔서 학위 말까지 연구자의 길을 걸을 수 있었습니다. 졸업 후에도 교수님의 가르침을 마음속에 새기고 연구에 임하겠습니다.

다음으로는 제 학위 심사를 맡아주신 박기호 교수님, 김남국 교수님, 공현중 교수님, 유병욱 박사님께 감사합니다. 그리고 저의 졸업 주제에 가장 많은 도움을 주신 안과학 교실의 박기호 교수님, 김영국 교수님, 하아늘 선생님께 감사 인사를 올립니다. 안과학 교실 덕분에 제가 가장 활발히 연구 할 수 있었고, 제 연구가 꽃을 피울 수 있었습니다. 정진욱 교수님, 박은우 선생님, 위서영 선생님께도 감사합니다. 피부과의 정진호 교수님, 조수익 선생님께 감사합니다. 좋은 주제를 제안해 주셔서 짧은 기간 동안 좋은 성과를 낼 수 있었습니다. 아직 진행 중이지만 서대현 교수님께도 감사인사를 올립니다. 또한 재난 대응 과제로 많은 도움을 주셨던, 신상도 교수님, 정주 교수님, 김태한 선생님께 감사드립니다. 마지막으로 생체 신호에 입문할 수 있도록 도움을 주신 신경외과 백선하 교수님, 박광현 선생님께 감사 인사를 올립니다. 공동

연구를 통해 협업과 소통을 배우고, 임상적 의학적인 관점에 대해 배울 수 있었던 귀중한 시간이었습니다.

연구실 생활을 마지막까지 함께 했던, 희진이형, 치현이형, 장재형, 희안누나, 승만이, 지은이 그리고 졸업생 분들 중 연구 및 생활에 가장 도움을 많이 주신 병욱이형, CMI 시절부터 함께 생활하고 같은 연구를 진행했던 동현이형 및 모든 졸업생 분들께 감사 인사를 올립니다. MBDL의 윤하형, 동아누나, 우상이, 찬훈이, 준희, 경진이 에게도 감사인사를 올립니다.

저에게 있어 박사 학위 기간은 새로운 도전의 연속이었습니다. 그 때마다 항상 응원해주신 어머니, 아버지 그리고 누나에게 감사합니다.