

**RETINAL DISORDERS DETECTION AND ANALYSIS  
FROM FUNDUS AND OPTICAL COHERENCE  
TOMOGRAPHY IMAGES USING DEEP LEARNING  
MODELS**

Thesis

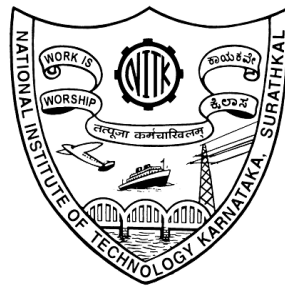
Submitted in partial fulfillment of the requirements for the degree of

**DOCTOR OF PHILOSOPHY**

by

**SMITHA A.**

(Reg No: 187118MA013)



DEPARTMENT OF MATHEMATICAL & COMPUTATIONAL SCIENCES

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA

SURATHKAL, MANGALORE - 575025

May, 2022



॥ चिदानन्दरूपः शिवोहं शिवोहं ॥

॥ नमस्ते शारदा देवी काश्मीरापुर - वासिनि ॥ त्वाम् अहं प्रार्थये नित्यं विद्याबुद्धिं च देहिमे ॥

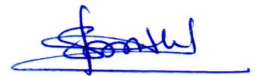
॥ मातृदेवो भव ॥ पितृदेवो भव ॥ आचार्यदेवो भव ॥

[*Dedicated to my parents, sister and husband.*]



## DECLARATION

I hereby *declare* that the Research Thesis entitled **RETINAL DISORDERS DETECTION AND ANALYSIS FROM FUNDUS AND OPTICAL COHERENCE TOMOGRAPHY IMAGES USING DEEP LEARNING MODELS** which is being submitted to the **National Institute of Technology Karnataka, Surathkal** in partial fulfillment of the requirements for the award of the Degree of **Doctor of Philosophy** in **Mathematical and Computational Sciences** is a *bonafide report of the research work carried out by me*. The material contained in this Research Thesis has not been submitted to any University or Institution for the award of any degree.



Smitha A

Reg. No: 187118MA013

Department of Mathematical and Computational Sciences

Place: NITK, Surathkal.

Date: 05 May 2022



## CERTIFICATE

This is to *certify* that the Research Thesis entitled **RETINAL DISORDERS DETECTION AND ANALYSIS FROM FUNDUS AND OPTICAL COHERENCE TOMOGRAPHY IMAGES USING DEEP LEARNING MODELS** submitted by **SMITHA A**, (Reg. No: 187118MA013) as the record of the research work carried out by her, is *accepted as the Research Thesis submission* in partial fulfillment of the requirements for the award of degree of **Doctor of Philosophy**.



Dr. JIDESH P.

Research Supervisor

Assistant Professor

Department of Mathematical and Computational Sciences

National Institute of Technology Karnataka



Chairman  
Chairman DRPC

Dept. of Mathematical and Computational Sciences  
National Institute of Technology Karnataka, Surathkal  
MANGALORE - 575 025





# ACKNOWLEDGMENTS

My journey of research at NITK has reached the stage of completion with the support of many individuals who have directly and indirectly played a significant role. I take this opportunity to thank everyone and acknowledge their contribution.

First and foremost, I am deeply indebted to my research supervisor *Dr. Jidesh P.*, faculty in the Department of MACS, NITK, for the unparalleled support, encouragement and research leads. I respect his immaculate work ethics which instilled the art of research and has been influential to me in all aspects. He remained proactive and paid heed to all my requirements even during lockdown of COVID-19 pandemic situation. His profound knowledge has helped me throughout my research work. Words fall short to recall the opportunities provided to attend various workshops, conferences, and International level research colloquiums. I consider it as a great privilege to work under his guidance.

I am extremely grateful to the Research Progress Assessment Committee members - *Dr. M. Venkatesan*, Department of Computer Science and Engineering and *Dr. Kedar-nath Senapathi*, Department of MACS for their insightful suggestions. I convey my deepest gratitude to *Dr. B.R.Shankar* and *Dr. Shyam S. Kamath* who were former Head of the Dept. of MACS for their relentless support. I recollect the helpful suggestions of *Dr. B.R.Shankar* when I was apprehensive to join as a research scholar. I also extend my sincere thanks to *Dr. R.Madhusudhan*, Head, Dept. of MACS for all the assistance and encouragement. My special thanks to *Dr. Srinivasa Rao Kola*, Secretary of Doctoral Research Programme Committee, Dept. of MACS for the co-operation throughout my research.

I extend my thanks to the Doctoral Thesis Assessment Committee members - *Dr. A.U. Narasimhadhan*, Assistant Professor, Dept. of Electronics and Communication, and *Dr. Partha Prathim Roy*, Associate Professor, IIT Roorkee, for their suggestions that helped me to improve the quality of my research work. I acknowledge them for spending their valuable time in reviewing my thesis.

I would like to acknowledge the efforts of *Dr. Santhosh George* and *Dr. Vishwanath K.P.* who taught me Linear Algebra and Numerical methods during the coursework.

They have never wavered their assistance. Their helpful advise and patience to answer my queries remain exceptional.

I express my deepest appreciation and heartfelt gratitude to *Mrs. Bhavani, Mrs. Sushma and Mr. Naveen*, non-teaching staff of Dept. of MACS. Be it briefing me about the rules and regulations or technical assistance, you all have always been instrumental, benevolent and co-operative. I extend my special thanks to all other teaching and non-teaching staff of the Department bonded together as mathematics family. It has been a great pleasure to be part of this MACS family. I am thankful to the head of Computer Science and Information Technology Department and the Central Computing Centre staff for their unabated assistance to access the high performance systems. It would have been impossible to complete the implementation of research objectives and submit the research articles on time, without their support.

I owe a great debt of honor to eminent scholars and research collaborators - *Dr. Vasudevan Lakshminarayanan*, Professor, University of Waterloo, Canada and *Dr. J. Jothi Balaji*, Ophthalmologist, Shankara Nethralaya, India, for tremendous insights. Every on-line discussion with them, including my guide as a mentor, played decisive role steering the research in right direction. The discussions helped me to be more confident and gain a different perception of the research work. I have greatly benefited from their warm encouragement and invaluable insights beyond the scope of my research work.

I would like to convey my deepest thanks to *Dr. Anupama Bappal*, Ophthalmologist, Mangalore, who took out time and put-in extra efforts to clarify and strengthen my understanding about the retinal imaging modalities. You have been persistent and amiable to provide me an opportunity to visit the retinal care center.

Every research work will be published only when the reviewers unanimously agree that the proposed work is acceptable. I extend my gratitude to all the unnamed reviewers and editors in chief of Biomedical Signal Processing and Control, Journal of Modern Optics, Multimedia Tools and Applications, and Springer Nature Computer Science. I also acknowledge the contributions of unannmed reviewers of International Conferences - CVIP, IHCI, and SPIE. The constructive critic and comments received aided me to elevate the quality of my work. I also thank the reviewers of this thesis report for

their time and valuable feedback.

I thank Ministry of Education, Government of India for providing financial support as Fellowship to carry out the research at NITK. I extend my thanks to *Mr. Jason Brownlee, Ms. Vineeta Das, Mr. Cheng-Bin Jin and Mr. Dhceeraj Rajaram Reddy* for helping me to resolve the implementation issues, when I started exploring deep learning as a juvenile researcher. I cannot begin to express my thanks to *Mr. Pruthviraj M.S.*, who sowed the seed of pursuing PhD and nurtured it within me all the way. You will remain as a tower of my strength. The meticulous comments, generous support, and illuminating discussions round-the-clock are revered always. I am thankful to *Mrs. Swathi B.P. and Mr. Akshay K.C.* for their embellished insights on Pattern recognition and Data structures. I could crack the entrance exams of NITK easily and the credits goes to you. My special thanks to *Dr. Febin J.P.*, and *Mrs. Veena Mayya*, who extensively helped me to carry out research. I am extremely grateful to *Ms. Jain M. Francis and Mrs. Architha Shasstry* for reviewing my thesis report with patience. I would like to recognize the assistance I received from Jain that aided to strengthen my mathematical concepts. I acknowledge *Mr. Abhinabha, Mrs. Sai Prasanna, Mr. Sushil Kumar, and Mrs. Athira* for their unconditional support. They have been a helping hand always. It would be incomplete if I do not mention the contribution of *Mrs. Madhura Bhat, Mrs. Namrata Chacko, and Mrs. Priya*. They always backed me emotionally. I enjoyed my stay at NITK because of my friends - *Mr. Natarajan, Mrs. Revathy, Ms. Sundari, Ms. Megha*, and all other fellow scholars. It is my pleasure to meet you all.

I thank my *parents and my elder sister* for their immense love and support that helped me to accelerate the research work. Without their cooperation, it would be impossible to reach this stage. The moral values you taught me has shaped me and influenced me positively since my childhood and I will remain indebted for everything. Finally, I would like to thank every one who directly or indirectly contributed to my success in career.

Place: NITK, Surathkal

Smitha A.

Date: 05 May 2022



# ABSTRACT

A brilliant vision gives us a luscious life. The retina of human eye plays a significant role in vision. Damage to any part of retina leads to visual impairments or total blindness. Diagnosis of retinal disorders is a daunting task for ophthalmologists as the devices are not equipped with automatic retinal analysis. The advent of deep learning has transformed the necessity of smart medical applications to a reality in recent years. However, the existing indigenous automatic retinal disorder detection system is designed to grade a single retinal disorder such as Diabetic Retinopathy.

Motivated by this, the proposed research aims to identify retinal disorders from multiple retina imaging modalities, namely fundus and Optical Coherence Tomography images. As fundus and Optical Coherence Tomography images differ in terms of image acquisition procedure, different artifacts affect the image quality. Adapting to the significant difference in image quality in these two modalities, two novel preprocessing approaches are proposed in Chapter 3 of the thesis. Histogram is used along with statistical analysis to assess the quality of the acquired retinal images. The proposed retinex based non-local total generalized variation restoration method enhances the fundus images increasing the visibility of the macula region. Realizing the fact that the speckle, inherent in Optical Coherence Tomography images are multiplicative in nature, a statistical analysis is incorporated to identify appropriate noise distribution. The non-local deep image prior, discussed in the thesis, despeckles Optical Coherence Tomography images eliminating the requirement of a large number of ground truth images for denoising. The proposed mathematical model using Bayesian MAP estimator and variational models are assessed through implementation. The qualitative and quantitative analysis presented in Chapter 3 of the thesis confirms that the proposed method outperforms other existing methods. The proposed model restores the image quality while retaining the edge and texture details in the image. Particularly, metrics such as Equivalent number of looks and entropy plots demonstrate that the proposed image restoration model works better than the other existing techniques.

Variants of Generative Adversarial Networks are proposed to classify the input retinal images into normal or abnormal categories. The abnormal categories include Age-

related Macular degeneration, Glaucoma, and Diabetic Macular Edema. Multiple publicly available repositories are preprocessed as described in Chapter 3 of the thesis. The preprocessed images are utilized to train the model and the results are presented in Chapter 4 of the thesis. Simultaneous segmentation and classification tasks are performed where the segmentation includes blood vessel extraction, optic disc region and fovea region extraction from fundus images. The performance of Generative Adversarial networks for various tasks such as segmentation and classification of retinal images is analysed in Chapter 4 of the thesis. The experimental analysis shows classification accuracy of upto 90% can be achieved proving the stability of a GAN amongst heterogeneous datasets. Other classification metrics such as F1-score and sensitivity are used to compare the proposed GAN model with other deep learning models.

In short, the thesis provides deeper insight into predominant retinal disorders, imaging modalities, existing state-of-the-art works on the retinal image analysis through Chapters 1 and 2. The significant contributions of the thesis are discussed in Chapters 3 and 4. Finally the conclusion and scope of future work is presented in Chapter 5. This research work acts as a cornerstone in developing an end-to-end standalone application.

**Keywords:** *Age-related macular degeneration, Fundus, Generative Adversarial Networks, Glaucoma, Optical Coherence Tomography, Non-local Total Variation, Retinex theory.*

**AMS Classification:** 68U10, 97R40, 92C50.

# Contents

<b>Abstract</b> . . . . .	i
<b>List of Figures</b> . . . . .	vi
<b>List of Tables</b> . . . . .	ix
<b>List of Abbreviations</b> . . . . .	xiii
<b>List of Symbols</b> . . . . .	xv
<b>1 INTRODUCTION</b>	<b>1</b>
1.1 RETINA . . . . .	2
1.2 RETINAL DISORDERS . . . . .	3
1.3 IMAGING MODALITIES . . . . .	8
1.3.1 Color Fundus Photography . . . . .	9
1.3.2 Optical Coherence Tomography . . . . .	11
1.4 THE NEED FOR NOVEL PREPROCESSING ALGORITHMS . . . . .	13
1.5 THE NEED FOR AUTOMATED ANALYSIS . . . . .	14
1.6 MOTIVATION . . . . .	15
1.7 OBJECTIVES . . . . .	16
1.8 CONTRIBUTION OF THE THESIS . . . . .	17
1.9 ORGANIZATION OF THE THESIS . . . . .	20
<b>2 A COMPREHENSIVE REVIEW OF RETINAL IMAGE ANALYSIS</b>	<b>23</b>
2.1 STATE-OF-THE-ART REVIEW PAPERS . . . . .	23
2.2 PREPROCESSING OF RETINAL IMAGES . . . . .	26
2.2.1 Traditional image restoration methods . . . . .	26
2.2.2 Variational image restoration models . . . . .	28
2.2.3 Non-local image restoration . . . . .	29
2.2.4 Deep learning architectures . . . . .	31
2.3 FUNDUS IMAGE ANALYSIS . . . . .	32
2.3.1 Classification . . . . .	34

2.4	OPTICAL COHERENCE TOMOGRAPHY IMAGE ANALYSIS . . . . .	36
2.4.1	Classification . . . . .	37
2.5	RESEARCH GAPS . . . . .	38
2.6	SUMMARY . . . . .	41
<b>3</b>	<b>RESTORATION AND ENHANCEMENT OF RETINAL IMAGES</b>	<b>43</b>
3.1	DATA DISTRIBUTION ANALYSIS . . . . .	44
3.2	BAYESIAN MAP ESTIMATE FOR GAMMA DISTRIBUTION . . . . .	48
3.3	RETINEX FRAMEWORK . . . . .	49
3.3.1	Need of L1 and L2 regularizers . . . . .	51
3.3.2	Proposed Non-local TGVR framework . . . . .	52
3.3.3	Numerical Implementation . . . . .	54
3.3.4	Experimental results . . . . .	56
3.4	NON-LOCAL DEEP IMAGE PRIOR . . . . .	61
3.4.1	Experimental results . . . . .	63
3.5	SUMMARY . . . . .	69
<b>4</b>	<b>GENERATIVE ADVERSARIAL NETWORKS FOR AUTOMATIC RETI-</b>	
	<b>NAL DISORDER DETECTION</b>	<b>73</b>
4.1	DATASETS . . . . .	73
4.2	GENERATIVE ADVERSARIAL NETWORKS-PRELIMINARIES . . . . .	74
4.3	GAN MODEL FOR FUNDUS IMAGES . . . . .	79
4.4	GAN MODEL FOR OCT IMAGES . . . . .	83
4.5	EXPERIMENTAL RESULTS . . . . .	87
4.5.1	Fundus image classifier . . . . .	88
4.5.2	Analysis of classifier model . . . . .	90
4.5.3	Optical Coherence Tomography (OCT) image classifier . . . . .	92
4.5.4	Design of Wasserstein GAN model . . . . .	96
4.6	SUMMARY . . . . .	98
<b>5</b>	<b>CONCLUSION AND FUTURE WORKS</b>	<b>101</b>
5.1	MULTI-MODAL ANALYSIS . . . . .	101
5.2	CONCLUSION . . . . .	103
5.3	FUTURE WORKS . . . . .	105



<b>APPENDIX I</b>	<b>107</b>
A.1 INVERSE PROBLEMS AND ILL-POSED PROBLEMS . . . . .	107
A.2 REVIEW OF VARIATIONAL FRAMEWORKS . . . . .	108
A.3 EULER-LAGRANGE METHOD . . . . .	110
A.4 SPLIT-BREGMAN APPROACH . . . . .	110
 <b>APPENDIX II</b>	 <b>111</b>
B.1 Deep Image Prior . . . . .	111
 <b>APPENDIX III</b>	 <b>114</b>
C.1 LAYERS OF DEEP LEARNING ARCHITECTURE . . . . .	114
C.2 ACTIVATION FUNCTIONS . . . . .	117
C.3 HYPERPARAMETERS . . . . .	118
 <b>APPENDIX IV</b>	 <b>118</b>
D.1 STANDARD METRICS FOR QUANTITATIVE ANALYSIS . . . . .	118
D.1.1 Jensen–Shannon divergence (JSD) . . . . .	118
D.1.2 Global Contrast Factor (GCF) . . . . .	119
D.1.3 Entropy . . . . .	119
D.1.4 Peak Signal to Noise Ratio (PSNR) . . . . .	119
D.1.5 Structural Similarity (SSIM) . . . . .	120
D.1.6 Contrast Noise Ratio (CNR) . . . . .	120
D.1.7 Equivalent Number of Looks (ENL) . . . . .	120
D.1.8 Wasserstein Distance . . . . .	121
D.1.9 Classification metrics . . . . .	121
D.2 GABOR FILTER . . . . .	123
 <b>BIBLIOGRAPHY</b>	 <b>124</b>
 <b>List of Publications</b>	 <b>146</b>



# List of Figures

1.1	Anatomy of human eye. . . . .	2
1.2	Retinal region around foveal pit. . . . .	4
1.3	Statistics of visual impairments. . . . .	6
1.4	Age-related Macular Degeneration stages. . . . .	7
1.5	Diabetic Macular Edema. . . . .	7
1.6	Glaucoma and healthy fundus image. . . . .	8
1.7	Multi-modal imaging for AMD. . . . .	9
1.8	Different devices to capture retina image. . . . .	10
1.9	Abnormalities visible from Color Fundus Photographs. . . . .	10
1.10	Working principle of direct ophthalmoscope. . . . .	11
1.11	Working principle of OCT. . . . .	12
1.12	Abnormalities visible on OCT images. . . . .	13
1.13	AI-enabled products. . . . .	15
1.14	Overall workflow. . . . .	17
2.1	Overview of existing literature on retinal image analysis. . . . .	24
2.2	Demonstration of non-local similarity computation. . . . .	29
3.1	Randomly selected degraded fundus images. . . . .	44
3.2	Randomly selected degraded OCT images. . . . .	45
3.3	Histogram comparison of CFP. . . . .	45
3.4	Histogram comparison of OCT images. . . . .	46
3.5	Analysis of data distribution in a sample OCT image. . . . .	47
3.6	Data distribution of another OCT image. . . . .	47
3.7	Input and enhanced fundus images. . . . .	58
3.8	Histogram of degraded and enhanced images. . . . .	59

3.9	Average contrast plot. . . . .	60
3.10	Global Contrast Factor. . . . .	60
3.11	Entropy plot. . . . .	60
3.12	Architecture of proposed non-local deep image prior model. . . . .	64
3.13	Results of despeckling synthetic image. . . . .	64
3.14	PSNR values after restoring synthetic image shown in Figure 3.13 (b). . . . .	65
3.15	SSIM values after restoring synthetic image shown in Figure 3.13 (b). . . . .	66
3.16	Results of despeckling OCT images. . . . .	66
3.17	Progression of non-local deep image prior model. . . . .	67
3.18	Zoomed view of despeckling OCT images. . . . .	68
3.19	Patches for quantitative analysis. . . . .	69
3.20	Average Contrast Noise Ratio for the input images shown in Figure 3.16. . . . .	70
3.21	Average ENL values for the input images shown in Figure 3.16. . . . .	70
4.1	Basic architecture of Generative Adversarial Network (GAN). . . . .	76
4.2	Block diagram for fundus image analysis. . . . .	79
4.3	Architecture of Vessel-Generative Adversarial Network (VGAN). . . . .	80
4.4	Segmentation outputs of fundus image. . . . .	82
4.5	Semi-supervised GAN architecture. . . . .	83
4.6	Block diagram for OCT image analysis. . . . .	84
4.7	Architecture of Wasserstein GAN. . . . .	85
4.8	Segmented output for ODIR dataset. . . . .	88
4.9	Classifier performance for multiple retinal disorder detection. . . . .	91
4.10	Graph depicting need of preprocessing. . . . .	91
4.11	Segmentation output of Wasserstein GAN. . . . .	93
4.12	Training and validation loss plot with epochs. . . . .	93
4.13	Confusion matrices of Wasserstein GAN. . . . .	95
4.14	Effect of varying kernel size and learning rate. . . . .	97
4.15	Effect of including Gabor feature. . . . .	98
5.1	Conceptual diagram for multi-modal analysis. . . . .	102
B.1	Representation of parameterized learning. . . . .	112

B.2	Denoising strategy in deep image prior. . . . .	113
C.1	Convolution operation. . . . .	114



# List of Tables

1.1	Some abnormalities in the retina. . . . .	5
2.1	Comparison of existing OCT image restoration methods. . . . .	33
2.2	Summary of existing works on fundus image analysis. . . . .	34
2.3	Summary of existing works on fundus image analysis using GANs. . . . .	35
2.4	Summary of existing OCT image classifiers. . . . .	39
3.1	PSNR values (in dB) of OCT images. . . . .	67
4.1	List of CFP datasets in public domain. . . . .	75
4.2	Details of OCT repositories in public domain. . . . .	76
4.3	Performance metrics for segmentation. . . . .	88
4.4	Performance metrics of the classifier for IDRID dataset. . . . .	89
4.5	Performance of proposed fundus image classifier on ODIR dataset. . . . .	91
4.6	Classification metrics on Dataset 1 (Reza et al. (2018)). . . . .	94
4.7	Classification metrics on Dataset 2 (Kermany et al. (2018a)). . . . .	94
5.1	Summary of existing works on multi-modal analysis. . . . .	102
C.1	Activation functions. . . . .	117





## List of Abbreviations

<b>AMD</b>	Age-related Macular Degeneration
<b>AUC</b>	Area under Region of Characteristic Curve
<b>CFP</b>	Color Fundus Photography
<b>CLAHE</b>	Contrast Limited Adaptive Histogram Equalization
<b>CNN</b>	Convolutional Neural Network
<b>CNV</b>	Choroidal Neovascularization
<b>CSC</b>	Central Serous Chorioretinopathy
<b>CNR</b>	Contrast Noise Ratio
<b>DIP</b>	Deep Image Prior
<b>DME</b>	Diabetic Macular Edema
<b>DR</b>	Diabetic Retinopathy
<b>DRIVE</b>	Digital Retinal Images for Vessel Extraction
<b>ENL</b>	Equivalent Number of Looks
<b>FFA</b>	Fundus Fluorescein Angiography
<b>GAN</b>	Generative Adversarial Network
<b>GCF</b>	Global Contrast Factor
<b>HRF</b>	High-Resolution Fundus
<b>IDRiD</b>	Indian Diabetic Retinopathy Image Dataset
<b>NLTVR</b>	Non-local Total Variational Retinex
<b>NLTGVR</b>	Non-local Total Generalized Variational Retinex
<b>OCT</b>	Optical Coherence Tomography

<b>ODIR</b>	Ocular Disease Recognition
<b>PRISMA</b>	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
<b>PSNR</b>	Peak Signal to Noise Ratio
<b>REFUGE</b>	Retinal Fundus Glaucoma Challenge
<b>SEAD</b>	Symptomatic Exudate-associated Derangements
<b>SSIM</b>	Structural Similarity
<b>STARE</b>	STructured Analysis of the Retina
<b>TGV</b>	Total Generalized Variation
<b>TV</b>	Total Variation
<b>VGAN</b>	Vessel-Generative Adversarial Network
<b>WHO</b>	World Health Organization
<b>WGAN</b>	Wasserstein Generative Adversarial Network

## List of Symbols

Symbol	Representation	Symbol	Representation
$U_0$ :	Observed degraded image.	$U$ :	Clean image.
$\hat{U}$ :	Restored image.	$u$ :	Intensity in log domain.
$R$ :	Reflectance.	$r$ :	Reflectance in log domain.
$L$ :	Illuminance.	$l$ :	Illuminance in log domain.
$N$ :	Noise distribution	$\Omega$ :	Image domain.
$G$ :	Generator.	$D$ :	Discriminator.
$K$ :	Number of classes in classification.	$W(\cdot)$ :	Weight function.
$X_{ul}$ :	Unlabeled set of images for segmentation.	$Y_{ul}$ :	Unlabeled set of groundtruth images for segmentation.
$X_l$ :	Labeled set of images for classification.	$Y_l$ :	Classification labels of $X_l$ .
$\mathbb{E}[\cdot]$ :	Standard Expectation notation.	$E(\cdot)$ :	Energy minimization.
$\rho(\cdot)$ :	Standard correlation function.	$\sigma$ :	Standard deviation.
$ \cdot $ :	Standard absolute operation.	$\ \cdot\ _2$ :	Standard $L_2$ norm.
$\nabla$ :	Gradient operation.	$H$ :	Histogram of image.
$c_1, c_2$ :	Constraints.	$a_1, a_2$ :	Auxiliary variables.
$p(\cdot)$ :	Probability of the event.	$f(\cdot)$ :	Function.
$\alpha, \zeta$ :	Shape, scale parameters.	$\Gamma(\cdot)$ :	Gamma function.
$\Phi(\cdot)$ :	Level set function.	$G_\sigma$ :	Gaussian kernel

$\lambda, \beta, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \gamma_7, \gamma_8$  denote regularization parameters.



# Chapter 1

## INTRODUCTION

Human eyes are the windows to experience the world around us visually. The complicated structure of the eye has inspired us to invent devices such as the digital camera. The anatomy of the human eye is shown in Figure 1.1. It has a cornea, pupil, iris, lens, retina, optic nerve, and sclera, which coordinate with the human brain to produce pictures of the objects. Specifically, the human eye could be divided into three parts - the anterior portion, which consists of cornea and sclera, the posterior side comprising of the ciliary body, choroid, and iris and the third part, vitreous chamber, filled with vitreous humor, comprises retina, macula, and optic disc (Khurana (2015)). The anterior and posterior chambers are filled with aqueous humor. The ciliary body is the muscle that assists the accommodation process of the lens. The ciliary body, iris, and the choroid region are together referred to as uvea. The converging lens present in the human eye is made of proteins.

When an object is illuminated, the lens directs the reflected rays from that object towards the inner walls of the retina. The retina is a vital part of the human eyes, responsible for the reception of light from the external source and conversion of this light into neural impulses (recognized by our brain). The rods and cones (photoreceptors) are responsible for color and brightness information. Further, a sharp central vision is obtained in a macula region within the retina, and an inverted image of the illuminated object is formed. The image information is sent to the brain through optic nerves where further processing is carried out. All these materialize in a span of 13 milliseconds (Trafton (2014)). Any image formed beyond the macula region appears as

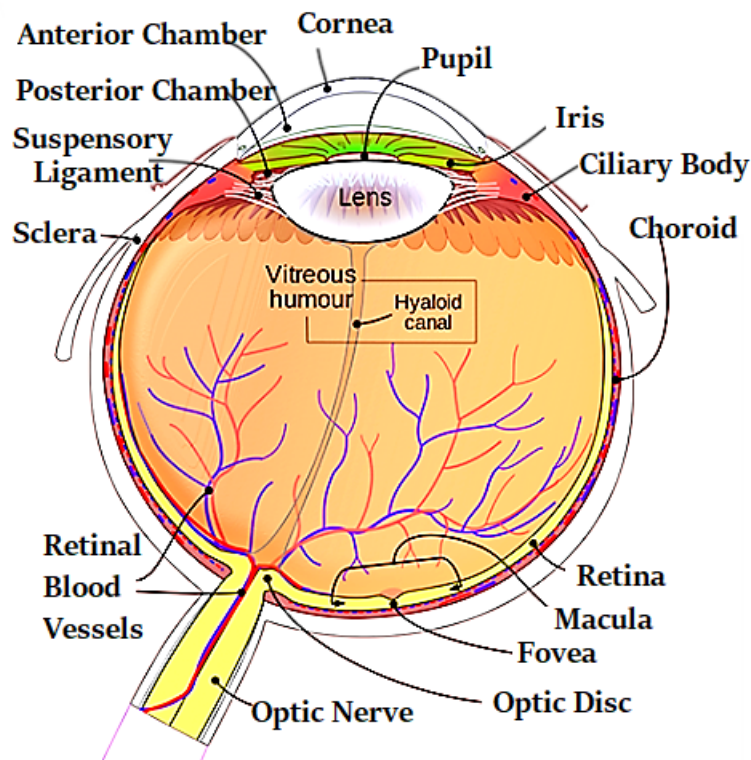


Image source: <https://www.umkelloggeye.org/conditions-treatments/anatomy-eye>

Figure 1.1: Anatomy of human eye.

cloudy/blurred vision.

## 1.1 RETINA

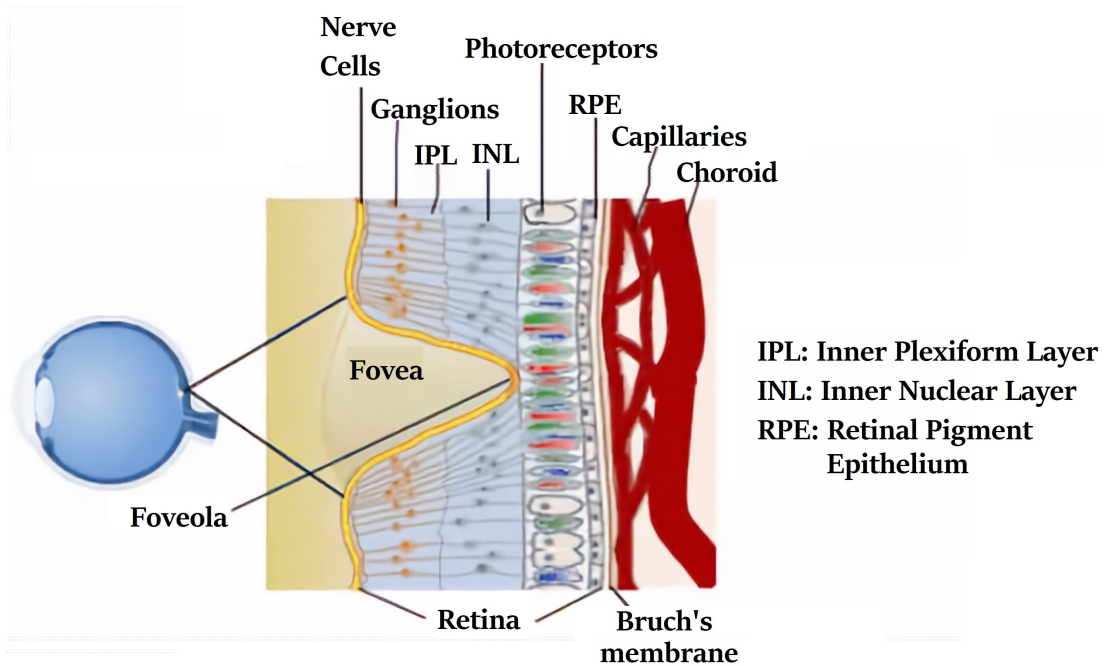
The retina is composed of several tissues. Derived from Latin, the word "rete" refers to a net. It covers nearly 65% of the back of the eye (Ehrlich et al. (2011)). The retina is a layer of tissue comprising the macula, blood vessels, and optic disc. The macula consists of a foveal pit. The macula region has approximately 5.5mm diameter (Helga et al. (2007)). Light entering the retinal layers triggers the photochemical reaction at the photoreceptor layer. The rods function in low light. It detects the motion, and it is absent in the fovea. The word "fovea" refers to a pit in the Latin language. In foveal region, the concentration of cone cells is high. Cones are responsible for color information, and hence the visual acuity is high in the central fovea. The region termed as a foveal avascular zone (with 0.5mm diameter), is devoid of blood vessels to ensure high vision. The photochemical reaction activates the bipolar cells and ganglion cells which carry the information in the form of neural impulses exiting from the optic disc

region. The occipital lobe of the brain now takes over to process the information and recognize the image.

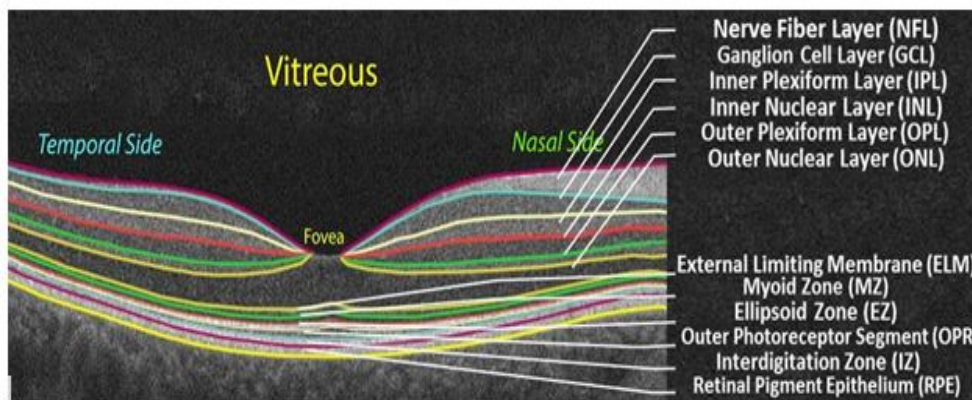
As shown in Figure 1.2 (a and b), the layer of tissues present in the retina perform various tasks. Each layer has a different capacity for reflecting and scattering the optic ray. The retinal layers are the retinal nerve fiber, ganglion cell, inner and outer plexiform, inner and outer nuclear, inner and external limiting membranes, retinal pigment epithelium, and vascular region defined by the choroid layer. The thickness of a normal human retinal nerve fiber layer ranges between  $45 \mu\text{m}$  and  $105 \mu\text{m}$  (Refer: Jonathan and Micieli (2019)). The optic disc region is filled with axons of ganglion cells, and it is from here that the neural impulses travel towards the brain. As the disc region is devoid of photoreceptors, it is also referred to as the *blind spot*. The central part of the optic disc is called optic cup region. The entire retinal layer is nourished by the blood vessels. They provide oxygen necessary for the functioning of the cells. The choroid layer is sandwiched between the retina and sclera. The choroid layer provides oxygen and essential nutrients to the neighboring retinal tissue layers. The coordination between various retinal tissues and the brain leads to a combined image formation, though the left and right eyes work independently, even in the low-illuminated condition. Thus, retinal layers play a significant role in vision.

## **1.2 RETINAL DISORDERS**

Loss of vision is a world of darkness. Blindness can be by birth, due to injuries, systemic diseases, or it could also be a natural phenomenon (age, genetic, etc). The retinal disorder is the condition where the blood vessels or the retinal tissue layers are damaged, leading to poor visibility or disrupted image formation. A plethora of disorders manifest themselves in the retina. Retinal disorders, if untreated, can lead to severe vision impairment or even total blindness. Ophthalmologists take approximately 5 to 10 minutes or even longer to examine the retina on a case-to-case basis. Certain commonly identified abnormalities in the retina are described in Table 1.1. Glaucoma, Diabetic Retinopathy (DR), Age-related Macular Degeneration (AMD), Diabetic Macular Edema (DME), Choroidal Neovascularization (CNV), and Retinitis pigmentosa



(a) Histology of retina.



(b) Retina visualized through tomography.

Image source: <https://www.everydaysight.com/fovea/> and Schaal et al. (2016)

Figure 1.2: Retinal region around foveal pit.

are some of the widespread retinal disorders.

Globally, the World Health Organization (WHO) fact sheet of 2018 reports that nearly 2.2 billion people or more suffered from visual impairments, of which 75% are preventable disorders (Refer: Bourne et al. (2017)). Overall, nearly 20.5% of the global blind population is constituted by the Indian population. In the year 2017, nearly 1,47048 were treated for Glaucoma and DR (Refer: Website (2017)). The economic impact of this scenario is summarized by Jitendra and Pradeep (2019). The statistics



Table 1.1: Some abnormalities in the retina.

<b>Abnormality</b>	<b>Description</b>
Hemorrhage	Abnormal bleeding of blood vessels.
Edema	Swollen regions and thickening of macula due to accumulation of fluid.
Drusen	Excessive deposition of lipids and proteins under the retinal layers.
Macular Cyst	Fluid-filled macular regions that obstructs blood vessels and shrinks the vitreous region.
Exudates	Protein depositions that results in fluid leakage.
Microaneurysms	Tiny red dots around yellow rings found on the fundus indicating the rupture of vascular network.

of visual impairment causes in people of age-group above 50 years is depicted in Figure 1.3. From Figure 1.3 (a), it is evident that the AMD is the primary cause of visual impairment globally, contributing about 37% of the total blind population. Other retinal disorders such as Glaucoma and DR are 15% and 17%, respectively. In India, the number of cataract cases is significant, with about 66% of the blind population. However, cataracts can be treated by surgery, and it is related to the lens of the eye. When the retinal disorders are considered together, Glaucoma and AMD are predominant, as depicted in Figure 1.3 (b). Majority of the retinal disorders, such as AMD, DME, and Glaucoma are progressive and irreversible. These disorders necessitate frequent visit to the eye care centers in order to monitor the progression of the disorders. Early diagnosis of these disorders can thus be highly beneficial. Considering the impact of these retinal disorders on the elderly community, our research work is limited to AMD, DME, and Glaucoma. A brief description of these retinal disorders is provided below (Raeba and Sobha (2014)).

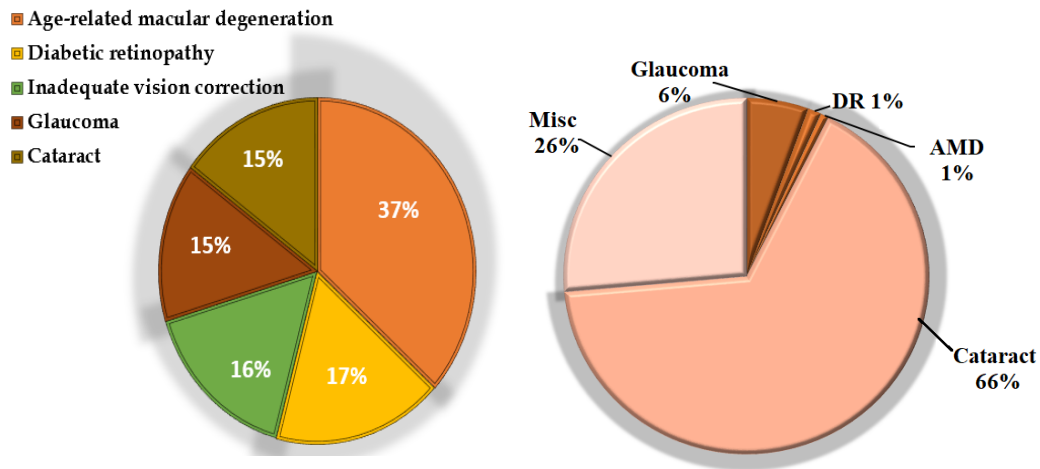
**AMD:**

**Modality:** Fundus Fluorescein Angiography (FFA), Optical Coherence Tomography OCT, Color Fundus Photography (CFP), and Amsler grid examination.

**Biomarkers:** Drusen, exudates, neovascular structures.

**Causes:** Genetics, age, high cholesterol.

**Description and grading according to ICD:**



(a) Global causes for visual impairments in people aged above 50 years. (Information from: <https://lowvision.preventblindness.org/2021/03/07/latest-statistics-on-prevalence-of-blindness-and-low-vision-worldwide/>).

(b) Causes of visual impairments in Indian population for people aged above 50 years: Source (NB&VI report - 2020).

Figure 1.3: Statistics of visual impairments.

1. The grading is between H35.30 and H35.35 (Refer International Statistical Classification of Diseases and Related Health Problems: <https://www.icd10data.com/ICD10CM/Codes>), depending on the type of AMD as shown in Figure 1.4
2. The early stage of AMD is diagnosed as Dry AMD or Atrophic AMD. Early AMD is caused by thinning of vessels near the macula or debris deposition of deteriorating tissue (Drusen).
3. The later stage of AMD is identified as Wet AMD or Neovascular Atrophy. Late AMD is caused by abnormal growth of vessels beneath the retina.

**DME:**

**Modality:** FFA, OCT

**Biomarkers:** Macular thickening or edema, exudates, fluid-filled cysts.

**Causes:** Diabetes Mellitus, damage from radiations.

**Description and grading according to ICD:**

1. The grading is E11.311 denoting type 2 diabetes with edema and retinopathy (Refer: <https://www.icd10data.com/ICD10CM/Codes>).

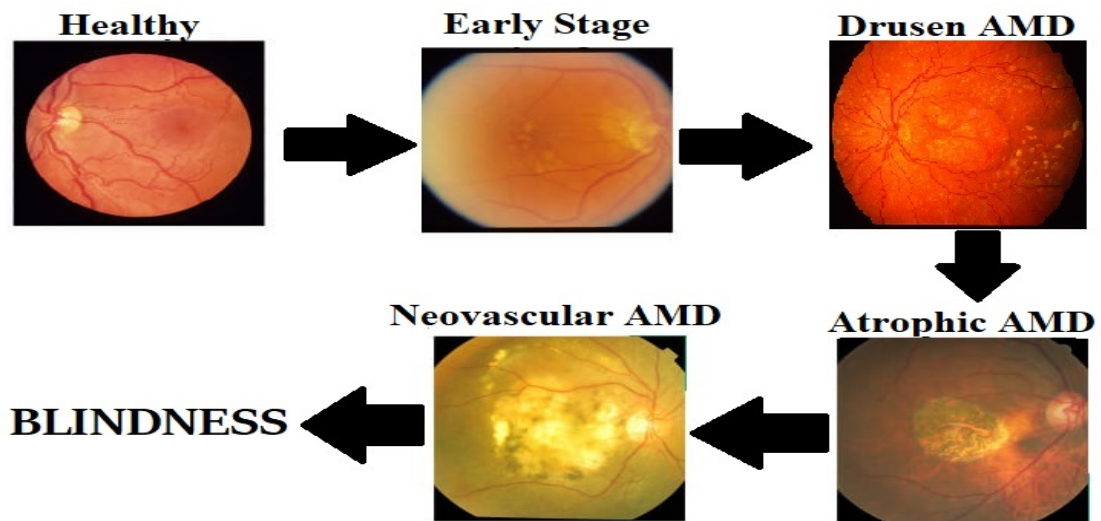


Image source: <https://imagebank.asrs.org/>

Figure 1.4: Age-related Macular Degeneration stages.

2. The condition of diabetic retinopathy graded from mild to the proliferative stage is either associated with edema in the macular region or without edema.
3. The DME is related to swollen macular region as depicted in Figure 1.5.

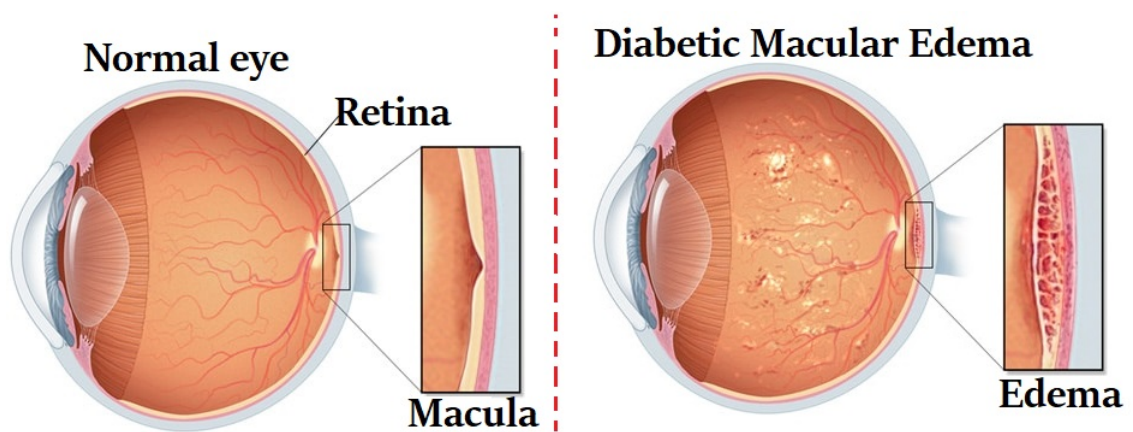


Image source: <https://www.mayoclinic.org/diseases-conditions/diabetic-retinopathy/multimedia/diabetic-macular-edema/img-20124558>

Figure 1.5: Diabetic Macular Edema.

**Glaucoma:**

**Modality:** Tonometry, CFP, OCT, Gonioscopy.

**Biomarkers:** Optic cup disc ratio, retinal nerve fiber layer thickness, interocular pressure

**Causes:** Variation in ocular pressure.

**Description and grading according to ICD:**

1. The ICD 10 grading for glaucoma is H40 (Refer International Statistical Classification of Diseases and Related Health Problems: <https://www.icd10data.com/ICD10CM/Codes>).
2. Figure 1.6 depicts the optic cup and optic disc variation with reference to Glaucoma.
3. The most common types of glaucoma are open-angle and angle-closure glaucoma. The open-angle refers to the angle between iris and cornea of the eye.

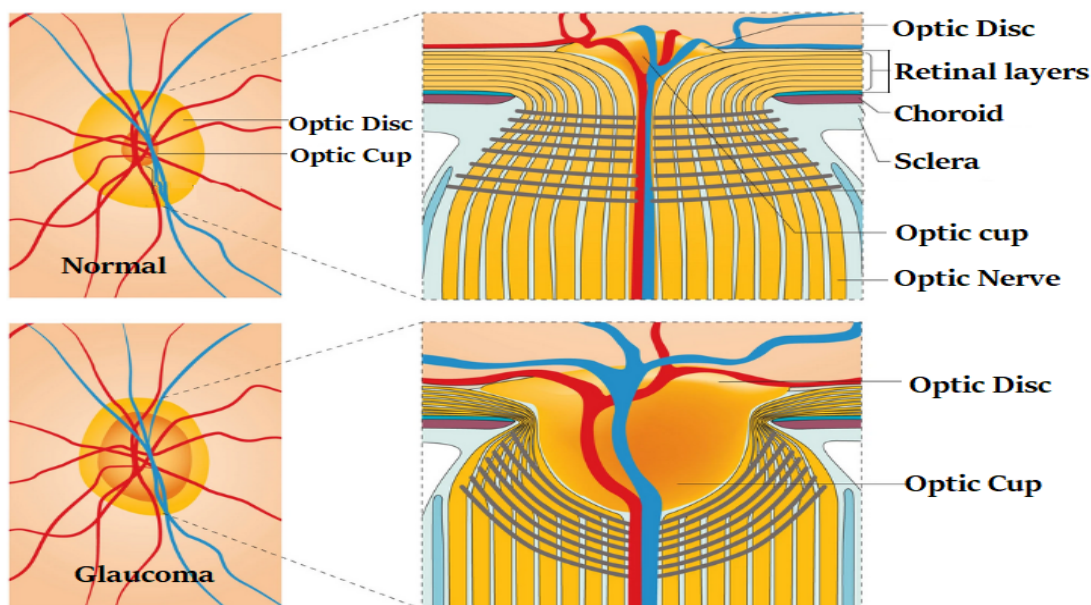


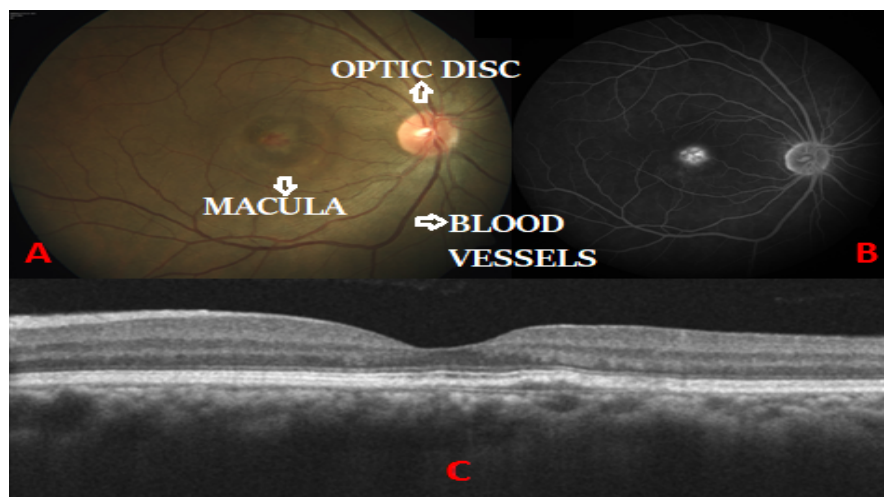
Image source: <https://www.nature.com/articles/nrdp201667/figures/1>

Figure 1.6: Glaucoma and healthy fundus image.

### 1.3 IMAGING MODALITIES

Ophthalmologists examine the retina using various imaging modalities like CFP, OCT, and Fundus Fluorescein Angiography (FFA) to discern the retinal abnormalities and diagnose the disorders. It is often necessary to rely on multiple imaging modalities to accurately detect the disorder on a case-to-case basis, as depicted in Figure 1.7. There

are several modern retina examining devices available in the market, which is elaborately explained by Panwar et al. (2016). However, retinal disorders are predominantly diagnosed using CFP. Ophthalmoscopes are used to get the fundus image and spot the clinical findings such as microaneurysms, hemorrhages, and exudates. Throughout the rest of the document, the term 'fundus' refers to images acquired using CFP. In general, fundus photographs do not give a detailed view of the retina layers and OCT are used for this purpose. The sample devices used to acquire images through these modalities are depicted in Figure 1.8. FFA is an invasive procedure, generally considered for detailed diagnosis on the progression of AMD. However, as the introduction of OCT has paved the way for multi-modal analysis with a wider focal area, our study is based on CFP and OCT. These modalities are described in the following subsections.



Images courtesy of Rajiv Raman et al. Sankara Nethralaya, India.

Figure 1.7: Images of the right eye of a 60-year-old woman with AMD. A) CFP (top-left), B) FFA (top-right), and C) OCT (bottom).

### 1.3.1 Color Fundus Photography

Fundus imaging is a technique to obtain 2D images of the retina. The essential structures visible from the fundus image are the optic disc, fovea, macula, and retinal blood vessels. The arteries and veins carry oxygenated and deoxygenated blood from and to the heart, respectively. Any slight rupture in the blood vessels or variation of the cup to disc ratio will lead to eye disorders. There could be hemorrhage, cysts, or lesions of different stages formed in the vessel tissues. If these are left untreated, it can cause



(a) Spectral Domain OCT of Zeiss.



(b) Ophthalmoscope of Heine.

Figure 1.8: Different devices to capture retina image.

blindness. The common abnormalities spotted on the fundus image are depicted in Figure 1.9. Ophthalmologists diagnose systemic diseases from fundus photography like DR, hypertensive retinopathy, possibilities of stroke. The fundus images can also identify other eye disorders like cataracts, myopia, and macular degeneration (Ryan et al. (2013)). Fundus imaging techniques are non-invasive. The illumination within the eye is not sufficient to visualize the fundus. So, an external beam of light is projected at a certain angle to obtain a good field of view of the retina using ophthalmoscopes.

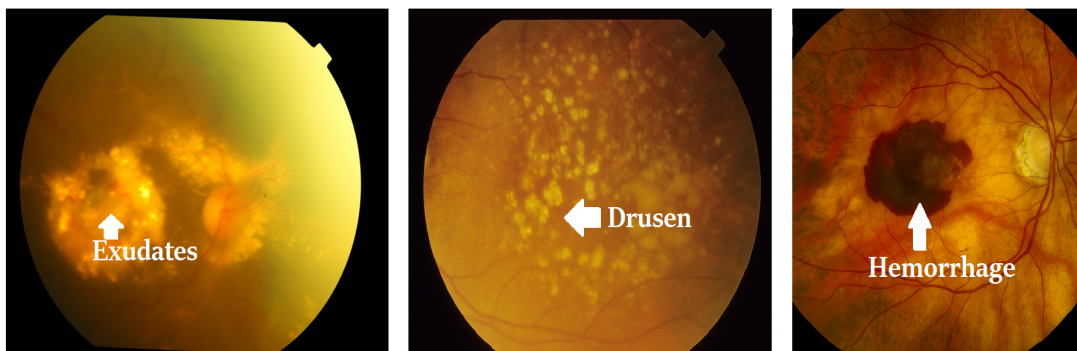


Image Source: <https://imagebank.asrs.org>

Figure 1.9: Abnormalities visible from Color Fundus Photographs.

There are two types of ophthalmoscopes: direct and indirect. The handheld instrument is called a direct ophthalmoscope. Figure 1.10 depicts the working principle of a direct ophthalmoscope (George and Michael (2005)). The handheld device has an illumination set up consisting of two lens, an aperture, and a mirror, to focus the beam of light on the patient's eye. The illuminated light is converged using lens 2. The light rays

diverge at the mirror and enter the patient's eye. The image of the patient's retina will be formed directly in the viewer's eye. The compensation lens allows the doctor to fine-tune the visualization of the patient's retina. As the device is restricted to a smaller field of focus, device calibration is necessary to obtain high-quality, high-contrast images. The indirect ophthalmoscope involves the external placement of a condensing lens and the device is worn like a headband by the viewer. Artifacts such as low contrast and blur are common in acquired fundus images. Thus, image enhancement or illumination inhomogeneity correction is needed prior to fundus images analysis.

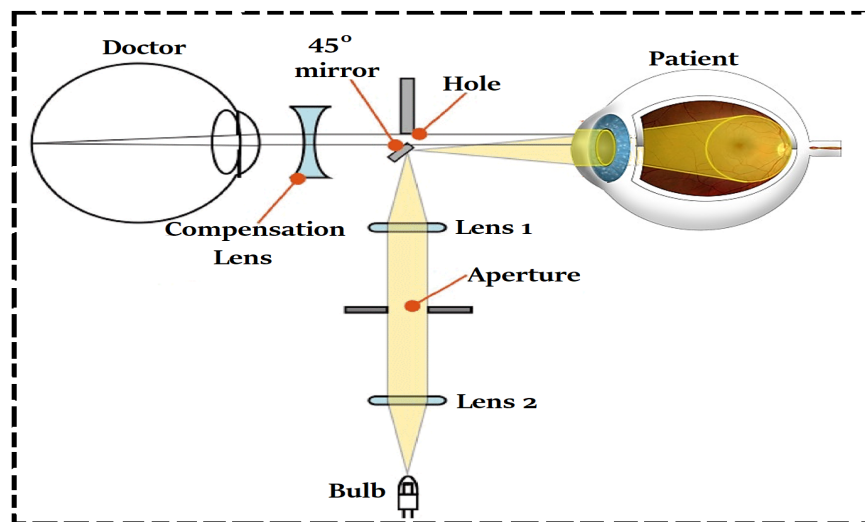


Image modified from: <https://www.cehjournal.org/article/understanding-and-caring-for-the-direct-ophthalmoscope/> and <https://ophthalmology.med.ubc.ca/patient-care/ophthalmic-photography/color-fundus-photography/>

Figure 1.10: Working principle of direct ophthalmoscope.

### 1.3.2 Optical Coherence Tomography

OCT is a non-invasive technique to study the superficial tissue layers of the retina. The primary focus is the macular region. The term OCT is called so because the light is used as a source instead of ultrasound. The word '*Tomography*' refers to the process of combining multiple image slices and obtaining a scanned image. The working principle is based on Michelson's interferometer, which is used to compute the time of flight. As depicted in Figure 1.11, a beam of light is split into two paths using a beam splitter. One beam is projected towards the patient's eye, and another beam is directed towards the reference arm. These two beams are reflected and collected at the detector with a

pinhole exit point. Further, based on the time taken to receive backscattered lights at the detector, the signal is processed, and the image is reconstructed. White intensity is assigned for the values where the backscattered light is nonzero, and black is assigned elsewhere. The beam of light is rotated in all directions, and thus, the intersection of backscattered beams produces the image of the object. Since the reconstructed image will be blurred, it is converted to the frequency domain and then filtered to obtain a sharp image of retinal layers.

There are different types of OCT imaging, namely, Time Domain, Frequency Domain (Swept-source and Spatial domain), and Phase Domain (Fujimoto et al. (2000); Yung et al. (1999)). Recent advancements include High Definition OCT and OCT-A (Angiography) variants (Malvika and Nadia (2019)). However, the fundamental operating principle (Michelson’s interferometer) remains the same, in all variants of the device. Since the two waves (reflected from patient’s eye and the reference arm) are coherent, fringes are formed due to constructive and destructive interference, resulting in a speckled appearance. Hence denoising these images is a preliminary step to restore the degraded images. The abnormalities such as drusen depositions, swelling in the

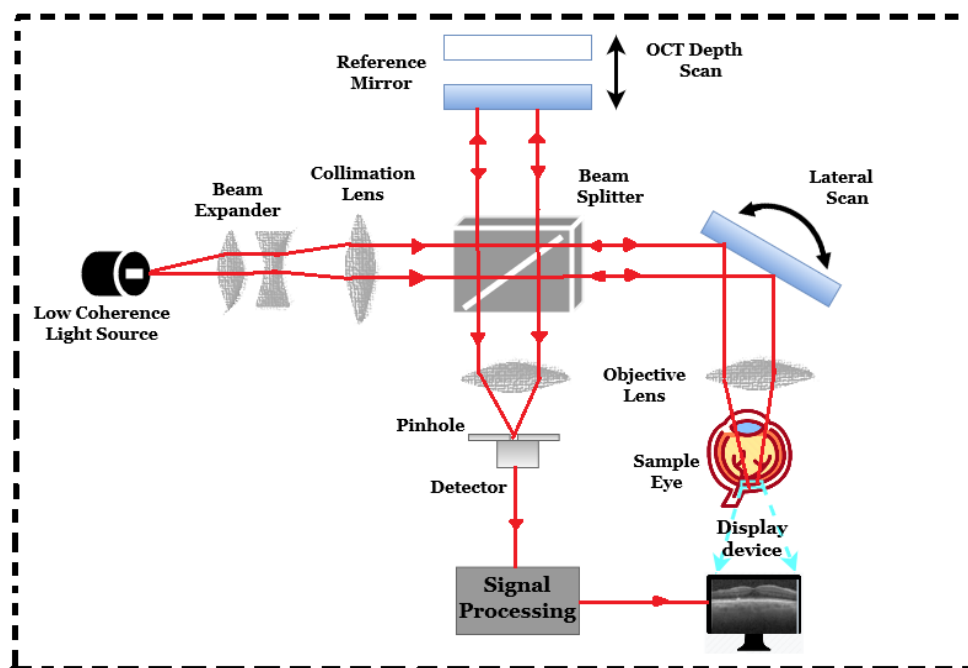


Figure 1.11: Working principle of OCT, (Modified and adapted from source: Hassan et al. (2019)).



retinal layers, growth of abnormal vessels, and fluid-filled cysts are located using OCT images as shown in Figure 1.12. The disorders diagnosed include AMD, Central Serous Chorioretinopathy (CSC) and DME.

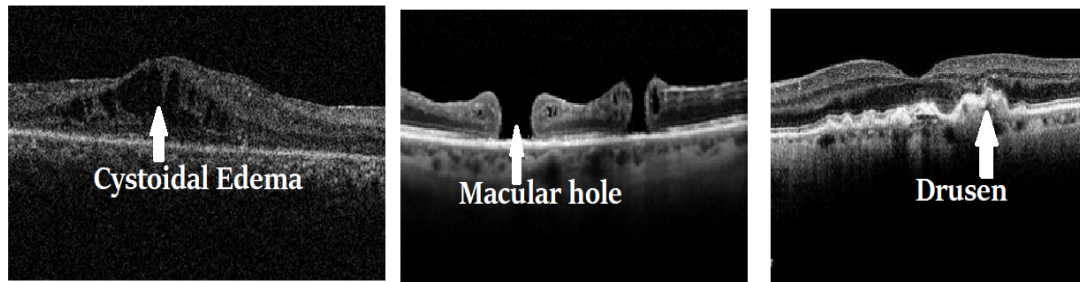


Image source: <https://imagebank.asrs.org>

Figure 1.12: Abnormalities visible on OCT images.

## 1.4 THE NEED FOR NOVEL PREPROCESSING ALGORITHMS

Overall, the two modalities (CFP and OCT) differ in terms of acquiring procedure, operating principle, structures observed, field of view, and degradation. The medical equipment utilized (such as OCT) possess limited features with respect to image quality enhancement. Certain artifacts are inevitable. Poor external room lighting, blurred image due to head movements and eye blinks are some of the instances that degrades the image quality. Plethora of existing image quality restoration algorithms blindly restore the image quality. Increasing the brightness and contrast of the image does not guarantee that the edges and other details of the image are preserved. Similarly, using traditional filtering techniques to eliminate the inherent speckle is inadequate. As ophthalmologists rely on multiple modalities to conclusively diagnose the retinal disorder, it is essential to integrate appropriate preprocessing techniques (adaptive) for these modalities. The major aim of the proposed thesis work is to implement such an adaptive preprocessing algorithm that outperforms existing techniques in terms of image restoration and edge preservation.

## 1.5 THE NEED FOR AUTOMATED ANALYSIS

WHO reports that the number of ophthalmologists available worldwide is inadequate to treat existing record of blindness. In the year 2015, 2,32,866 ophthalmologists were available for service among 194 countries (Resnikoff et al. (2019)). Considering more than one billion population of India, merely 15,000 ophthalmologists and about 45,000 optometrists were serving the country by the year 2019, against the required 1,25,000 optometrists, according to a guest blog page (Sandeep (2019)). Therefore, the primary focus of researchers has been to exploit the latest technologies like deep learning algorithms and offer cloud-enabled, cost-effective, smart, personalized solutions to diagnose eye disorders in such a way that it can reach people in remote locations and also expedite the overall diagnosis process.

There are some user-friendly, cost-effective devices available in the market, which the common man can use to capture the retinal fundus images and send the reports to the nearest ophthalmologist. Some examples for these include D-EYE ([https://www.d-eyecare.com/en\\_IN/product](https://www.d-eyecare.com/en_IN/product)) found by a company in Italy costing approximately 400 USD, PEEK ([https://www.peekvision.org/en\\_GB/peek-solutions/peek-retina](https://www.peekvision.org/en_GB/peek-solutions/peek-retina)) for people in Kenya, as shown in Figure 1.13. Biomedical engineers at Duke University found an innovative solution by developing a low-cost, portable OCT scanner that promises to bring the vision-saving technology to under-served regions throughout the world, thus preventing blindness. For the Indian market, a fundus camera set up that can be integrated into a smartphone is developed by researchers of IIT Delhi - which approximately costs around 50,000 INR, including the ophthalmoscope (Paul and Kumar (2015)). Furthermore, in Karnataka, a start-up company named 'Artelus' has been successful to screen thousands of patients with their innovative doctor-friendly device and applications (with ophthalmologists level of accuracy) Haloi (2018). A comparative analysis of some of the smartphone-based fundus cameras, including the cost and latest features incorporated is given by Mir et al. (2020). An insight of challenges in these AI-enabled devices, such as lower image resolution, poor image quality, lesser field of view, etc are also highlighted in their paper.



Figure 1.13: AI-enabled products.

Often, people may not visit ophthalmologists until they suffer from poor visual acuity. This negligence might lead to severe damage of the eye. The release of retinal images in the public domain in the form of grand challenges imparts the need to develop an automatic retinal disorder detection system. In other words, though the technology and the imaging equipment are upgraded in today's digital world, the imaging devices are not completely automated. So far, there is no indigenous clinically validated application that can be used by end-users to check whether there is an urgent need to refer an ophthalmologist. The primary implication from the above discussion is that there is a need to address preventable blindness. Incorporating automatic detection of certain eye disorders can assist ophthalmologists in treating patients with visual impairments proactively. There is still some room for further research.

## 1.6 MOTIVATION

The health of an individual is reflected in the eyes. At present, the clinical findings are manually examined and diagnosed by medical experts. In order to expedite the diagnosis procedure, developing an AI-enabled application that can be integrated into the image acquiring device to detect multiple disorders becomes the need of the hour. Though several researchers have proposed various deep learning models for automatic detection of retinal disorders, there is no end-to-end system to identify the disorders using multiple imaging modalities. The high impact of retinal disorders and the illustrious benefits of deep learning is the motivation to explore this domain.

## 1.7 OBJECTIVES

Medical images are severely degraded during the image acquisition procedure. It is necessary to analyse the noise distribution in the retinal images acquired from multiple repositories and appropriately propose an adaptive restoration algorithm. It is essential to validate the results qualitatively and quantitatively using metrics such as Peak Signal to Noise Ratio (PSNR), and Structural Similarity (SSIM). Further the preprocessed images are used to train a semi-supervised GAN architecture that can perform simultaneous segmentation and classification using retinal images. The performance of the proposed models is quantitatively assessed by comparing it with transfer learning models<sup>1</sup> and other relevant deep learning architectures. The segmentation and classification are assessed on heterogeneous dataset using metrics such as dice coefficient, sensitivity, and confusion metrics, and loss plots. Collectively, the research work acts as a baseline for the multi-modal retinal image analysis to automatically detect retinal disorders. The overall objectives of the thesis are formulated below:

1. To devise an image restoration algorithm that can enhance fundus images and denoise OCT images based on the probability mass distribution of the data.
2. To design an appropriate GAN model that can classify fundus and OCT images into normal, glaucoma, AMD, and DME.
3. To assess the performance of the proposed model with state-of-the-art deep learning models in terms of classification metrics such as accuracy, sensitivity, specificity, Area under Region of Characteristic Curve (AUC), and F1-score.

The above-mentioned objectives are met using the overall workflow as shown in Figure 1.14. Accordingly, given any one of the two imaging modalities (CFP or OCT), the proposed work will enhance and denoise the input image in preprocessing stage, perform segmentation, simultaneously classify the input images into normal and abnormal conditions using the deep learning model. State-of-the-art Convolutional Neural Net-

---

<sup>1</sup>Transfer learning is a special technique of deep learning, where a pretrained model is reused for solving a similar, but a different problem. The model is originally trained on a large set of images. Further, it is fine-tuned using relatively smaller number of training images. It saves training time.

work (CNN) models are used to compare the performance of the proposed model in terms of classification metrics using multiple datasets.

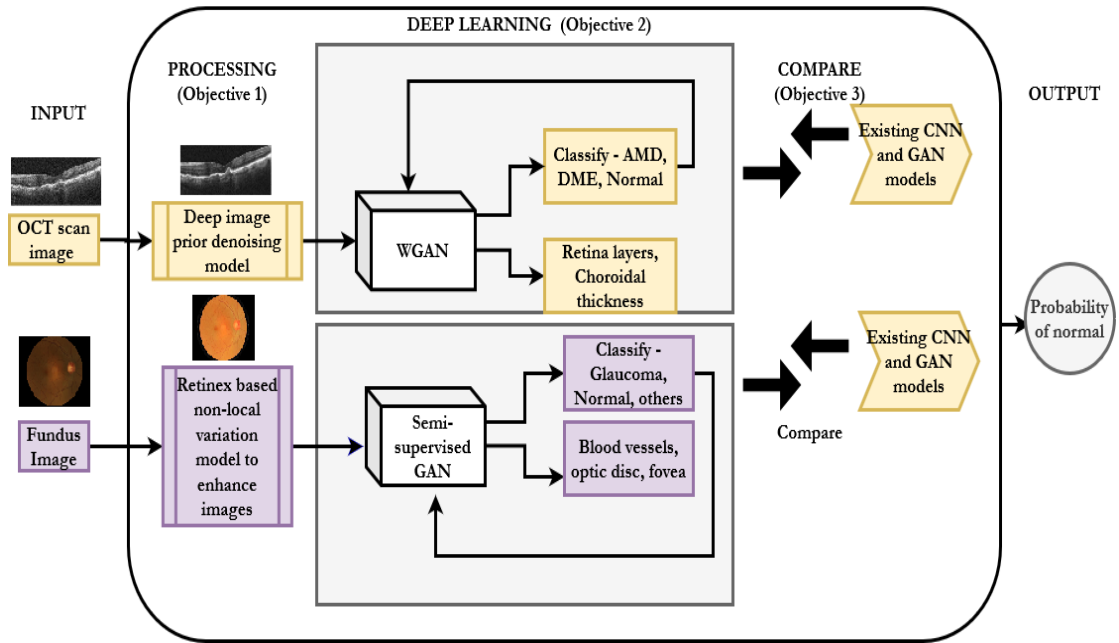


Figure 1.14: Overall workflow.

## 1.8 CONTRIBUTION OF THE THESIS

Considering the above mentioned factors, the thesis contributes to develop an end-to-end AI-enabled retinal disorder detection system. The research work aims to segment prominent regions from retinal images and identify the probability of having an eye disorder using deep learning architecture combined with image processing techniques. Two different modalities, namely CFP and OCT are chosen due to the availability of large datasets in the public domain. The retinal disorders studied include glaucoma and AMD, as these are the widespread retinal disorders affecting people of age group above 60 years, according to the statistics depicted in Figure 1.3. Acquiring images from multiple sources has a major setback. These repositories differ significantly in terms of image quality, resolution, and the number of images. Hence enhancing the image quality is of utmost importance, which is handled in the proposed work using Non-local Total Generalized Variational Retinex (NLTVGR) and Deep Image Prior (DIP) models. The GAN model used for classification in the research work is compared with

state-of-the-art models using image quality metrics and classification metrics.

The two imaging modalities (CFP and OCT), scan different parts of the eyes. These imaging modalities have been analyzed in-depth in many research works lately. Though these images are analyzed independently for ophthalmic disorder detection and analysis systems, a combined effort (by using both modalities) has not been explored to a deserving level. Since these images give different aspects of information about human eye, a combined analysis can reveal many disorders that are difficult to analyze otherwise. Deep learning models generate their own feature maps and maps the input to its corresponding output. Although some image acquiring devices (like ophthalmoscope and OCT) are equipped with retinal layer thickness profile generation, there is still room to incorporate better image enhancement algorithms. Inspired by AI-enabled smartphone integrated applications like D-Eye, PEEK Retina, Eye Art etc, that are capable of grading DR, our aim is to develop an application that can enhance and segment retinal structures on multi-modal images. Further, the application would classify the normal and abnormal retinal images. The major contributions of the thesis are:

- A novel perceptually inspired NLTGVR framework is proposed to enhance fundus images and address the illumination inhomogeneity.
- A novel non-local DIP model is proposed to despeckle severely degraded OCT images under the assumption that the speckle follows Gamma distribution.
- The variants of GAN model - semi-supervised and Wasserstein GAN are trained to automatically classify the retinal disorders using heterogeneous datasets.

**Choice of GAN:** GANs are adversarial in nature. It was introduced and predominantly used to generate realistic fake images. Particularly, it is beneficial to generate synthetic medical images, as it is a daunting task to obtain massive medical images for experimental purposes. However, the adversarial nature is explored in supervised learning aspect in recent years. GANs are complex and burdensome to train due to the adversarial nature. Thus, the process of training a GAN is exploited to perform multiple tasks such as classification along side the generation of fake images. The variants of GANs including semi-supervised GAN have demonstrated tremendous improvement

in accuracy compared to the traditional convolutional neural networks in recent years Haque (2020). In general, a CNN is employed for classification and a U-net model is recommended for segmentation of images. GANs comprises both CNN and U-net architecture within the generator and discriminator modules. Thus, it is possible to alter the loss functions and achieve multitasking (classification, segmentation, and fake image generation) using the same GAN architecture and partially available labeled data. The ability to learn from limited data, reliability to perform multitasking are the primary reasons to select GAN among other deep learning architectures like CNN, for the proposed research work.

The above-mentioned contributions further aids to address the following research questions.

1. Does the retinex model aid to enhance the features of the fundus images irrespective of capturing device?
2. Will the deep image prior model minimize the influence of speckle inherent in OCT images?
3. Can the GANs be effective or give stable performance in the segmentation and classification of heterogeneous retinal images acquired from multiple locations?
4. Is there any scope for implementing automatic multi-modal retinal disorder detection using deep learning network architecture?

The novelty of this research work is:

1. To analyze the data distribution of retinal images acquired from multiple resources and then propose an adaptive restoration model.
2. To train the GAN model in order to perform both segmentation and classification using limited availability of ground-truth images.
3. The concept of Gamma loss integrated to deep image prior model reduces the effect of speckle in OCT images.

The significance of the proposed work is limited to the development of an automatic retinal disorder detection system using a deep learning model. It will be useful for ophthalmologists to identify various structures from different modalities, which can be further quantified using appropriate annotation tools. The automatic prediction of the retinal condition expedites the retinal examination at the primary or secondary eye care hospitals.

## **1.9 ORGANIZATION OF THE THESIS**

Retinal imaging and analysis has been the limelight of research since the inception of deep learning. Relevant works in this domain are summarized in Chapter 2. Considering the rapid evolution of technologies and transformations in the retinal examining modalities, existing contributions since last five years are elaborated. The primary focus has been to identify the research gaps in CFP and OCT image analysis. Though several other variants of modalities exist, these modality images are available in public domain in large number. Therefore, the focus of Chapter 2 is to obtain a comparative analysis of different preprocessing strategies performed to enhance and restore the retinal images. Some of the review works serve as a bedrock to strengthen the concepts of retinal image analysis.

Chapter 3 is reserved for the preprocessing methods to restore the degraded retinal images. Histogram is used to assess the quality of the acquired retinal images. A perceptually inspired NLTGVR is proposed to enhance the contrast of the RGB fundus images. Using the proposed algorithm, the edges and details are well preserved in the retinal images. Furthermore realizing the fact that the ground-truth clean images of OCT is unavailable in reality, a DIP approach is proposed to obliterate the speckle. A mathematical approach is derived using Bayesian estimation to handle the multiplicative noise distribution.

Chapter 4 discusses the automatic diagnosis of disorders such as AMD, DME, and Glaucoma using CFP and OCT repositories. The deep learning models namely semi-supervised GAN, and Wasserstein GAN are employed to sequentially train the segmentation and classification tasks using images acquired from different modalities. The



transfer learning approach is a simple and effective method for classification of retinal disorders. Even though, these models offer greater reliability and higher range of accuracy in classification, the performance is stable only when the test images are processed and belong to the same distribution set as that of the training images. The performance of the proposed GAN with transfer learning models is also discussed.

Finally, the conclusion is presented in chapter 5. It outlines how this research work augments existing contributions and discusses the scope of future works. Additionally, a conceptual model is designed for automatic classification of Glaucoma from multi-modal images. The conclusion chapter provides a brief summary of how the objectives set at the beginning of the research are met using novel algorithms.



## **Chapter 2**

# **A COMPREHENSIVE REVIEW OF RETINAL IMAGE ANALYSIS**

This chapter introduces the state-of-the-art research on retinal image analysis restricting the study to the domain of our interest. Figure 2.1 captures the topics considered in this research, which is self-explanatory. A Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) approach (Page et al. (2021)) is followed to screen and identify relevant articles. The role of deep learning, retinal imaging modalities, information about public repositories, and the scope of research works in retinal imaging are discussed through existing review papers in section 2.1. Selected papers related to restoration of degraded retinal images, segmentation of prominent features, and classification of the retinal disorders into normal or abnormal categories are discussed in further sections.

### **2.1 STATE-OF-THE-ART REVIEW PAPERS**

The scope and application of artificial intelligence and deep learning in ophthalmology is discussed by Ting et al. (2019). For a juvenile researcher, it offers ample information about fundus and OCT images, types of diseases that are of utmost importance, and potential challenges. The application of GAN in medical imaging is elaborately discussed by Yi et al. (2019). The authors compared various modified GAN architectures that can be suitable for segmentation and classification purposes in medical imaging. Despite the massive contribution in this domain, limited availability of retinal images or ground-truth for segmentation, computational restrictions, non-standardized image acquisition

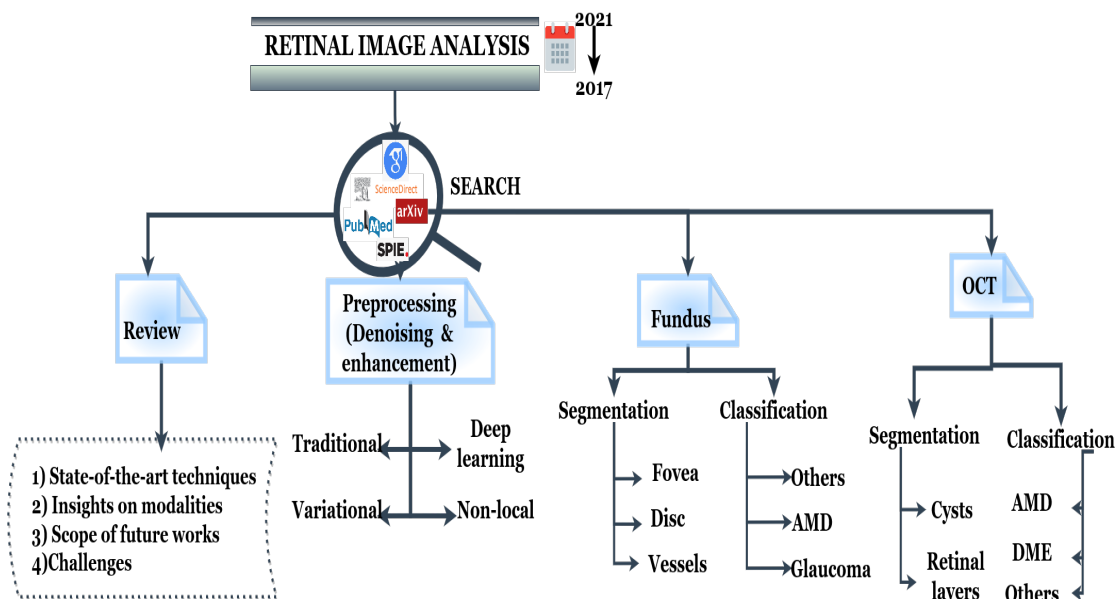


Figure 2.1: Overview of existing literature on retinal image analysis.

techniques, and inconsistency in the performance metrics prevail as the challenges to the researchers (Yanagihara et al. (2020)). Amidst these prevailing challenges, several attempts are in progress to automate the retinal disorder detection primarily using two modalities: CFP and OCT.

An overview of retinal disorders like DR, Glaucoma, and AMD has been provided by Abràmoff et al. (2010) and independently by Hossein et al. (2016). They explained the benefits of retinal vessel extraction, lesion detection, optic cup disc segmentation, Symptomatic Exudate-associated Derangements (SEAD) segmentation from CFP images, and the applications of multi-modal imaging. They highlighted the future scope of these imaging techniques and state the need to obtain a low-cost, highly efficient imaging system. Various datasets available in public domain that can be accessed to study eye diseases like Glaucoma, DR and AMD are outlined in recent literature (Sengupta et al. (2018)). They compared the state-of-the-art methodologies that classify these diseases and discussed the feasible future works in ophthalmology. Various methods to assess the quality of fundus images is surveyed by Raj et al. (2019). They state that the CFP images are severely degraded due to illumination inhomogeneity. So, contrast enhancement is often necessary to train the automatic segmentation or classification models.

The retinal disorders like, AMD, DME, and CSC are reviewed by El-Baz (2018). They illustrated the sample OCT scans for each of these disorders. They explained the need to monitor the retinal layer thickness for early detection of vision disorders. A similar study was done by Garcia-Layana et al. (2017). Here it is stated that the retinal layers in OCT images possess different reflectivity patterns that can help to diagnose eye disorders by spotting the fluid depositions in the retinal layers. Structures such as retinal layers, cyst regions, edema, and choroidal thickness are assessed from the OCT images. OCT images are subjected to speckle due to a series of constructive and destructive interference (Schmitt et al. (1999)).

The speckle is multiplicative in nature, as illustrated in existing literature (Aubert and Aujol (2008)). Further, contemporary works suggest that the OCT-A is highly susceptible to bulk motion noise (Spaide et al. (2015)). This is challenging to remove due to the fact that the incident light and the blood vessel distribution on the image are perpendicular to each other. Certain other artifacts affecting OCT images were also listed by Chhablani et al. (2014). However, they stated that the clinically significant artifact, such as speckle noise, will hinder the segmentation and automatic classification mechanisms. As clean data can highly influence the performance of deep learning models or automated analysis systems (Vikram (2020)), it is necessary to incorporate an image enhancement method appropriately.

Recall from Chapter 1, that CFP and OCT modalities differ significantly in terms of image acquisition procedure, image quality, features extracted, and disorders identified. However, both are degraded and suffer multiple artifacts that hinder automated analysis. Considering these factors, retinal image analysis can be broadly classified into three categories, namely: denoising and enhancement, segmentation, and classification. These tasks are carried out either using traditional image processing techniques or machine learning and deep learning methods. Therefore, the remaining content of this chapter is divided into the following sections. We first discuss various preprocessing models available to restore both CFP and OCT images in section 2.2. This is followed by the discussion of the latest works on classification of retinal disorders.

## **2.2 PREPROCESSING OF RETINAL IMAGES**

Denoising and Contrast enhancement are two different tasks. Contrast enhancement is a process of modifying the range of intensities used to represent the image, such that the prominent features are visible clearly. Denoising, on the other hand, refers to the technique of suppressing the variations (that makes the image noisy) in the image. Throughout the rest of the content in the thesis, the term image restoration refers to the process of transforming a degraded image into a visually appealing one either in terms of denoising, enhancement, or both of these tasks.

### **2.2.1 Traditional image restoration methods**

The effect of applying Contrast Limited Adaptive Histogram Equalization (CLAHE) on fundus images was assessed by Alwazzan et al. (2021). Accordingly, the green channel was extracted as it carries maximum information. The Weiner filter was applied before CLAHE to reduce any noise present in the green channel. The enhanced images were quantified using contrast improvement index and linear index of fuzziness. They showed that this method is effective to enhance the severely degraded CFP images. Despite the qualitative and quantitative analysis, the effect of pre-processing is unexplored for automatic segmentation and classification.

A similar attempt was done by Sonali et al. (2019). However, in their approach the denoising filters and CLAHE were applied on all the three channels (R,G, and B) independently, and the final fundus image was constructed by combining the enhanced channels. The quantitative metrics used to assess the quality of enhanced images include edge preserving index and SSIM. Though CLAHE is an efficient method to enhance degraded images, the major setback of this method is that it enhances the noise component and it is primarily applicable for grayscale images.

Denoising depends on certain prior assumptions about the noise distribution in the image. Statistical models were developed to analyze the noise distribution of OCT images Jorjandi et al. (2017). Compared to the fundus images, the OCT images are highly noisy, and the speckle noise is removed using traditional methods such as homomorphic Weiner filter (Franceschetti et al. (1995)), Lee filter (Lee (1981)), Kuan filter

(Kuan et al. (1985)), and other filtering methods or transformation methods. Some other methods include anisotropic diffusion (Weickert (1998)) and variational models (Duan et al. (2016)).

Various denoising methods of OCT images are summarized by Muxingzi et al. (2017). Accordingly, the denoising methods can be broadly categorized into a single frame approach or a multi-frame approach. A single-frame approach can be further divided into spatial-domain or wavelet-domain. Variational methods, probabilistic approaches, and partial differential equation-based modeling are some examples of spatial-domain denoising. On the other hand, Gaussian Mixture Model, Block-matching and 3D filtering (BM3D), and dictionary-based learning are instances of denoising in a wavelet-domain. Considering the fact that several B-scans can be acquired for a particular region of interest, averaging multiple B-frames can be an approach to generate a cleaner version of the noisy OCT images, referred as a multi-frame approach. Another approach to solve speckle noise is to consider transforming the input to the log domain, which subsequently converts the multiplicative noise into an additive noise but fails categorically to capture the inherent features of the speckles.

Overall, a significant contribution is made to denoise and enhance retinal image quality. However, removal of inherent speckle noise in OCT images can be treated as an inverse problem, and more specifically, it is an ill-posed problem. A problem is referred to as a well-posed problem if it satisfies existence, uniqueness, and continuity conditions for its solution (Hadamard (1953)). Failing to meet any of these conditions makes it an ill-posed problem. Further, an inverse problem is related to computing the cause of event, using the effect. For instance, given a noisy input image, the objective is to find the cleaner image. The details regarding inverse and ill-posed problems are highlighted in Appendix A.1. Solving such ill-posed problems is challenging, and often regularization methods (Rudin et al. (1992)) are used.

Regularization approach is a process of solving an inverse problem by addition of regularization parameters. The variants include  $L_1, L_2$ , and Tikhonov regularization. Details of these frameworks are elaborated in Appendix A.1 and A.2. Variational methods including Total Variation (TV) approach are another category of regulariza-

tion methods. These methods are extensively used due to its ability to retain the edge information while smoothing the data. These models can be tuned to various noise distributions such as Gaussian, Gamma, Rayleigh, and Poisson. Unlike traditional denoising approach, despeckling is considered as a daunting task due to the multiplicative nature of the speckle. In a nutshell, formulating the image restoration framework specifically for despeckling and enhancement using precise statistical modeling is still an open research problem that can be addressed using variational frameworks.

### **2.2.2 Variational image restoration models**

The variational models (Vese and Guyader (2015)) capable of incorporating the nature of distortion in the design of the framework are observed to be more efficient in restoring the data, especially in data correlated environments. The idea of TV denoising is to consider the ill-posed problem as an energy minimization functional, with terms that constitute of data fidelity and regularity such that the denoising is achieved without leading to over-smoothing. A scalar positive value acts as a regularization parameter to monitor the smoothness levels. Certain other preliminaries of the variational frameworks is presented in Appendix A.2.

A plethora of variants are introduced in the literature for the TV regularization. Perceptually inspired models have shown a considerable improvement to the existing image restoration paradigms for ill-posed problems. For instance, retinex theory (Bertalmio et al. (2009)) based models were incorporated in a variational framework to handle various contrast degradation aspects of the input images (Liu et al. (2017)). A total variational functional based on block matching was defined by Shengjian et al. (2019). The TV retinex models perform linear approximations of the input data, eventually leading to the patch-like appearance of the resultant images (Rudin et al. (1992)). This was addressed effectively using Total Generalized Variation (TGV) models (Kim and Kudo (2020)). Further, to improve the restoration ability of the TGV models, non-local TGV models were proposed in the literature (Ranftl et al. (2014); Jidesh and Febin (2020)).



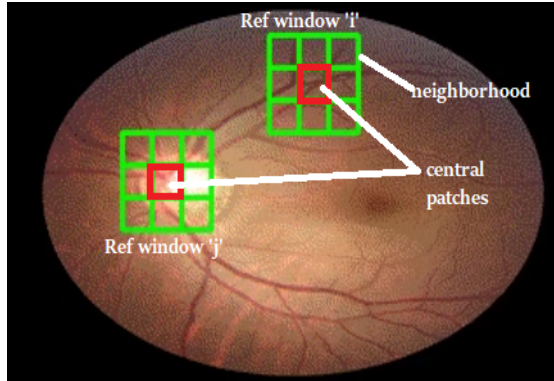


Figure 2.2: Demonstration of non-local similarity computation.

### 2.2.3 Non-local image restoration

The variational models are subjected to piecewise approximations and it is local in nature. This tends to ignore the detailed features or textures in the data. Non-local variational models are motivated by the initial works of Buades et al. (2005) that is referred as non-local means filtering approach. Traditional filtering techniques merely perform averaging of the pixel neighborhood. Unlike these filtering technique, the non-local means, computes a weighted sum considering all the pixels in the image. The non-local variational image restoration framework is widely used to despeckle medical images. The idea of non-local gradient computation ( $\nabla_{NL}$ ) is to break down the entire image of size  $width \times height$  into smaller windows. Within a window, the weight associated with every central patch is computed based on the intensities of similar patches in its neighborhood. This process is repeated for every pixel in the entire image as shown in Figure 2.2. Accordingly, two reference windows ( $i$  and  $j$ ) are depicted around the central patches given by red colored box. The green patches represent the neighborhood. The size of the patch and the neighborhood considered can be varied intuitively. It can be observed that increasing the neighborhood (window size), increases the influence of dissimilar patches, increasing the computational burden. The benefit of incorporating non-local computation is that the similar patches add more weight and contribute to enhancing the region of interest. Since the neighborhood strongly influences the given region of interest, non-local restoration methods preserve the edge information without blindly over-smoothing the image globally. The computational burden can be greatly minimized using parallel programming architectures or other fast numerical methods.

Mathematically, the non-local gradient  $\nabla_{NL}U(x,y)$  for a pair of points  $(x,y)$  is defined as the vector of all discrete derivatives:

$$\nabla_{NL}U(x,y) := (U(y) - U(x))\sqrt{W(x,y)}, \text{ for } y \in \Omega. \quad (2.1)$$

The norm of the non-local gradient of  $U$  at  $x$  is defined as:

$$|\nabla_{NL}U|(x) := \sqrt{\int_{\Omega} (U(y) - U(x))^2 W(x,y) dy}. \quad (2.2)$$

where the weight function  $W(x,y)$  gives the similarity between  $x$  and  $y$  computed using:

$$W(x,y) = \frac{1}{T_x} \int_{\Omega} e^{-\frac{G_{\sigma}|q(N_x)-q(N_y)|^2}{h^2}} dy. \quad (2.3)$$

where  $G_{\sigma}$  is Gaussian kernel,  $h$  is filtering parameter,  $T_x$  is normalizing constant,  $N_x$  and  $N_y$  denote the neighborhood centered at  $x,y$ , respectively.

The non-local variational framework is proposed by Gilboa and Osher (2009). A patch-based algorithm was proposed using Gaussianization transform for OCT denoising and enhancement by Zahra and Hossein (2017). The major limitation of non-local variational approach in Gilboa and Osher (2009) is that it considers the noise is additive white Gaussian in nature. Recently, Anoop et al. (2019) portrayed various literature specifically focusing on speckle noise formulated as a Gaussian, a Rayleigh, or a Gamma distribution. They render qualitative and quantitative results of denoising OCT images using TGV, anisotropic diffusion, iterative adaptive unbiased filter, wavelet transform, and non-local means filtering techniques. Contrast degradation and illumination inhomogeneity are not addressed in their work.

An optimized non-local linear minimum mean square error-based approach is proposed by Rahimizadeh et al. (2020). They suggest that the traditional methods fail to perform well when the image intensity varies in a particular region of the image. This issue is addressed by them using non-local methods. In all variational models of restoration, it is necessary to have some prior knowledge about the noise distribution or appropriate constraints that have to be introduced to monitor the data fidelity. Deep learning methods, on the other hand, rule out the human intervention of feature extraction, by automatically identifying the feature maps for a given problem.

## 2.2.4 Deep learning architectures

Deep learning has been a game-changer in various fields that have aided in resolving numerous computationally expensive problems. A wide range of deep learning architectures have been devised in the last few decades to achieve realistic results of image restoration. In general, all these architectures involve a series of convolution, down-sampling, and up-sampling, and nonlinear activation function operations applied to the input images. Overall, every deep learning model aims to solve the optimization problem by formulating a loss function. Since there are no straightforward techniques to design a convolutional neural network, the performance greatly depends on the training dataset, hyperparameters, and loss functions. Therefore, several combinations are experimentally verified towards restoring CFP and OCT images.

A deep learning network is proposed by Shen et al. (2021), to enhance the quality of fundus images. However, it relies on the availability of massive ground-truth images. To achieve this, they have considered synthetically degrading the image quality. A modified residual densely connected U-net model is used to enhance the degraded CFP images by Raj et al. (2022). The fundus images are synthetically degraded and an ensemble model is constructed using several U-net models where the output of each U-net is concatenated and fed to a final convolutional layer. Multiple U-net modules are used to restore the illumination inhomogeneity, and haze affected images. Mean absolute error and SSIM is used as loss function.

Prioritizing the edge preservation in OCT images, a modified PCANET approach is proposed by Yu et al. (2018). Accordingly, the deep learning network is trained to extract features from preprocessed input images. The images are filtered using an optimized Bayesian non-local means filter in the preprocessing stage. The features extracted from PCANET are further fed to another PCANET to reconstruct a cleaner image using a similarity weight matrix. Quantitative results prove that their method successfully despeckles the noisy images while preserving edge details.

Brunet model is trained to enhance the OCT images acquired from multiple devices by Apostolopoulos et al. (2020). Although their method performs denoising and enhancement of the OCT images, it does not take into consideration any prior information

about the noise distribution. Obtaining ground-truth images is a major challenge for medical datasets. In this regard, a multi-frame averaging technique has proved to be exceptional.

A content loss that can be added to adversarial loss is proposed by Zhao et al. (2020). Accordingly, the content loss denotes the pixel-wise mean square error. The modified GAN is coined as speckle modulation GAN and experimental analysis is performed on OCT images.

Recently, another GAN-based denoising was performed in Sourya et al. (2021). Here the speckle is assumed to follow multiplicative Gamma noise. The retinal layer information is preserved using Wavelet transform integrated to the Generator module. The requirement of ground-truth images to train the model is the setback of this method. Recent advancements in deep learning paved way to elevate the image denoising process with the availability of massive computational power.

As discussed earlier, image denoising using a deep learning approach usually requires an enormous amount of input training data. This notion has been wiped out by a novel approach, described as Deep Image Prior (Ulyanov et al. (2020)). Lately, the concept of the deep image prior was explored on a progressively growing Generative Adversarial Network (GAN) (Refer Sidorenko (2018); Pan et al. (2020)). In addition, an automatic training halt mode is defined in their work. This is formulated using SSIM. The overall idea is to grow the network deeper if the SSIM has improved and fall back to the previous depth, otherwise. Furthermore, the deep image prior model is extended to non-local domain to include an explicit regularizer (Fan et al. (2020)). Some advantages and disadvantages of existing denoising and enhancement methods are captured in Table 2.1.

## **2.3 FUNDUS IMAGE ANALYSIS**

As mentioned earlier in Chapter 1, various structures are segmented from the fundus images. Though automatic segmentation of abnormalities such as exudates and hemorrhages is vital, the deep-learning approach necessitates enormous ground-truth images, which are unavailable in the public domain. Thus, the automatic segmentation of blood vessels and optic disc region is the major highlight in the literature, due to the avail-

Table 2.1: Comparison of existing OCT image restoration methods (\*D: Denoising, E: Enhancement, CNN: Convolutional Neural Network).

Author	Task	Method	Advantages	Limitations
Mehdizadeh et al. (2021)	D	DnCNN is used to despeckle OCT images where the CNN consists of pretrained VGG models.	Blurring effect on the images are addressed using the deep feature loss. This is quantified using perceptual sharpness index.	The higher order variation is ignored in their work to preserve the details of the image and avoid piecewise approximations.
Guo et al. (2021)	D	Non-local GAN is designed that learns to denoise in unsupervised manner.	Model is tested on multiple datasets. The non-local means layer included in the generator preserves the edge information.	Contrast enhancement is not considered. Unsupervised deep learning models are computationally expensive.
Song et al. (2019)	D	Various filtering methods are assessed including median filtering, wavelet threshold method, Wiener filtering and anisotropic diffusion.	Compared to other methods, filtering is computationally simple method to restore the OCT images. Anisotropic diffusion method was found to perform better than the other filtering methods.	Speckle is assumed to follow Rayleigh distribution which may not be applicable for most of the methods. Filtering in spatial domain is local in nature and result in over-smoothing.
Lou et al. (2021)	E	The method consists of noise suppression, edge detection and applying the sigmoid energy equation on the fused image.	The edges in the OCT image are sharpened using the sigmoid energy equation and the algorithm is tested on multiple datasets.	The reflectance and illumination terms are treated in same pattern which limits the contrast enhancement ability to local domain.
Tao et al. (2021)	E	Retinex model is considered to enhance the OCT images under the assumption that the data distribution follows Gaussian law.	The retinex model has significantly increased the contrast of severely degraded images and visually pleasing results are obtained to segment the edema region simultaneously.	The data distribution is assumed to be Gaussian and the benefits of non-local restoration and higher-order regularization terms are not considered.
Liu et al. (2018)	E	Shock filtering method to enhance the image under the assumption that the data distribution in image follows Gamma law.	Similar to non-local means method, a self-similarity measure is used to control the effect of smoothing as well as retain sharpening effects.	The robustness and scalability of model to restore severely degraded OCT images is uncertain.
Huang et al. (2020)	D & E	Anisotropic TV is used to despeckle and retinex model is applied for contrast enhancement of OCT image in logarithmic domain.	Method addresses the adaptive enhancement considering the fact that the background region of OCT image has noise dominant region and recover the speckle noise map.	The TV model excludes the speckle following Gamma distribution.
Xu et al. (2020)	D & E	OCTNet deep learning model is used to obtain noise free, enhanced, high quality images.	The model performs well even if the background region is dominant near the retinal layers.	The performance of the model greatly depends on the quality of ground truth images used for training.

ability of ground-truth images. In this direction, the comparative analysis of the deep learning implementations available for CFP is shown in Table 2.2. Some of the works

utilize GANs for segmentation, generation or classification of visual disorders as given in Table 2.3.

Table 2.2: Summary of existing works on fundus image analysis.

Sl. No.	Paper details	Datasets	Task	Description	Results	Advantages	Disadvantages
1	Meyer et al. (2018)	REFUGE	Localization of fovea and optic disc.	Pixel-wise distance is computed and a regression based approach is used with U-net architecture.	Median of Euclidean distance: 0.94 to 1.12	Joint detection of optic disc and fovea is favourable for multiple disorder detection.	Optic cup to disc ratio is not measured and blood vessels are not segmented.
2	Lam et al. (2018)	Kaggle, eOphta	Retinal Lesion Detection.	Sliding window based patches were extracted and classification was performed using standard convolution neural networks like VGG, GoogleNet, Resnet and Inception	Accuracy: 0.74 to 0.96	Multiple datasets are analysed.	Only microaneurysms are detected. It could be extended to detect multiple indicators.
3	Leopold et al. (2019)	DRIVE, STARE and CHASE-DB1	Retinal vessel segmentation.	Fully residual autoencoder with gated residual streams termed as PixelBNN.	Accuracy: 0.9106, Sensitivity: 0.69, Specificity: 0.95, AUC: 0.82	Histogram enhancement is performed as a preprocessing step using CLAHE. A parallel input stream is used along with skip connections, batch normalization in the architecture.	Although good results are obtained, the dataset considered consists of minimal set of images.
4	Yan et al. (2018)	DRIVE, STARE, CHASE-DB	Retinal vessel segmentation.	The model consists of three stages. Thick segmenter, thin segmenter and the fusion part.	Accuracy: 0.95 to 0.97 Sensitivity: 0.76, Specificity: 0.95	Both thick and thin vessels are segmented separately. This increases the visibility of retinal vessels.	The performance on poor quality images is unknown.

### 2.3.1 Classification

A comprehensive study on recent works in fundus image analysis is given in Sarki et al. (2020). It highlights the existing limitations in deep learning model-based automatic screening systems. A generic eye disease diagnosis mechanism and automatic conversion method of acquired data into a structured form that can simplify the patient case analysis are explained by Malik et al. (2019). It emphasizes the need for automatic screening of retinal disorders. The feature of multi-class and multi-label is utilized to

Table 2.3: Summary of existing works on fundus image analysis using GANs.

Sl. No.	Paper details	Datasets	Task	Description	Results	Advantages	Disadvantages
1	Son et al. (2017)	DRIVE, and STARE	Fundus vessel segmentation	U net based generator GAN model	Dice Coefficient: 0.829 for DRIVE and .834 for STARE	The authors performed pixel based, patch based and image based discrimination.	Results fail to detect thin vessels from fundus images. DRIVE and STARE are small datasets.
2	Bellemo et al. (2018)	DRIVE, STARE, MESSIDOR	Generate fundus images	A review of available GANs used to generate fundus images	-	Lack of large datasets necessitates generation of synthetic images. This paper compares available GAN architectures for generating fundus images.	The detailed biomarkers generated as fake images might not be up to the clinical standard.
3	Lecouat et al. (2018)	IDRiD	Diabetic Retinopathy	Patch based semi-supervised GAN enables visualization of prominent features	AUC: 75.8 to 84.5	Semi-supervised GANs are more suitable when variations are present in the dataset.	Authors used small dataset of 149 images for training. They do not discuss about the accuracy of the model when it is used for classification.

automate retinal disorder detection using Convolutional Neural Networks (CNN) (Gour and Khanna (2021)). Here, the authors experimented with two methods. The first one is to consider left and right fundus images in parallel, while the second method is to concatenate both images and then feed the CNN. The pre-trained models like Resnet (He et al. (2016)), Inceptionv3 (Szegedy et al. (2016)), Mobilenet, VGG-16 (Simonyan and Zisserman (2015)) are used to analyze the performance of the classifier with the different optimizers, namely, Adam and Stochastic Gradient Descent. They concluded that the VGG16 model achieves higher performance than the other models.

The Resnet model is used extensively as a backbone network in a similar approach and spatial correlation module (Islam et al. (2019); Li et al. (2020a)). The output from this module is then either concatenated, multiplied, or added to analyse the effect of merging the left and right fundus images. Their implementations illustrate Resnet 101 outperformed Resnet 18, Resnet 34, and Resnet 50 with the concatenate operation among left and right fundus images. The multiple labels associated with some fundus images are retained in their method. Yet another paper demonstrates the results of transfer learning on the Ocular Disease Recognition (ODIR) dataset using VGG-

16, Resnet-18, Resnet-50, ResNext-50, SE-Resnet-50, SE-ResNext-50, Inceptionv4, Densenet (Huang et al. (2017)), and CaffeNet (Li et al. (2021)). In their method, the left and right fundus images are considered independently. The implementation results in their paper show that the Inception v4 outperforms all other models. The above-discussed papers primarily focus on analyzing the effect of using both left and right fundus images to automatically detect retinal disorders. Most of the papers compare and contrast the performance of various pre-trained models without incorporating the data cleaning procedures or data preprocessing techniques. To further highlight the quantitative results, Densenet, Inception, Resnet, MobileNet (Howard et al. (2017)), Efficientnet (Tan and Le (2019)), VGG, and Xception (Chollet (2017)) models were engaged in training the classifier (Wang et al. (2020a)). Here, each fundus image was distinctly passed to a grayscale histogram equalization module and a color histogram equalization module before feeding the classifier. The results obtained show that Efficientnet can successfully learn a majority of the features and distinguish the retinal disorders.

A unique way of using Hessian-based fibermetric filter was proposed by Ram and Reyes-Aldasoro (2020). The filtered fundus images were then fed to a Convolutional Neural Network (CNN) to validate the accuracy of the classifier considering a various number of classes. The results presented in their paper included two classes (Normal and Cataract), three classes (Normal, Cataract, AMD), and four classes (Normal, Cataract, AMD, Myopia), respectively. However, they noted that the classifier underperformed when the number of categories was beyond 4. This is mainly due to the unbalanced number of images in the dataset. Similar analysis of various classes along with data augmentation on the fly was attempted by Jordi et al. (2019); Meller (2020).

## **2.4 OPTICAL COHERENCE TOMOGRAPHY IMAGE ANALYSIS**

A large and growing body of literature has described the role of deep learning for automatic detection of retinal disorders from OCT images in recent years. A semi-supervised deep learning model is proposed by Sedai et al. (2019), to extract the retinal layer from a partially labeled dataset. The deep learning model is comprised of a



teacher-student module designed using Dense-U-net architecture. The teacher module was trained using the Bayesian deep learning concept, and it controlled the learning of the student module. The preprocessing steps included data augmentation. This method is greatly beneficial when the segmentation labels are unavailable in large numbers. The downside of this method is that it does not involve denoising OCT images, and the number of trainable parameters is large due to the use of Dense-U-net architecture.

Deep Retina (Li et al. (2020b)) is another recent work on retinal layer segmentation. Here, an encoder-decoder architecture was tested with a focal loss function. The ground-truth images for segmentation were obtained using manual segmentation by experts. The encoder network involved a pre-trained Xception-65 model. Although this model provides a sensitivity of up to 92% on the public repository without any preprocessing module, there is a scope to further reduce the complexity of the model.

A Squeeze and Excitation U-Net (SEU-Net) deep learning model was implemented by Xuehua et al. (2021). Accordingly, the SEU-Net is an improvised version of U-net model following an encoder-decoder architecture. It was used in combination with the graph search method to segment 9 retinal layers. The model was trained using 10 3D-retinal data, and the performance was assessed using mean absolute error and root mean square error. Though the deep learning model performance was improved by incorporating a post-processing method, the performance of the model on severely affected retinal layers was not discussed.

### **2.4.1 Classification**

A plethora of convolutional neural network (CNN) models are designed for classification purposes (See Negi et al. (2020); Van Hulle et al. (2016); Sharma et al. (2018)). Transfer learning methods were applied to classify the OCT images into normal and abnormal categories using a private dataset (Refer Serener and Serte (2019); Wang and Wang (2019)). The classification accuracy of above 98% is achieved using transfer learning. However, the performance of transfer learning models greatly depends on the distribution of the training data set. Therefore, there is a need to have a greater resemblance between the quality of images used to obtain the pre-trained model as well as the testing dataset.

To further deal with the overfitting and biased decision of the pre-trained models, an ensemble model was proposed by Paul et al. (2020). Accordingly, VGG-16, Densenet, Inception-v3, and a custom designed convolutional neural network models are first fine-tuned to classify the OCT images. The second step includes merging the feature maps of the pretrained model and assigning a non-linear weight based on the performance of each model. The concatenated output is fed to a neural network with fully connected neurons. Though the classification metrics are in the acceptable range, combining several pre-trained models significantly increase the complexity of the model, thereby increasing the number of trainable parameters.

When a limited labeled dataset is available, semi-supervised GAN is appropriate to classify the OCT images. This was implemented by Das et al. (2020). AOCT-NET (Alqudah (2020)) and OctNET (Sunija et al. (2021)) are recent CNN models with simple architecture comprising a series of convolution and max-pool layers, capable of classifying the input OCT images into normal or abnormal using public repositories. The accuracy of up to 99% was achieved without any preprocessing module. Although the model depth is less than 6 layers, the performance of the model on severely degraded OCT images is uncertain as data denoising was not considered in their works. Overall, the majority of the works on OCT image analysis exhibit exemplar performance in terms of accuracy when trained for a single task. The comparative analysis of the deep learning implementations available for OCT image segmentation and classification is shown in Table 2.4.

## **2.5 RESEARCH GAPS**

Recent advancements in deep learning paved way to elevate retinal image restoration and classification with the availability of massive computational power. Nevertheless, the following limitations are identified from the related works discussed so far.

- A plethora of existing works on retinal image analysis tend to ignore preprocessing when using deep learning models. Though deep learning models give stable and reliable performance, when the training images are raw and heterogeneous, the performance of automatic classification and segmentation can be severely de-

Table 2.4: Summary of existing OCT image classifiers.

Sl. No.	Paper details	Datasets	Task	Description	Results	Advantages	Disadvantages
1	Girish et al. (2018)	OPTIMA cyst challenge	Segmentation of intraretinal cyst	The method comprises of preprocessing stage that includes image noise removal followed by segmentation stage. Fully convolutional network with U-net architecture is used for segmentation.	Dice coefficient: 0.71	The preprocessing technique incorporated equalize the OCT images across multiple vendors as fast non-local means method is used.	Multiple disorder classification is not done. Penalizing False positives result in smaller cysts getting missed.
2	Asgari et al. (2019)	Private hospital, Austria	Segmentation of drusen, RPE, Bruch's Membrane region, and the background region in OCT scans	Modified U-net architecture with spatial pyramid pooling	Dice coefficient: 0.74	The concept of pyramid module with dimensionality reduction strategies are beneficial.	Segmentation is performed for drusen to detect the AMD stages. Other clinical findings are not taken into consideration.
3	Li et al. (2019)	Duke university dataset	CNV, DME, DRUSEN, and NORMAL.	Integrating hand-crafted features like curvature and thickness of retina layers into the Inception v3 model	Accuracy: 0.94 to 0.99, Sensitivity:0.97, Specificity:0.99	Extracting SIFT and Gabor features, concatenating the features to the deep learning model is the novelty.	Model is tested on only one dataset. Denoising is not incorporated.
4	Wang et al. (2019a)	Zhongshan Ophthalmic Center, China	Classification of glaucoma	Performs segmentation followed by classification to detect the glaucoma	Accuracy: 0.814, AUC: 0.864	The RNFL thickness estimated from segmentation process is used as an input to classify the retinal disorder.	The authors concentrated only on one retinal disorder and one image modality.
5	Kermany et al. (2018b)	Various hospitals, China	DRUSEN, CNV, DME, NORMAL	Transfer learning using Inception model	Accuracy : 0.966	Pretrained models are finetuned for OCT image dataset. Images from various hospital are considered to achieve good results	The details of the size of drusen or other clinical findings is not estimated.
6	Fauw et al. (2018)	Moorfields Hospital, UK	Classification of retinal disorders.	Segmentation and classification is performed using OCT images.	AUC: 0.99	Authors of this paper define classification labels as urgent, semi-urgent, routine, and observation based on the referral needed for a patient.	The authors do not mention about the quality of the images considered.
7	Roy et al. (2017)	Duke university SD-OCT dataset for DME	Retina layers and fluid segmentation	Fully convolutional network consisting of contracting encoder and expanding decoder path without integrating hand-crafted features.	Dice Coefficient: 0.99	Stability against imbalanced classes as a composite loss function comprising of a weighted logistic regression loss along with Dice loss is used.	Although the dice coefficient value obtained is acceptable, reproducible segmentation results could not be achieved. Further, disorder detection is not performed.

graded.

- The distortion in OCT images is anticipated to be due to speckle. Speckle is

multiplicative in nature and therefore, it can be formulated as a data correlated noise. In other words, a small perturbation in speckle affected data leads to amplification of intensity points in the data. Despeckling is thus a challenging task. The existing image restoration algorithms blindly attempt to denoise and enhance the contrast of images without considering the noise distribution in the data. As Gaussian noise is additive and data-independent in nature, it does not capture the inherent properties of the speckle. Therefore, it is necessary to propose a restoration model that is adaptive to various distortions in the image.

- As the deep learning based restoration models rely on availability of the ground-truth images and plethora of training images, there is a need to propose a restoration model that can eliminate this requirement using concept such as deep image prior.
- Transfer learning methods are widely applied to classify multiple retinal disorders. Though the classification metrics of transfer learning models achieve promising accuracy levels, these models may not offer reliable results for heterogeneous datasets, especially when the test dataset significantly differs from training dataset. The adversarial nature of GAN makes it tunable for a variety of applications such as generation of images, segmentation, and classification even with partially available labeled data. Therefore, it is the need of the hour to analyse the stability of GANs to classify multi-modal retinal images into normal or abnormal.

Though a wide range of studies are available on each of these sub-domains, proposing an end-to-end analysis for multiple retinal disorder detection is an open research problem. This is performed through the proposed research work. Considering the above mentioned limitations and the ability of GANs to deal with semi-supervised data, the performance of GANs for the dual task of segmentation and classification from CFP and OCT images is explored through this research work.

## 2.6 SUMMARY

The retinal image analysis has been a limelight of research since the times of yore. Exhaustive research publications are available related to the diverse imaging modalities and disorders under study. The surge of articles has transformed the field of telemedicine to the reality. The diverse methodologies and non-standardized quantification metrics often tend to leave a juvenile researcher perplexed to identify the research gaps. This chapter highlights some of the relevant works in retinal image analysis (Indian and International works), considering the last five years as a major focus. Deep learning methods are the spotlight due to the availability of massive computational resources. Though the training time is more, often the deep learning methods have outperformed traditional methods in terms of metrics such as accuracy.

The common underlying theme for automatic retinal disorder detection mechanism is to acquire training images from relevant imaging modalities, perform preprocessing to enhance the contrast and denoise the images, further train the deep learning model using various loss functions and deep learning architecture. The focus of such automated diagnosis methods is directed towards classification metrics of the deep learning model or the dice coefficient for segmentation. The proposed work in the thesis shifts the direction of importance towards restoring the images without oversmoothing or loss of texture and edge details. When the details of the image is well preserved, it assists to identify significant features during the segmentation and classification tasks. Variational methods associated with deep image prior concept are potential techniques to restore image quality in this regard. Furthermore, the concept of semi-supervised GAN architecture from the existing literature can be potentially expanded to perform dual task of segmentation and classification of retinal images.

This chapter provides an insight into assorted preprocessing, segmentation, and classification techniques trained and tested on private and public datasets. Majority of these methods are not reproducible due to the incomplete implementation details or due to the mismatch between input datasets. Therefore, a handful of works are considered for comparative analysis in this thesis. Apart from stating the pros and cons for every method, this chapter provides a bird's eye view of variational methods and

non-local restoration frameworks. Finally, the chapter is concluded by identifying the research gaps that are addressed through the proposed works.

## Chapter 3

# RESTORATION AND ENHANCEMENT OF RETINAL IMAGES

This Chapter proposes two different image restoration frameworks based on the fact that the two modalities differ in image acquiring procedure. Recall from Chapter 1 that CFP images suffer from contrast degradation. A novel retinex based non-local total generalized variational framework is proposed in section 3.3 for contrast enhancement. Further, realizing that speckle is inherent in grayscale OCT images, an adaptive non-local deep image prior model is proposed in section 3.4. These two preprocessing techniques restore the images such that the edge and texture related fine details are well preserved.

As a first step, several CFP and OCT images are acquired from public repositories. These images collectively differ in terms of resolution since various devices are used to capture these images. Details of the datasets are elaborated in Chapter 4. Nevertheless, some of the degraded CFPs are demonstrated in Figure 3.1. The first row of Figure 3.1 corresponds to images of patients with Glaucoma. The second row depicts the AMD condition, and the last row is the fundus of normal retina. As evident from this Figure, the illumination is non-uniform. These images are randomly selected from the ODIR-2019 grand challenge dataset. A fundus image is regarded as degraded if the structures such as the optic disc region, macula region, or blood vessels are not prominently visible. Certain retinal disorders such as cataract also mask the structures on the fundus. However, since our study is restricted to retinal disorders such as Glaucoma and AMD, the degradation is assumed to be due to illumination inhomogeneity and pres-

ence of drusen deposits. Thus the restoration algorithm should enhance the features while retaining the structures such as optic disc and macula.

Similar to CFP, OCT images are distorted in terms of quality based on the dataset. Certain instances of noisy OCT images are depicted in Figure 3.2. The OCT-B scans of AMD cases are depicted in the first row of the Figure 3.2, while the successive rows correspond to the DME and Normal conditions, respectively. The OCT images are considered noisy if the retinal layers are not distinguishable. The degradation is assumed to be due to speckle, as mentioned earlier in Chapter 2. A detailed data distribution analysis of fundus and OCT images is presented in next section.

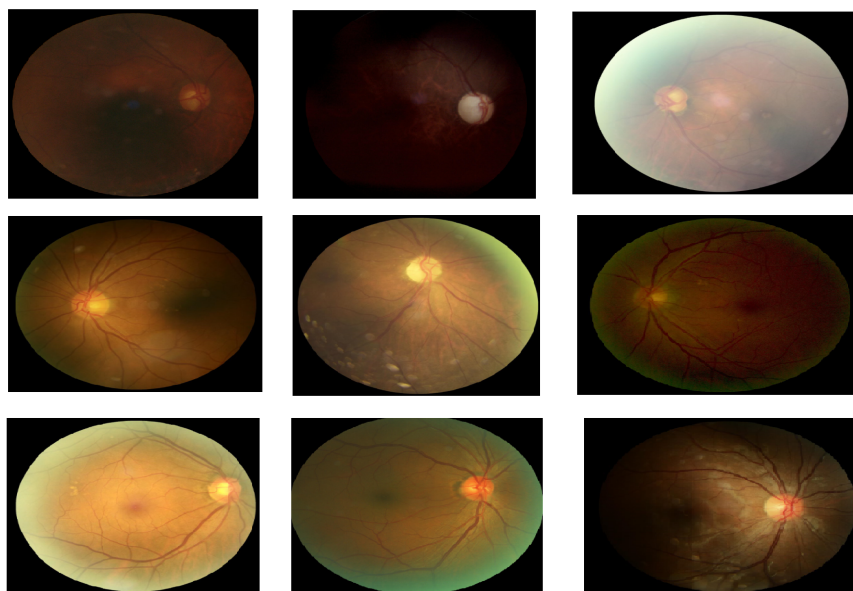


Figure 3.1: Randomly selected degraded fundus images.

### 3.1 DATA DISTRIBUTION ANALYSIS

In general, the histogram plot of a homogeneous region can be used to identify the contrast information and data distribution. Hence, the histogram of a relatively cleaner fundus image and a degraded image is provided in Figure 3.3. Clearly, it can be concluded from Figure 3.3(a) that the well-illuminated fundus image has a greater range of intensities representing the image. On the other hand, the contrast limited fundus image has a smaller spread of the RGB channel intensities (Figure 3.3(b)). The X-axis in the histogram plot of the entire fundus image, represents the intensity values that can range



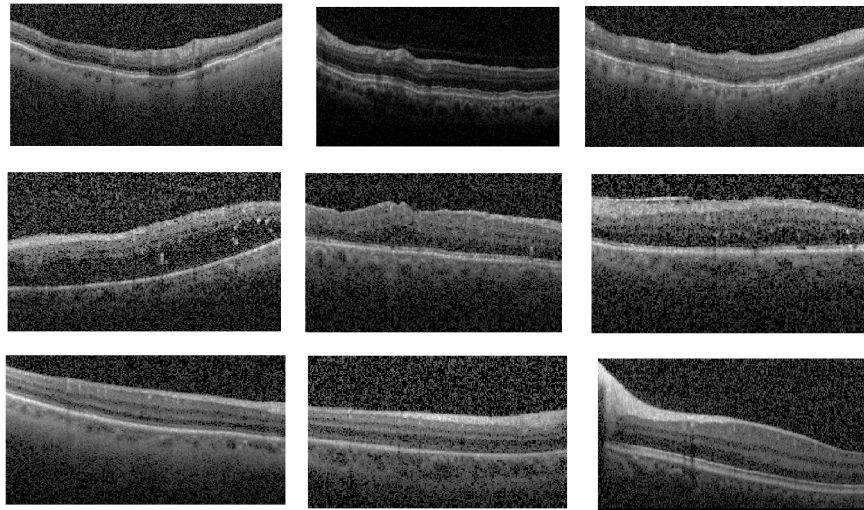
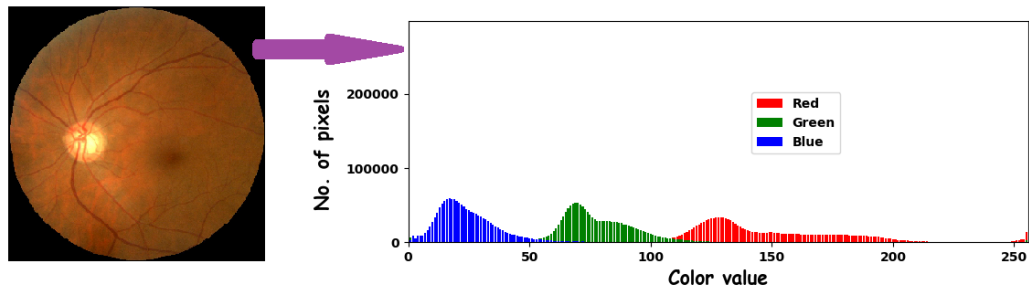


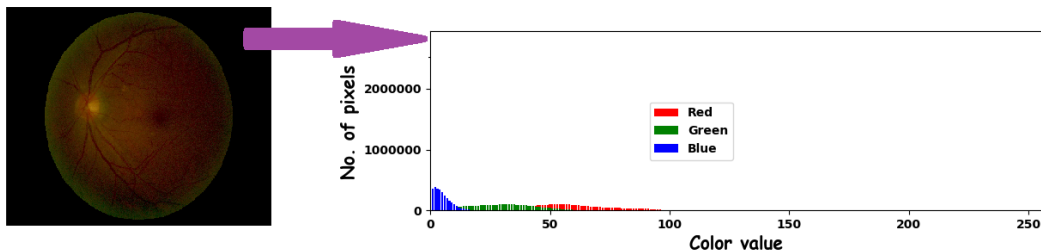
Figure 3.2: Randomly selected degraded OCT images.

between [0-255]. The Y-axis denotes the frequency corresponding to a particular intensity value. Furthermore, realizing the working principle of ophthalmoscopes (used to acquire CFP, discussed in Chapter 1), from the existing literature and our experimental study, it is observed that the data in CFP follows Gaussian distribution.

A similar histogram plot for OCT images is shown in Figure 3.4. The histogram of the



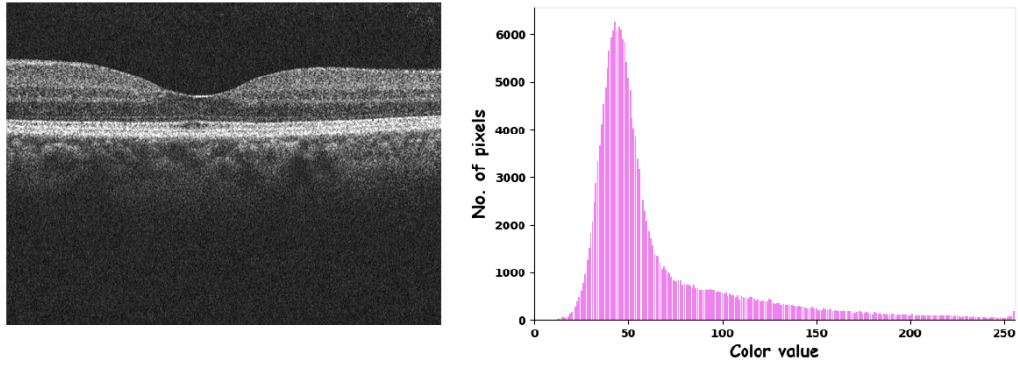
(a) Well-illuminated RGB fundus image.



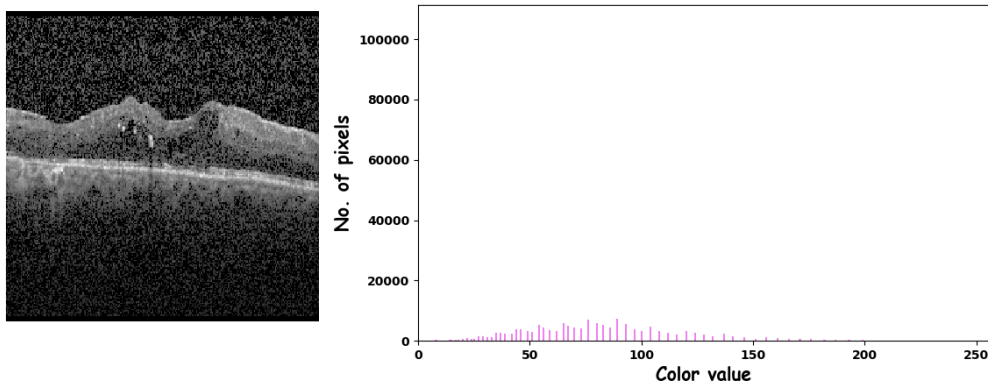
(b) RGB fundus image with low contrast.

Figure 3.3: Histogram comparison of CFP.

image shown in subfigure 3.4(a) follows a particular distribution and the retinal layers



(a) OCT image with distinguishable retina layers.



(b) Degraded OCT image.

Figure 3.4: Histogram comparison of OCT images.

are relatively distinguishable. It corresponds to the OCTID repository (Gholami et al. (2020)) available in public domain, as discussed in Chapter 4. OCT image depicted in subfigure 3.4 (b) corresponds to a different public repository (Reza et al. (2018)) and here the influence of speckle is evident from its histogram plot.

Since OCT images are noisy and grayscale in nature, various formulations are proposed in existing literature, as discussed in Chapter 2. Therefore, the OCT images considered in the proposed research work, is analysed statistically to obtain the prior information about the noise model. Motivated by Febin et al. (2018), the data distribution of image is identified using a similar procedure. Accordingly, a small homogeneous region of variable size is randomly selected from multiple locations on OCT images. The normalized patch intensity values are fitted with standard distribution functions, namely, 'Gaussian', 'Gamma', and 'Rayleigh', and the histogram of the data is plotted against the standard curves. One example is shown in Figure 3.5(a). Here, the patch

size is  $49 \times 52$ . The histogram plot appears to be much closer to the Gamma distribution from Figure 3.5(b). The Jensen–Shannon divergence (JSD) (Fuglede and Topsoe (2004)) is estimated between the actual data and the fitted distribution. Refer Appendix D.1.1 for the details of JSD. The smaller the distance, the higher is the similarity of the data with the corresponding distribution. For most of the patches considered from different OCT images, a lower value of JSD is recorded for Gamma distribution as shown in Figure 3.5(c). The p-value hypothesis test (Glen (2014)) is yet another way to iden-

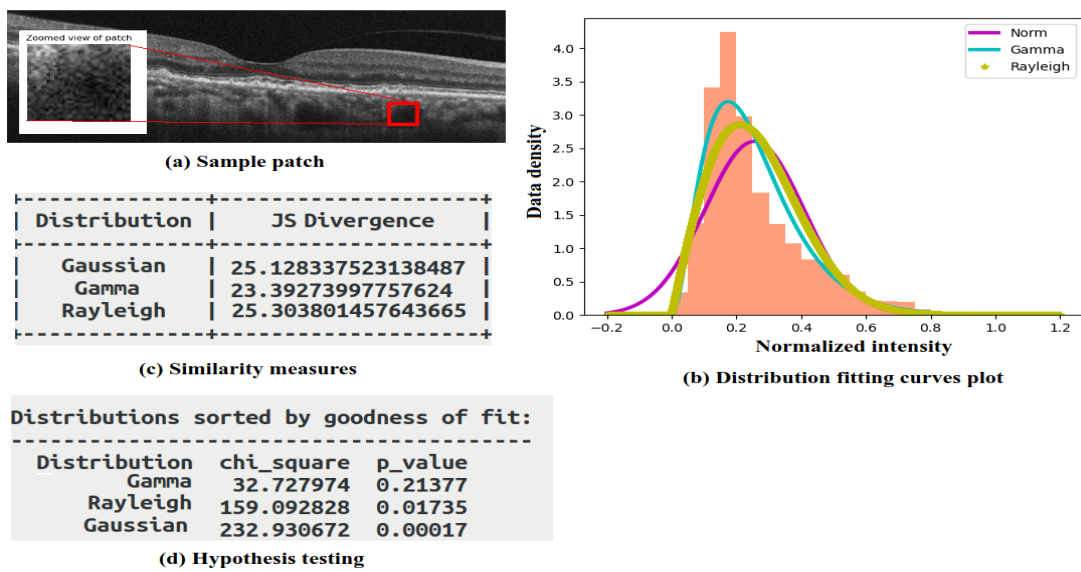


Figure 3.5: Analysis of data distribution in a sample OCT image.

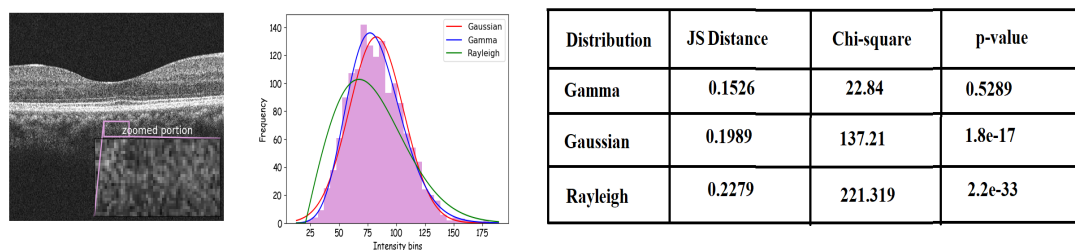


Figure 3.6: Data distribution of another OCT image.

tify the data distribution. The NULL hypothesis is: distribution follows Gamma. A goodness of fit test is performed to confirm the similarity using Chi-square test. The results are summarized in Figure 3.5(d). Accordingly, a higher  $p$  – value indicates that the NULL hypothesis has to be accepted. In our study, 95% Confidence Interval is used. This confirms that there is a close proximity of data distribution with Gamma.

A similar instance is depicted in Figure 3.6 for another randomly selected OCT image. Further, Bayesian maximum apriori estimate is employed to formulate the data fidelity term, under the assumption that speckles in OCT images follow Gamma distribution.

### 3.2 BAYESIAN MAP ESTIMATE FOR GAMMA DISTRIBUTION

As the noise is multiplicative in its characteristics, it follows the following formulation:

$$U_0 = U * N. \quad (3.1)$$

where ' $U_0$ ' & ' $U$ ' are noisy and clean images, respectively,  $*$  denotes a normal point-wise multiplication, ' $N$ ' denotes data correlated noise.

Since the speckle noise of OCT images closely approximates Gamma distribution, the probability density function  $f(U_0)$  is given as:

$$f(U_0) = \frac{U_0^{\alpha-1} e^{-\frac{U_0}{\zeta}}}{\Gamma(\alpha)\zeta^\alpha}. \quad (3.2)$$

where  $\Gamma(\alpha)$  is standard Gamma function,  $\alpha$  &  $\zeta > 0$  are the shape and scale parameters, respectively.

The variational framework for Gamma distribution is derived using Bayesian Maximum apriori estimate (Aubert and Aujol (2008); Balaji and Jidesh (2017)). The formulation is restated here to ensure continuity in reading. The posterior estimate for the noise density function is given as:

$$p(U_0|U) = \frac{p(U|U_0)p(U_0)}{p(U)}, \quad (3.3)$$

where  $p(U)$  denotes the probability function,  $p(U_0|U)$  and  $p(U|U_0)$  are posterior and likelihood estimates, and  $p(U_0)$  is a prior term associated with Gibbs's prior denoted as:

$$p(U_0) = \frac{e^{-\lambda\phi(U_0)}}{T_n}, \text{ where } T_n \text{ is a normalizing constant.} \quad (3.4)$$

The objective is to find the functional that maximizes the posterior estimate. Considering the log domain, maximizing the posterior is equivalent to minimizing the negative log function, given as:

$$\max_{U_0} \log p(U_0|U) = \min_{U_0} \left\{ -\log p(U|U_0) - \log p(U_0) + \log p(U) \right\}, \quad (3.5)$$

Under the assumption that the noise samples on each pixel are independent and identically distributed, the likelihood estimate is given by the product of sum terms denoted

as follows:

$$p(U|U_0) = \prod_i p(U_i|U_{0i}) \quad (3.6)$$

Since  $\log p(U)$  is a constant, one shall ignore its contribution while estimating the maximum posterior. Thus, we can rewrite (3.5) as:

$$\max_{U_0} \log p(U_0|U) = \min_{U_0} \left\{ -\log \frac{U_0^{\alpha-1} e^{-\frac{U_0}{\zeta}}}{\Gamma(\alpha)\zeta^\alpha} - \log \frac{e^{-\lambda\phi(U_0)}}{T_n} \right\}, \quad (3.7)$$

This can be simplified further assuming that the contribution of constant terms are negligible. The final optimization model amounts to:

$$\max_{U_0} \log p(U_0|U) = \min_{U_0} \left\{ \int_{\Omega} \left( \log U_0 + \frac{U}{U_0} \right) d\Omega + \lambda \int_{\Omega} \phi(U_0) d\Omega \right\}, \quad (3.8)$$

where  $\Omega = dx dy$  represents the entire image domain.

With necessary preliminary understanding about the data distribution in fundus and OCT images, section 3.3 highlights the retinex framework and non-local variational model to address the contrast degradation of fundus images.

### 3.3 RETINEX FRAMEWORK

Variational retinex models (Kimmel et al. (2003)) has been phenomenal in alleviating the illumination inhomogeneity. Therefore, a perceptually inspired non-local retinex framework (Zosso et al. (2015)) is adopted to enhance fundus images in our proposed works. The advantage of using non-local models over traditional image enhancement methods, like CLAHE is that the non-local methods preserve the texture and other edge details present in images. Therefore, the quality of image is enhanced without losing significant information. The retinex based non-local total bounded variational framework is used to restore ultrasound images in earlier works (Febin and Jidesh (2021)) which is modified to fit the requirements of fundus image enhancement. For the sake of completeness, the retinex based image restoration model is explained here. The retinex theory states that the intensity function can be decomposed into reflectance ( $R$ ) and illuminance ( $L$ ), respectively. The symbolic representation of this fact is:

$$U = R * L, \quad (3.9)$$

where  $U$  is the observed intensity (luminance),  $0 < R < 1$  is the reflectance component, and  $0 < L < \infty$  denotes the illuminance component, respectively, defined in the image

domain  $\Omega \subseteq \mathbb{R}^2$ , and  $*$  represents pointwise multiplication operation. Considering the log domain, (3.9) can be rewritten as:

$$u = r + l. \quad (3.10)$$

where  $u = \log U, r = \log R, l = \log L$ . Here the illuminance is the property of the source, and it is spatially smooth function, while reflectance, substantially depends on the illuminated surface, usually represented by a piecewise constant function. Therefore, these two components have to be treated independently to restore images. In general, in a retinex based TV image restoration framework (elaborated in Appendix A.2), the illumination inhomogeneity is minimized by considering the input image as an energy functional where the objective is to estimate the minimum of the functional that satisfies a set of constraints. In general, the awareness about limited dynamic range of the reflectance is set up as a constraint. This constraint minimization problem is converted to an unconstrained one by introducing an additional penalty term for data fidelity which is controlled by a regularization parameter. The formulation (introduced by Ng and Wang (2011)) is given as:

$$E(l, r) = \min_{l, r} \left\{ \int_{\Omega} \|\nabla r\| d\Omega + \gamma_1 \int_{\Omega} \|\nabla l\|_2^2 d\Omega + \gamma_2 \int_{\Omega} (l - r - u)^2 d\Omega + \gamma_3 \int_{\Omega} l^2 d\Omega \right\}. \quad (3.11)$$

where  $d\Omega = dx dy$ ,  $\|\nabla r\|$  denotes the TV norm of the reflectance, second term  $\|\nabla l\|_2^2$  is the squared  $L_2$  norm of illuminance,  $\gamma_1, \gamma_2, \& \gamma_3$  are regularization parameters. The term  $(l - r - u)$  denotes the data fidelity, and the term associated with  $\gamma_3$  is an additional term for theoretical setting. However, the piecewise approximation causes visual discrepancies in the output. Moreover, the ordinary TV semi-norm is local in nature. Therefore, the TV model is extended to a non-local domain (adapted from Febin and Jidesh (2021)), by incorporating a weighted norm of the reflectance term. The energy minimization model is denoted as:

$$E(l, r) = \min_{l, r} \left\{ \gamma_1 \int_{\Omega} W(r) \|\nabla_{NLTV} r\| d\Omega + \gamma_2 \int_{\Omega} \left( \exp(r) - \frac{1}{2} \right)^2 d\Omega + \gamma_3 \int_{\Omega} \|\nabla l\|_2^2 d\Omega + \gamma_4 \int_{\Omega} (r + l - u)^2 d\Omega \right\}. \quad (3.12)$$

where the first term associated with  $\gamma_1$ , corresponds to the non-local weighted norm of reflectance, where  $W(r) = 1 + \frac{1}{1 + \|\nabla r\|}$ . The reflectance term is conditioned to be close to average value as in Febin and Jidesh (2021); Li et al. (2012) using  $\int_{\Omega} \left( \exp(r) - \frac{1}{2} \right)^2 d\Omega$

term. As illuminance is spatially smooth,  $L_2$  norm is introduced to the model for the illumination correction. Minimization of error can be obtained by introducing the constraint  $\int_{\Omega} (r+l-u)^2 d\Omega$  for data fidelity.  $\gamma_1, \gamma_2, \gamma_3, \& \gamma_4$  are regularization parameters. This functional is solved using Split Bregman (SB) method (Goldstein and Osher (2009); Goldstein et al. (2009); Liu and Huang (2010)) and Euler-Lagrange equation. Refer Appendix A.4 for details of SB scheme and Appendix A.3 for Euler-Lagrange method. Applying SB technique, (3.12) yields three sub-problems denoted as:

$$E(r) = \min_r \left\{ \gamma_1 \int_{\Omega} W(r) \|\nabla r\| d\Omega + \gamma_2 \int_{\Omega} \left( \exp(r) - \frac{1}{2} \right)^2 d\Omega + \gamma_4 \int_{\Omega} (r+l-u)^2 d\Omega \right\}. \quad (3.13)$$

$$E(u) = \min_u \left\{ \gamma_4 \int_{\Omega} (r+l-u)^2 d\Omega \right\}. \quad (3.14)$$

$$E(l) = \min_l \left\{ \gamma_3 \int_{\Omega} \|\nabla l\|_2^2 d\Omega + \gamma_4 \int_{\Omega} (r+l-u)^2 d\Omega \right\}. \quad (3.15)$$

These sub-problems are solved iteratively to obtain minimal solution, that represents the restored image. Since the input image is a 3-channel RGB image, in the proposed approach, the input fundus image (RGB) is initially converted to Hue, Saturation, Values (HSV) components. As the variations in the image are preserved in the V component of HSV image, this is used for further processing (Refer Kimmel et al. (2003) for the concept of HSV retinex framework). Finally, the RGB image is obtained using reverse mapping of color space HSV to RGB. It is to be noted that the H and S components are unaltered. The final enhanced RGB fundus image is used for classification. This method is referred to as Non-local Total Variational Retinex (NLTVR) in the rest of the document.

### 3.3.1 Need of L1 and L2 regularizers

Image denoising is treated as an energy minimization problem and solved using mathematical models. In general, the ill posed problem such as denoising is formulated using norm minimization where the objective is to find the minimal solution that satisfies the given constraints. For image denoising, the constraint is to preserve the data fidelity while minimizing the noise (abrupt changes). Adding regularization terms will assist to convert the constraint minimization problem to an unconstrained problem. The L2-norm (Tikhonov regularization) removes the noise content and might result in over-

smoothing and loss of edge details leading to a blurred image. Hence L1-norm is used as it has edge-preserving ability. **List of assumptions for the proposed work:**

- If the image is generated from a physical process, its intensity values are proportional to the energy radiated by a physical source. Therefore the image function is finite and is expressed as a combination of illumination and reflection functions.
- The illumination function is spatially smooth.
- The reflection function is piecewise constant. It is bounded between 0 and 1.
- A noisy image has higher total variation than a clean image. Thus total variation minimization is visualized as a convex minimization problem.
- The speckle is multiplicative in nature and follows Gamma distribution.
- The choice of architecture does not have any significant impact on the results for a deep image prior model.

### 3.3.2 Proposed Non-local TGVR framework

Recall from Chapter 2 and Appendix A.2, that the non-local total variational restoration models are beneficial for texture preserved image restoration. Among the higher-order TGV models, second-order TGV model is extensively used for various image processing applications in the literature Kim and Kudo (2020). Since most existing Total Variation model considers only the first-order derivative, it suffers from problems such as staircase artifacts and loss in smooth intensity changes for textures and low-contrast objects, which is a major limitation in improving image quality. The first-order derivative is too sensitive to the pixel values, even linear intensity changes are detected as false edges, which leads to staircase artifacts. The higher-order derivatives possess a potential risk that, as the order of differentiation is larger, its ability to enhance image edges is smaller leading to an image blurring problem. In the proposed work, we use a combination of first and second-order TGV to preserve smooth intensity changes well. The magnitude of first and second-order terms is controlled using a regulating function to eventually improve the natural outlook of the restored results. The corresponding



NLTGVR model includes two regularization terms: a non-local total variation (NLTV) and a non-local total k-split variation (NLTKV). The NLTGV gradient  $\nabla_{NLTGVR}$  is calculated using the following equation,

$$\nabla_{NLTGVR} = t\nabla_{NLTV} + \frac{(1-t)}{8}\nabla_{NLTKV}, \quad (3.16)$$

where  $t$  is a parameter to control the magnitude of first and second-order derivatives. The first order derivative  $\nabla_{NLTV}$  is estimated as in Gilboa and Osher (2009) and second-order derivative  $\nabla_{NLTKV}$  of pixels  $x_i$  and  $x_j$  is calculated as:

$$\nabla_{NLTKV} = \sum_{k=1}^8 |(u_i - u_{i_k}) - (u_j - u_{j_k})|, \quad (3.17)$$

where the value of  $k$  changes from 1 to 8 to find the difference between eight neighboring pixels (see Kim and Kudo (2020) for more details). For convenience,  $\nabla_{NLTGVR}$  of  $r$  and  $u$  are represented as  $\tilde{\nabla}r$  and  $\tilde{\nabla}u$ , respectively in this document. As stated earlier (in section 3.1), from the data distribution analysis and histogram of fundus images, it is evident that contrast upgradation is necessary. However, without loss of generality, it is assumed that the data follows Gaussian distribution. To address any additive white Gaussian noise effectively, a weighted non-local total generalized bounded variation of intensity image  $u$  is introduced as given below:

$$\int_{\Omega} (W(u)\|\tilde{\nabla}u\| + \beta\|u\|_2^2)d\Omega, \quad (3.18)$$

where  $\|\tilde{\nabla}u\|$  denotes the TV norm computed with non-local gradients, and a high value of  $\beta$  adds more smoothing into the result by employing  $L_2$  norm (for further details refer Febin and Jidesh (2021)). To enhance contrast of the image, a constraint to enforce the histogram equalization (see Wang et al. (2020b)) is also included in the proposed model, given as:

$$\int_l^L (H_U(\tau) - H_c(\tau))^2 d\tau. \quad (3.19)$$

Here, the pixels are assumed to have integer values  $\tau$  in the range  $[l, L]$  (for an 8 bit gray scale image, it is in the range  $[0, 255]$ ),  $H_c$  is the cumulative histogram of a uniform histogram image which is calculated as below:

$$H_c(\tau) = \frac{\tau - l}{L - l}, \quad (3.20)$$

and  $H_U$  is the cumulative histogram of output image  $U$  which is estimated as follows

$$H_U(\tau) = \frac{1}{M} \sum_{i \in \Omega} \chi_{[l, \tau]} U_i, \quad (3.21)$$

with

$$\chi_{[l,\tau]}(U_i) = \begin{cases} 1, & U_i \in [l, \tau] \\ 0, & \text{otherwise} \end{cases}. \quad (3.22)$$

where  $M$  is the total number of points in the image domain  $\Omega$ . The overall proposed model for optimization is designed as:

$$\begin{aligned} E(l, r, u) = \min_{l, r, u} \{ & \gamma_1 \int_{\Omega} W(r) \|\tilde{\nabla} r\| d\Omega + \gamma_2 \int_{\Omega} \left( \exp(r) - \frac{1}{2} \right)^2 d\Omega \\ & + \gamma_3 \int_{\Omega} \|\nabla l\|_2^2 d\Omega + \gamma_4 \int_{\Omega} (r - u + l)^2 d\Omega \\ & + \gamma_5 \int_{\Omega} (W(u) \|\tilde{\nabla} u\| + \beta \|u\|_2^2) d\Omega + \gamma_6 \int_l^L (H_U(\tau) - H_c(\tau))^2 d\tau \}, \end{aligned} \quad (3.23)$$

where  $W(z)$  for  $z \in \{u, r\}$ , is the weight function evaluated as  $1 + \frac{1}{1 + \|\nabla z\|}$  as explained in Febin and Jidesh (2021). It aids to preserve the fine details. The weight increases as the magnitude of gradient decreases and reaches a maximum value two, for a homogeneous region. Equation (3.23) can further be split into separate minimization problems as below:

$$\begin{aligned} E(r) = \min_r \{ & \gamma_1 \int_{\Omega} W(r) \|\tilde{\nabla} r\| d\Omega + \gamma_2 \int_{\Omega} \left( \exp(r) - \frac{1}{2} \right)^2 d\Omega \\ & + \gamma_4 \int_{\Omega} (r - u + l)^2 d\Omega \}, \end{aligned} \quad (3.24)$$

$$\begin{aligned} E(u) = \min_u \{ & \gamma_4 \int_{\Omega} (r - u + l)^2 d\Omega + \gamma_5 \int_{\Omega} (W(u) \|\tilde{\nabla} u\| + \beta \|u\|_2^2) d\Omega \\ & + \gamma_6 \int_l^L (H_U(\tau) - H_c(\tau))^2 d\tau \}, \end{aligned} \quad (3.25)$$

and

$$E(l) = \min_l \left\{ \gamma_3 \int_{\Omega} \|\nabla l\|_2^2 d\Omega + \gamma_4 \int_{\Omega} (r - u + l)^2 d\Omega \right\}. \quad (3.26)$$

### 3.3.3 Numerical Implementation

The above stated minimization problems can be solved using an efficient numerical implementation. Here we use the SB iteration technique for obtaining a fast convergence and reducing the parameter sensitivity as mentioned earlier. According to SB model, new constraints  $c_1 = \tilde{\nabla} r$ ,  $c_2 = \tilde{\nabla} u$  and auxiliary variables  $a_1, a_2$  are introduced to reform

the minimization problem in (3.24) and (3.25) as given below:

$$E(r) = \min_r \left\{ \gamma_1 \int_{\Omega} W(r) \|c_1\| d\Omega + \gamma_2 \int_{\Omega} \left( \exp(r) - \frac{1}{3} \right)^2 d\Omega \right. \\ \left. + \gamma_4 \int_{\Omega} (r - u + l)^2 d\Omega + \gamma_7 \int_{\Omega} \|c_1 - \tilde{\nabla}r - a_1\|_2^2 d\Omega \right\}, \quad (3.27)$$

$$E(u) = \min_u \left\{ \gamma_4 \int_{\Omega} (r - u + l)^2 d\Omega + \gamma_5 \int_{\Omega} (W(u) \|c_2\| + \beta \|u\|_2^2) d\Omega \right. \\ \left. + \gamma_6 \int_l^L (H_U(\tau) - H_c(\tau))^2 d\tau \right. \\ \left. + \gamma_8 \int_{\Omega} \|c_2 - \tilde{\nabla}u - a_2\|_2^2 d\Omega \right\}. \quad (3.28)$$

For solving the equations (3.27) and (3.28) they are split into following sub-problems:

$$r^{k+1} = \min_r \left\{ \gamma_2 \int_{\Omega} \left( \exp(r) - \frac{1}{2} \right)^2 d\Omega + \gamma_4 \int_{\Omega} (r - u + l)^2 d\Omega \right. \\ \left. + \gamma_7 \int_{\Omega} \|c_1 - \tilde{\nabla}r - a_1^k\|_2^2 d\Omega \right\}, \quad (3.29)$$

$$c_1^{k+1} = \min_r \left\{ \gamma_1 \int_{\Omega} W(r) \|c_1\| + \gamma_7 \int_{\Omega} \|c_1 - \tilde{\nabla}r - a_1^k\|_2^2 d\Omega \right\}, \quad (3.30)$$

$$u^{k+1} = \min_u \left\{ \gamma_4 \int_{\Omega} (r - u + l)^2 d\Omega + \beta \|u\|_2^2 + \gamma_6 \int_l^L (H_U(\tau) - H_c(\tau))^2 d\tau \right. \\ \left. + \gamma_8 \int_{\Omega} \|c_2 - \tilde{\nabla}u - a_2^k\|_2^2 d\Omega \right\}, \quad (3.31)$$

and

$$c_2^{k+1} = \min_u \left\{ \gamma_5 \int_{\Omega} W(u) \|c_2\| d\Omega + \gamma_8 \int_{\Omega} \|c_2 - \tilde{\nabla}u - a_2^k\|_2^2 d\Omega \right\}, \quad (3.32)$$

The auxiliary variables  $a_1$  and  $a_2$  are revised in each iteration as given below:

$$a_1^{k+1} = a_1^k + (\tilde{\nabla}r - c_1^{k+1}). \quad (3.33)$$

and

$$a_2^{k+1} = a_2^k + (\tilde{\nabla}u - c_2^{k+1}). \quad (3.34)$$

Equations (3.30) and (3.32) are solved using shrinkage formula. For any  $V$  and  $\Lambda$  the shrinkage operation is as given below:

$$\text{shrinkage}(V, \Lambda) = \frac{V}{|V|} \max(|V| - \Lambda, 0), \quad (3.35)$$

and using the same, (3.30) and (3.32) can be rewritten as follows

$$c_1^{k+1} = \text{shrinkage} \left( \tilde{\nabla}r + a_1^k, \frac{W(r)}{\gamma_7} \right), \quad (3.36)$$

and

$$c_2^{k+1} = \text{shrinkage} \left( \tilde{\nabla}u + a_2^k, \frac{W(u)}{\gamma_8} \right). \quad (3.37)$$

To solve minimization problems given in equations (3.29) and (3.31), Euler-Lagrange equation is applied and later it is solved in Fourier domain as stated below (assuming  $\gamma_2 = \gamma_3 = 1$ ):

$$r^{k+1} = \mathcal{F}^{-1} \left( \frac{\gamma_4 \mathcal{F}(u-l) - \gamma_9 \mathcal{F}(\exp(r) (\exp(r) - 1/2)) - \gamma_7 \mathcal{F}(\nabla \cdot (c_1 - \tilde{\nabla} r - a_1^k))}{\gamma_4} \right) \quad (3.38)$$

and

$$u^{k+1} = \mathcal{F}^{-1} \left( \frac{-\gamma_4 \mathcal{F}(r+l) + \gamma_6 \mathcal{F}(H_U(U) - H_c(U)) - \gamma_8 \mathcal{F}(\nabla \cdot (c_2 - \tilde{\nabla} u - a_2^k))}{\mathcal{F}(\beta - \gamma_4)} \right) \quad (3.39)$$

where  $F$  denoted the Fourier Transform and  $\mathcal{F}^{-1}$  is the inverse Fourier Transform. The step by step procedure for solving the model is explained in Algorithm 1.

---

**Algorithm 1:** Algorithm for proposed NLTGVR restoration framework.

---

**Input:**  $U_0 \leftarrow$  Noisy image

**Output:** Restored image  $U$

- 1 Initialize  $k = 1, \varepsilon = 0.0001, c_1^1 = 0, a_1^1 = 0, c_2^1 = 0, a_2^1 = 0, u^1 = \log(U_0), l^1 = \max(u^1)$ , and  $r^1 = (l^1 - u^1)$
  - 2 **while**  $\frac{|u^k - u^{k-1}|}{u^k} < \varepsilon$  **do**
  - 3     Calculate  $c_1^{k+1}$  using equation (3.36).
  - 4     Estimate  $r^{k+1}$  as in (3.38).
  - 5     Update  $a_1^{k+1}$  using (D.8).
  - 6      $c_2^{k+1}$  is calculated using equation (3.37).
  - 7     Estimate  $u^{k+1}$  as in (3.39).
  - 8     Update  $a_2^{k+1}$  using (3.34).
  - 9     Calculate  $l^{k+1}$  using the following equation  $l^{k+1} = \mathcal{F}^{-1} \left( \frac{\gamma_4 \mathcal{F}(u-r)}{\mathcal{F}(\gamma_4 + \gamma_3 \Delta)} \right)$ .
  - 10 **end**
  - 11 Update  $U$  as exponential of  $e^{(l+r)}$ .
- 

### 3.3.4 Experimental results

The implementation of retinex based non-local total generalized variational model to enhance the fundus images is implemented using Matlab2018b, on core i5-6200U CPU

@2.3GHz processor with 4 GB RAM. The visual results of enhancing randomly selected fundus images from the ODIR-2019 dataset are presented in Figure 3.7. Here, the first column represents input images. The second column denotes the output of CLAHE. The third and the fourth column depicts the output obtained using NLTVR and the proposed NLTGVR respectively. It is evident that applying CLAHE on these images does not necessarily address the illumination inhomogeneity. The proposed NLTGVR framework handles this issue while preserving the contrast in the optic disc, fovea, and blood vessel region. Visually, the non-local methods provide similar results. However, in some images, the staircase effect is visible at the boundary regions in the case of total variational retinex models which is smoothed out in total generalized variational model.

### **Result Analysis:**

Further, to quantitatively assess the image enhancement method, the histogram is plotted for these images as given in Figure 3.8. Each histogram plot corresponds to the images shown in Figure 3.7. The x-axis denotes the pixel intensity and the y-axis represents the number of pixels. Comparing the first column with the fourth column, we can infer that the NLTGVR maps the input image with the higher range of intensities. The CLAHE equalizes the histogram. Furthermore, a total of 50 CFP images are considered for quantitative analysis. The statistical measure of Contrast is computed as:

$$\mathbb{C} = \frac{\max - \min}{\max + \min} \quad (3.40)$$

where max and min are the maximum and minimum intensity values in a randomly selected patch on the image. From every fundus image, five different patches of size  $50 \times 50$  are selected around blood vessels, optic disc, and macula region.

The box plot shown in Figure 3.9, represents the average contrast values of the input fundus images, images enhanced using CLAHE, NLTVR, and the proposed NLTGVR methods. Here, the whiskers of the box plot corresponds to the maximum and minimum recorded contrast values. The horizontal strike line inside the boxes denote the median value. Comparing the CLAHE method with the other non-local enhancement methods, it can be inferred that the CLAHE has a greater spread of the contrast. This denotes

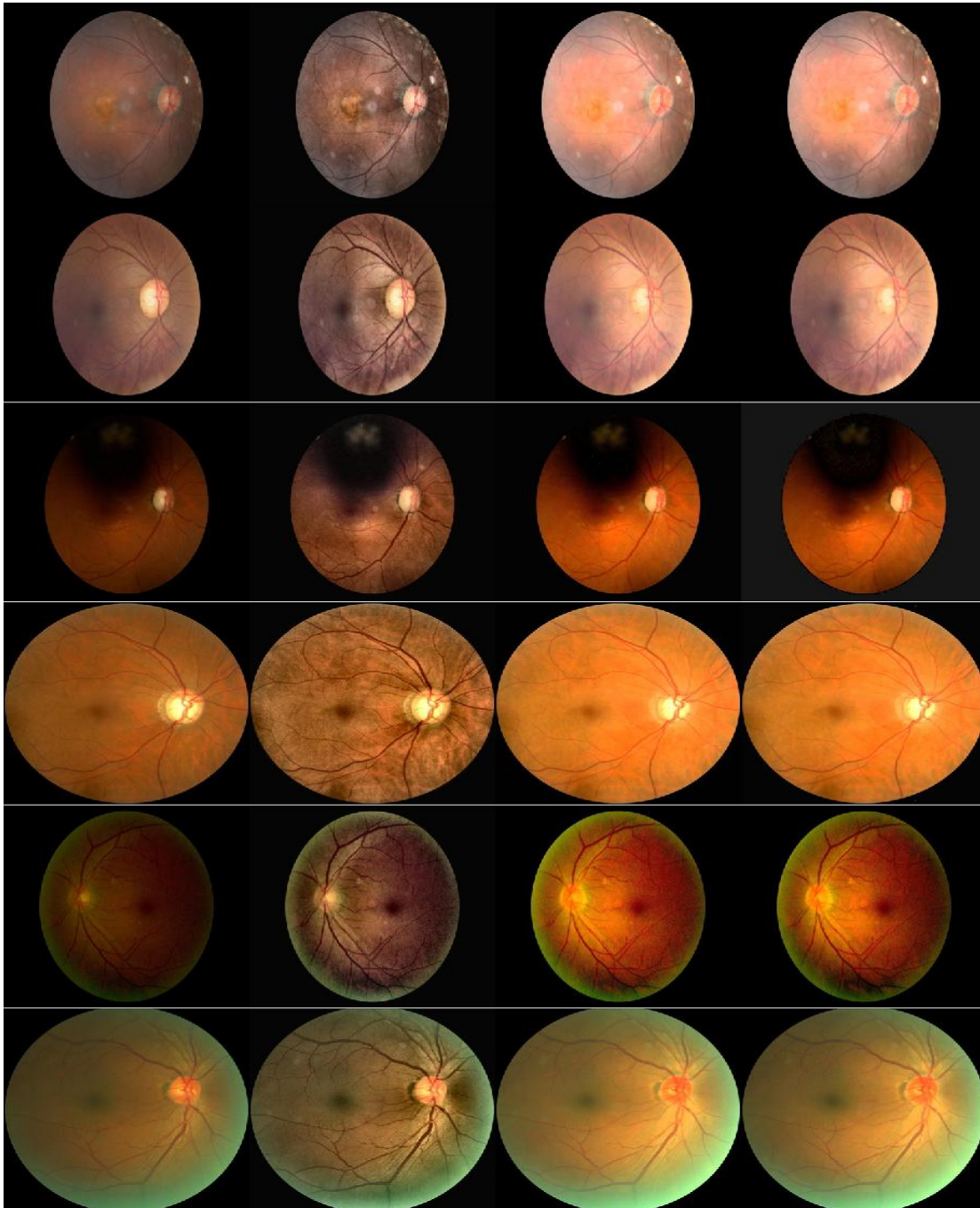


Figure 3.7: Input and enhanced fundus images.

that certain structures are over-smoothed in some images. For instance, refer to the visual output presented in sixth row of Figure 3.7. Here the CLAHE enhanced image has darkened the region on the right side of the optic disc that appears to resemble the macular region. However, there is a marginal difference between the proposed NLTVR method and the NLTVR method. Since the median value is approximately

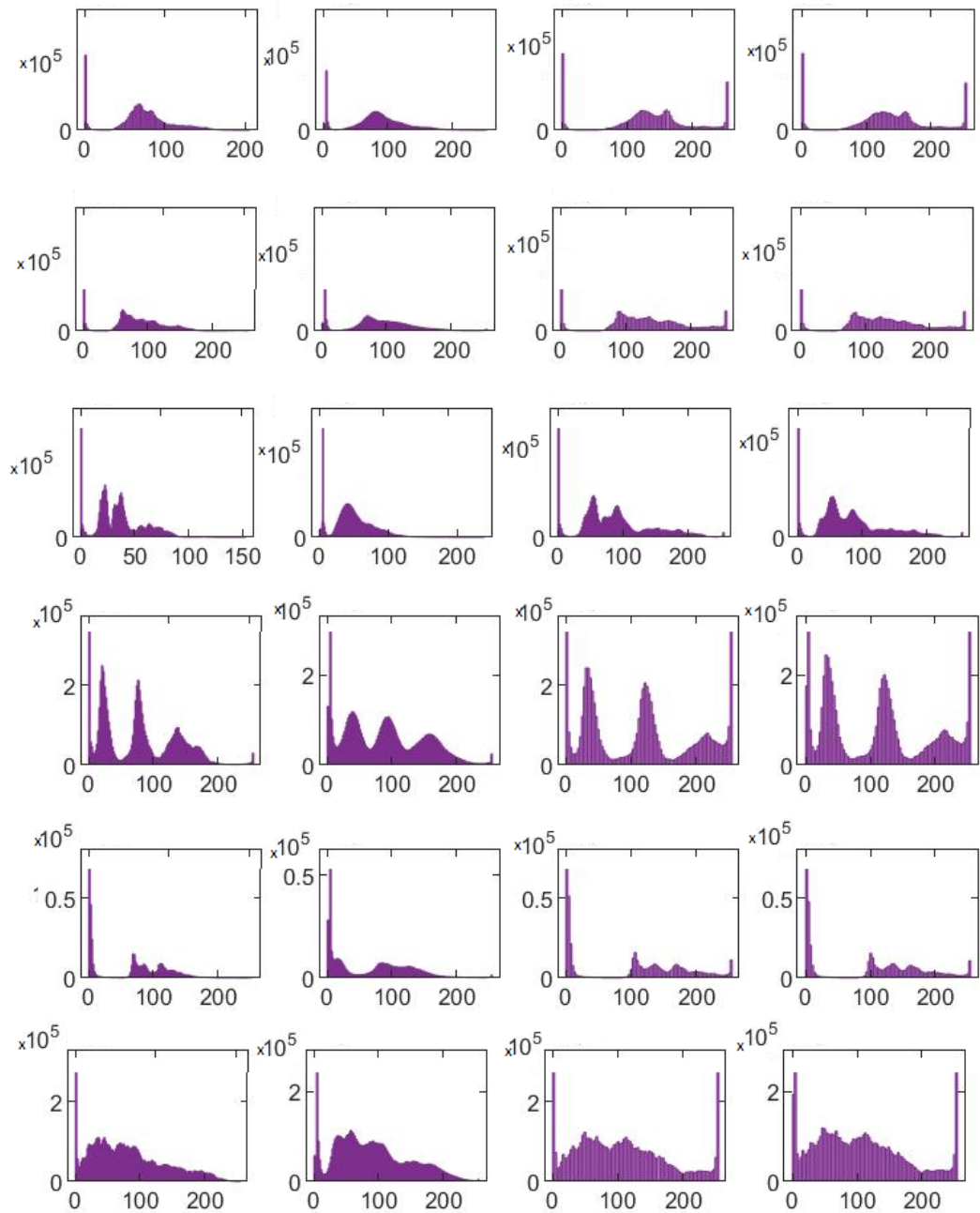


Figure 3.8: Histogram of degraded and enhanced images.

same as that of the input images, it confirms that the texture and edge details are well preserved in these non-local restoration methods.

Global Contrast Factor (GCF) is measured to obtain the overall contrast enhancement factor as recommended by Matković et al. (2005). Refer Appendix D.1.2 for further details. The average values are shown in the form of a horizontal bar chart for

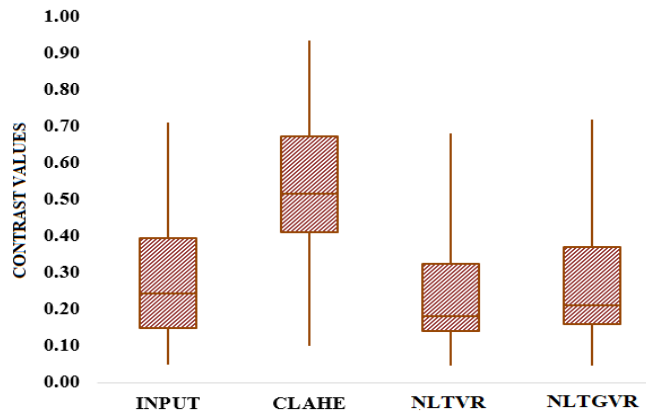


Figure 3.9: Average contrast plot.

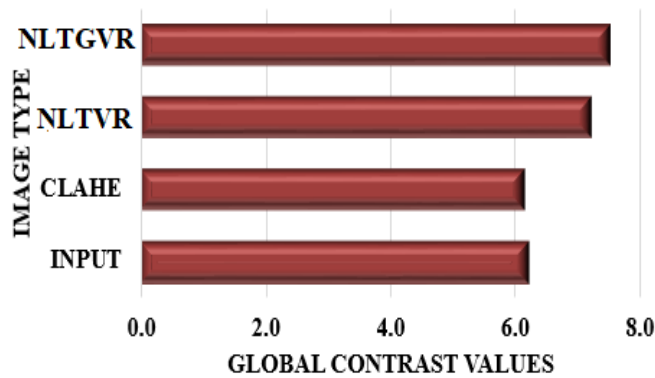


Figure 3.10: Global Contrast Factor.

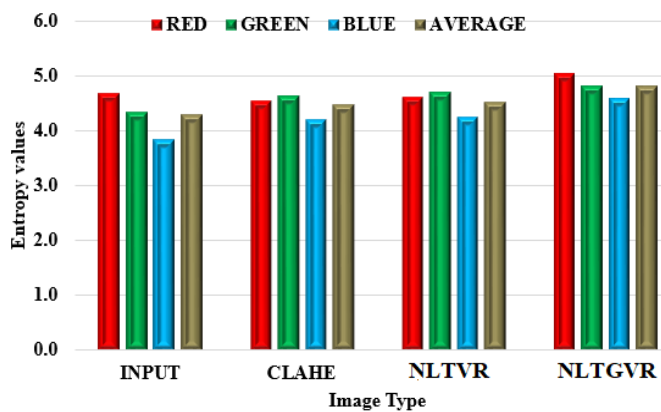


Figure 3.11: Entropy plot.

the unprocessed image and enhanced images in Figure 3.10. Global Contrast Factor is expressed as a weighted sum of the contrast in the neighborhood. Though the approach of non-local enhancement method is to consider the neighborhood patches to relatively enhance the images, Figure 3.10 shows that the proposed method enhances the fundus



images well.

Finally, entropy (Refer Appendix D.1.3.) is recorded to obtain the amount of information retained by each channel (Red, Green, Blue) of the RGB fundus image. Entropy is a measure of amount of information present in the image. A higher entropy value denotes that the structures such as blood vessels and the abnormalities are well preserved. Since the entropy is measured on a grayscale image, in general, for a RGB fundus image, the entropy is recorded for each channel separately. Accordingly, from Figure 3.11, the red channel has greater information compared to the green and blue channels in the input image. This relation holds good for the proposed method as well. CLAHE on the other hand, tends to equalize the pixel intensities. Overall, the average entropy value is highest in the fundus image enhanced using the proposed method. There is no significant difference in the average entropy values of the fundus images enhanced using NLTVR and CLAHE methods.

As the name suggests, CFP is a color image (RGB) and it is affected by contrast degradation more than the noise. The above mentioned qualitative and quantitative analysis are in favour of the proposed retinex based non-local total generalized variation restoration model to address the contrast degradation in fundus images. Unlike CFP, OCT images are grayscale and from the working principle of OCT devices and existing literature, it is evident that the speckle is inherent. Despeckling algorithm must adapt to the data distribution and therefore a different preprocessing approach is designed for OCT images. In the next section, the design of a non-local deep image prior model for restoring OCT images is elaborated.

### **3.4 NON-LOCAL DEEP IMAGE PRIOR**

Deep learning based restoration models are limelight of research in recent years. Compared to traditional methods like filtering and spatial transformation based methods, the deep learning models have the ability to automatically learn the features necessary to map the given input to the output. This rules out the need of prior knowledge to some extent. However, incorporating explicit prior, ensures the model learns right features

and the model can produce reliable and stable results. In the deep learning approach, the input to the model will be noisy image ' $U_0$ '. Let ' $\hat{U}$ ' be the predicted denoised image from the deep learning model after a particular set of iterations. The traditional transfer learning methods of despeckling are primarily based on reducing the error or the dissimilarity between the groundtruth clean image and the generated image. Thus the model learns the hyperparameters that have to be applied on the unseen test image from the training process. However, in the concept of deep image prior (elaborated in Appendix B.1), instead of using plenty of training images, the training image is kept constant, and the hyperparameters are searched during the learning process. Since the ground truth image is not available, it is assumed that the process of mapping the reconstructed image always passes closely to the domain of the clean image, as illustrated in Appendix B.1. Therefore, the training is halted when an optimal result (visually pleasing) is obtained. This way, the deep image prior model differs from traditional models by eliminating the need for groundtruth images and the need of plenty of training images. Mathematically, it represents mapping of  $\hat{U}$  and  $U$  such that  $f(\theta) = U$ , where  $\theta$  is hyperparameterization. At every iteration,  $\theta$  is searched, and the corresponding output image is reconstructed before computing the loss between them, forming an indirect mapping to traditional loss function. The idea of DIP is to use implicit regularization. Further Fan et al. (2020) used explicit regularization that minimizes the self-similarity of the difference image. This is elaborately explained below.

Initially, at each iteration (epoch of training a deep learning model), a difference image  $d$  between the predicted image ' $\hat{U}$ ' and input noisy image is computed as:

$$d = |U - \hat{U}|. \quad (3.41)$$

Then, the similarity between patches of the difference image is estimated using block matching algorithm, as:

$$\rho_{d_i, d_j} = \sigma_{d_i d_j} / \sigma_{d_i} \sigma_{d_j}. \quad (3.42)$$

where  $\rho$  represents the correlation coefficient and  $\sigma$  is standard deviation, ' $d_i$ ' and ' $d_j$ ' are patches centered at  $i$  and  $j$  respectively. Since despeckling is a process of minimizing the dissimilarity, the uncorrelation factor is estimated as follows:

$$\rho^+(\hat{U}_i) = 1 - \frac{1}{l} \sum_{k=1}^l (\rho_{d_i, d_j})_k. \quad (3.43)$$

Furthermore, the uncorrelated values are sorted and the first few values are multiplied with original image intensity value. This is given as:

$$\rho_s^+(\hat{U}) = U * \rho^+(\hat{U}_i) \quad (3.44)$$

Combining the explicit regularization with the mean square error term for image reconstruction, the objective function proposed by Fan et al. (2020) is:

$$\mathcal{L} = \underset{\theta}{\operatorname{argmin}}(\|U - \hat{U}\|_2^2 + \lambda \|1 - \rho_s^+(\hat{U})\|_2^2). \quad (3.45)$$

As the mean square error term may not address the speckle property, the proposed non-local deep image prior model is tuned to handle the Gamma noise. The modified objective function is given as:

$$\mathcal{L} = \underset{\theta}{\operatorname{argmin}}(\|\log \hat{U} + \frac{U}{\hat{U}}\|_2 + \lambda \|1 - \rho_s^+(\hat{U})\|_2^2). \quad (3.46)$$

where  $\lambda$  acts as a regularization parameter and controls the effects of over-smoothing,  $\theta$  denotes the hyperparameterization (weights and bias of the model). The first term is specific to denoise the Gamma noise derived using Bayesian MAP as explained in section 3.2, and the second term is to estimate the similarity of the reconstructed image.  $\lambda$  is the regularization parameter. The deep learning model in proposed approach is a U-net (following the similar architecture as given by Fan et al. (2020)). It has a series of convolution layers doubling at each layer-16, 32, 64, and 128 with a filter size of  $3 \times 3$ . A stride of 4 is set at the last layer. The overall depth of the network is 4 layers. The generic architecture diagram is given in Figure 3.12. Adam optimizer is employed to solve the above objective function and the number of iterations vary between 1000 and 2000. Based on the empirical analysis,  $\lambda$  value in equation (3.29) is set as 0.25 for most of the images as it was providing considerably good restoration of the input data.

### 3.4.1 Experimental results

The implementation of proposed non-local deep image prior model to despeckle OCT images is implemented using Pytorch and python. The machine type is n1-standard-8 (8 vCPUs, 30 GB memory), with 1 NVIDIA Tesla V100 GPU created using Google Cloud Platform. The CPU configuration employed is Intel Xeon E5-2698 v4 @2.2GHz, with 51.2MB cache size, 20 cores, and has 265GB RAM. Apart from OCT image, one synthetic image is also considered to estimate the performance, as shown in Figure 3.13(a).

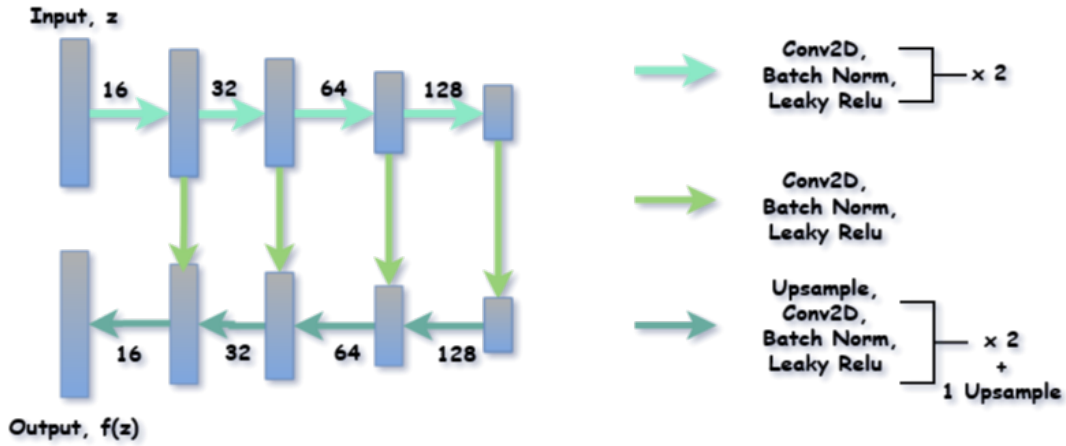


Figure 3.12: Architecture of proposed non-local deep image prior model.

A clean image represents the original synthetic image before incorporating the noise. For a normalized image of size  $128 \times 128$ , the speckle noise is introduced with mean 1, and the variance 0.1 (Figure 3.13(b)). To handle the division by zero error, a small constant of 0.0001 is added wherever necessary. The proposed approach produces the output as shown in Figure 3.13(d). These images show that the proposed approach generates a clean image relatively better than the non-local deep image prior method shown in Figure 3.13(c) (Fan et al. (2020)). Some of the related works are chosen for quan-

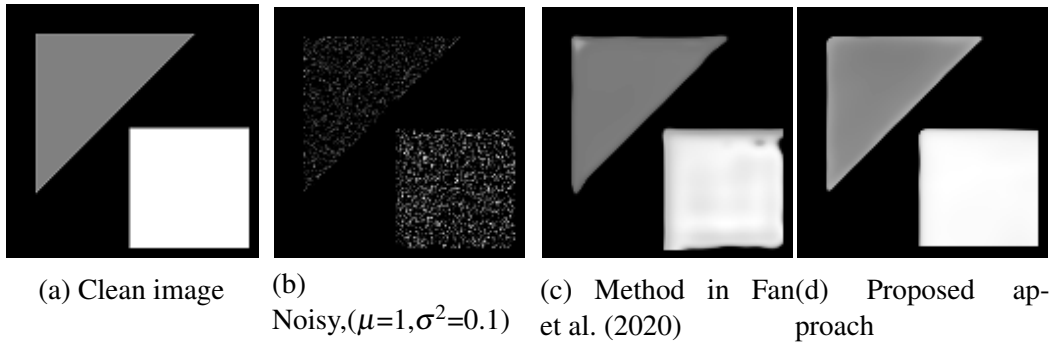


Figure 3.13: Results of despeckling synthetic image.

titative analysis of the proposed approach. Accordingly, methods discussed in Jidesh and Banothu (2018), Muxingzi et al. (2017), Ulyanov et al. (2020), Fan et al. (2020), and Sukesh and Sivaswamy (2018) are henceforth referred as "*method\_1*", "*method\_2*", "*method\_3*", "*method\_4*", and "*method\_5*" respectively. The quantitative metrics used to assess the image quality are defined in Appendix D.1. The PSNR value recorded after despeckling the synthetic image using various methods is plotted using the bar

chart in Figure 3.14. A higher value of PSNR indicates a better image quality. From this Figure, it is clear that the proposed approach performs better despeckling under the multiplicative noise framework. SSIM (Wang et al. (2004)) is another metric for measuring the similarity between the denoised image and a clean image. It is used to measure the luminance, contrast, and structure of two different images. The value of SSIM for normalized images ranges between 0 and 1. Index value 0 indicates that the two images are not similar, while a value of 1 represents the two images are perfectly similar and the structures are well preserved. Since the expected clean image is available for synthetic image, SSIM is measured, and the plot of the same is given in Figure 3.15. The bar chart depicts SSIM value close to 1 for the proposed approach. Therefore, the noise reduction and details preservation of the proposed model is duly demonstrated in these figures and an inference about the efficiency of the model under consideration is compared therein. Visual results for despeckling OCT images using

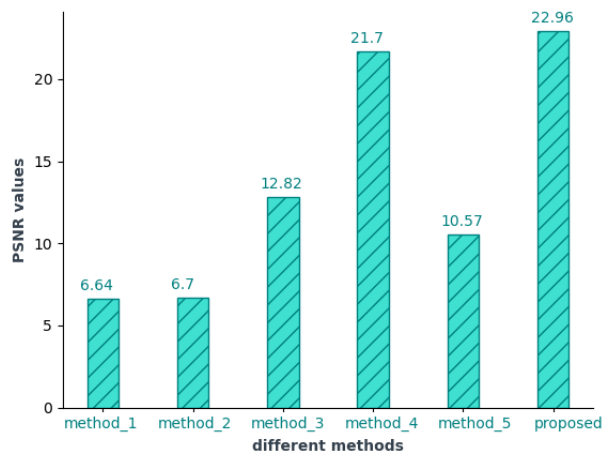


Figure 3.14: PSNR values after restoring synthetic image shown in Figure 3.13 (b).

the proposed approach are depicted in Figure 3.16. The first row represents the input images. Successive rows display the best results of despeckling using mean square error (MSE) loss function (Ulyanov et al. (2020)) and the proposed Gamma loss function, respectively. It can be observed that the retinal layers are well preserved using the proposed approach. The progressive transformation of noise vector in latent space to the denoised image is given in Figure 3.17. The reference input for this is the image in Figure 3.16(c). As it can be seen beyond 1000 iterations, the predicted image resembles the input noisy OCT image. Similarly, multiple images were analysed, which

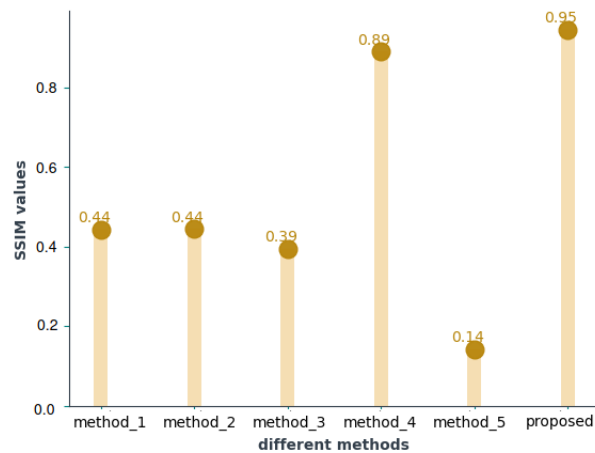


Figure 3.15: SSIM values after restoring synthetic image shown in Figure 3.13 (b).

were acquired from public domain (Fang et al. (2012)). For every image, the number of iterations needed to obtain visually appealing results is varied. Table 3.1 summarizes the PSNR values obtained for all three images shown in Figure 3.16. The PSNR was tracked during every iteration, and the final highest value obtained is recorded. It is evident from this table that the proposed approach is on par with existing methods. The

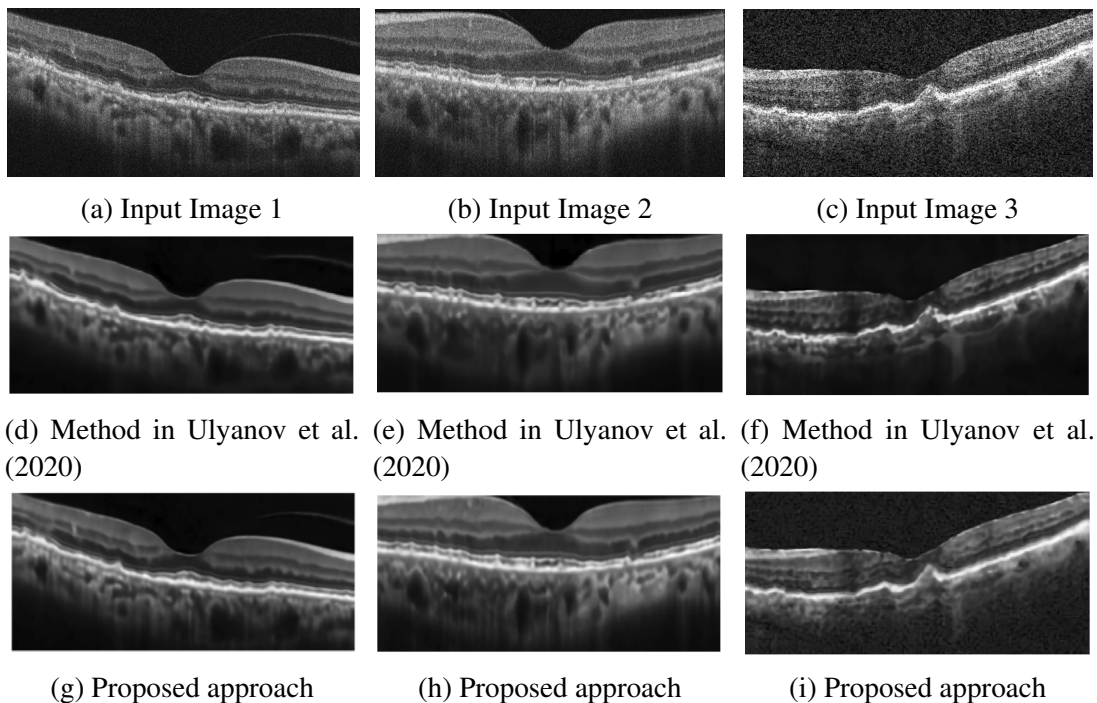


Figure 3.16: Three different OCT images are shown in row 1 from left to right. Second row corresponds to the deep image prior despeckling using mean square error loss function. The best output obtained using proposed gamma loss formulation is depicted in row 3.

Table 3.1: PSNR values (in dB) of OCT images.

Image no.	Method in Ulyanov et al. (2020)	Proposed Method
1	26.43	26.65
2	25.795	26.10
3	21.756	22.031

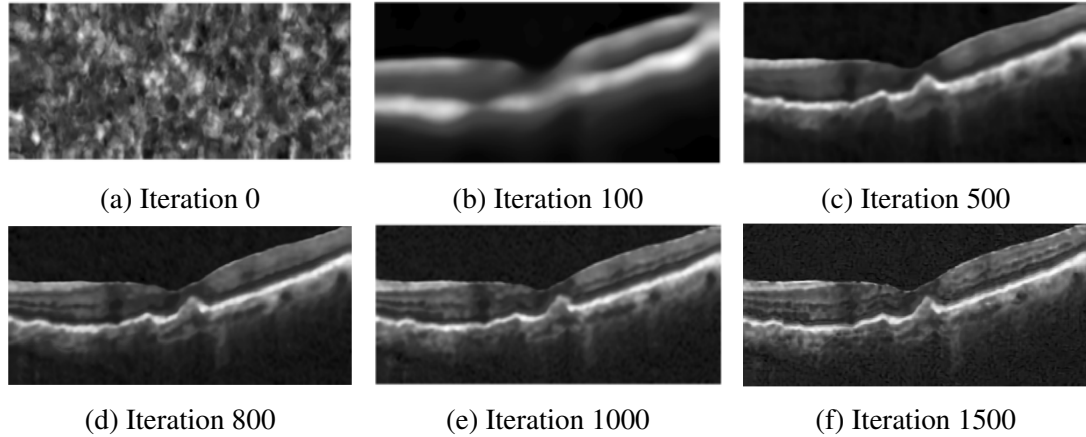


Figure 3.17: Progression of despeckling using Gamma loss framework for given input image. The noise in iteration 0 is transformed to a cleaner version of the input image by 1500 iterations.

proposed method is compared visually with some of the existing works, as mentioned earlier. The output obtained for input image Figure 3.16(b) using various approaches is depicted in Figure 3.18. To highlight the clarity of retinal layers, the output image is cropped around the fovea region. Here, Figure 3.18(a) represents the input noisy OCT image. Outputs shown in Figures 3.18(b), 3.18(c), 3.18(d), 3.18(e), and 3.18(f) corresponds to "*method\_1*", "*method\_2*", "*method\_3*", "*method\_4*", and "*method\_5*". The proposed framework given by (3.29) is coined as "*proposed*" henceforth, and the corresponding result is given in Figure 3.18(g). From these figures, we can observe that the image enhancement is qualitatively better in Jidesh and Banothu (2018). However, the gray level values are chopped out in the resultant image making it more closer to a bi-level image. The methods Ulyanov et al. (2020) and Fan et al. (2020) over-smooth the retinal layers. It is because the stopping condition is not defined by any constraint. The proposed approach preserves the retinal layers and abnormalities relatively better than the comparative methods while despeckling the input OCT image. Statistical analysis of the proposed approach is performed using the standard image quality metrics,

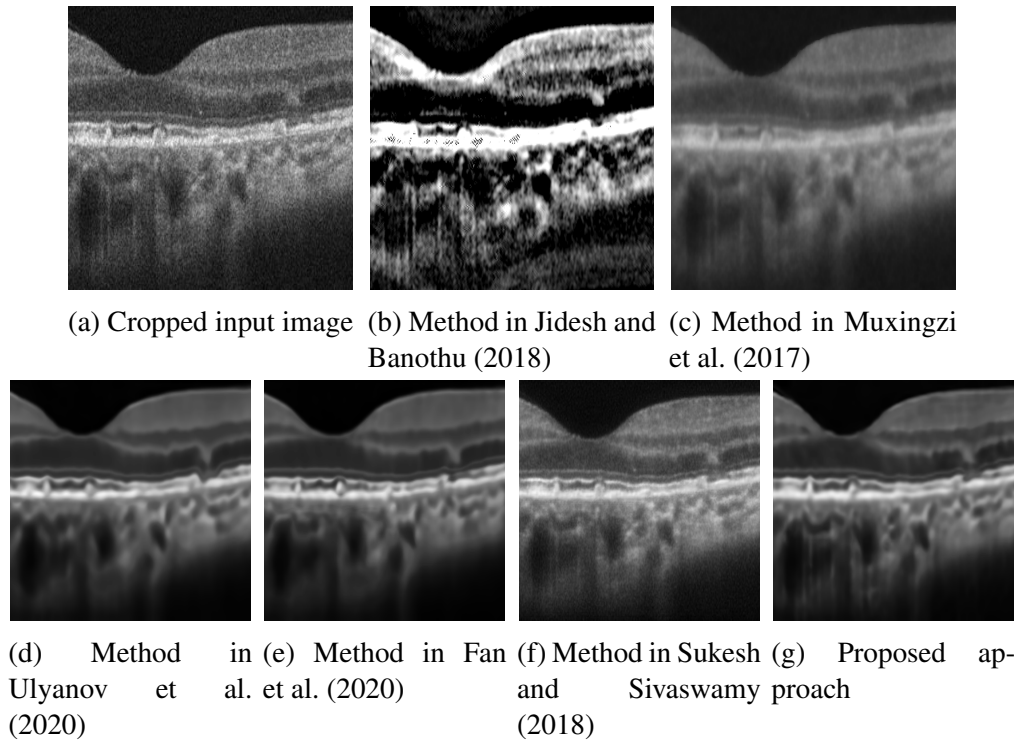
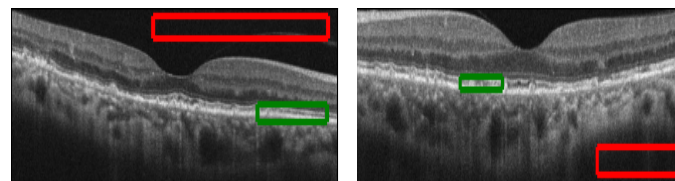


Figure 3.18: Partially cropped output images obtained using various methods corresponding to input image (Figure 3.16(b)) to visually highlight the retinal layer region.

namely, Contrast Noise Ratio (CNR) and Equivalent Number of Looks (ENL), which are most widely used, as discussed in the overview section. CNR is an image quality assessment metric that is similar to SNR, but it eliminates the bias influence. Ideally, a higher CNR value signifies a better contrast ratio in the image. ENL is a statistical measure that is computed as a ratio between the mean and the variance. Since the noise in an image implies high variance, the higher value of ENL signifies a cleaner image. Further, to evaluate CNR, homogeneous patches are randomly selected from the output image obtained using different methods ("*method\_1*", "*method\_2*", "*method\_3*", "*method\_4*", and "*method\_5*"). Two such homogeneous region selections are depicted in Figure 3.19 (for input image Figures 3.16(a) and 3.16(b)). As shown here, the red block indicates the background region with dark pixels, and the green block denotes the foreground region with bright pixels. Although multiple images and multiple homogeneous regions were analyzed experimentally, the results of randomly selected four patches are depicted in this section. It is further noted that the results obtained for the other patches follow a similar pattern. Figure 3.20 is a graph plot of CNR values in



dB. In the Figure, the horizontal axis denotes various patches. In other words, the notations  $im1\_p1$ ,  $im1\_p2$ ,  $im2\_p1$ , &  $im2\_p2$  represent set of background and foreground patches (p1, p2) selected from the set of images (im1, im2). It is ensured that the same patch is considered across all output images obtained using different methods. The bars represent various methods ( $method\_1 - 5$ ) along with the proposed one. Clearly, the CNR value of the proposed method is higher for most patches.  $Method\_4$  also provides competitive results. The ENL values obtained for randomly selected homogeneous patches are obtained and presented in the form of a bar graph in Figure 3.21. As higher value indicates better image quality, the proposed method projected higher ENL values compared to  $method\_4$ . It was observed that, for most patches considered, the statistical values were in similar scale. In a nutshell, the proposed model outperforms other comparative models in terms of both visual and statistical quantification as observed from the comparative results. The performance of the model is well justified by the theoretical design of the model. Unlike most other works in the literature, the proposed one duly considers the distribution of noise in the input data, which accounts for its better performance in the comparative analysis.



(a) Patches selected from in- (b) Patches from input image  
 put image (Figure 3.16(a)) (Figure 3.16(b))

Figure 3.19: Patches of two images, red box indicates the background region and green box indicate the foreground region. Here red block and green block together is denoted as a patch (p1 or p2).

### 3.5 SUMMARY

Restoration of degraded CFP and OCT images is the fundamental requirement, prior to the development of automatic disorder detection models. As stated in Chapter 2, existing restoration methods tend to ignore the noise distribution in the image. Moreover, these methods might result in over-smoothing when tested on images from multiple repositories.

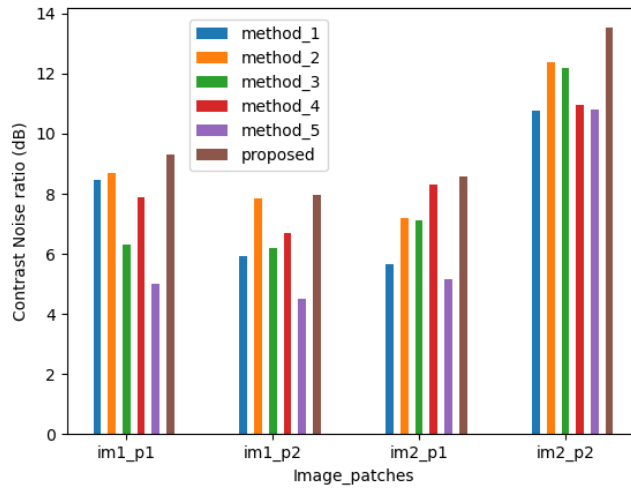


Figure 3.20: Average Contrast Noise Ratio for the input images shown in Figure 3.16.

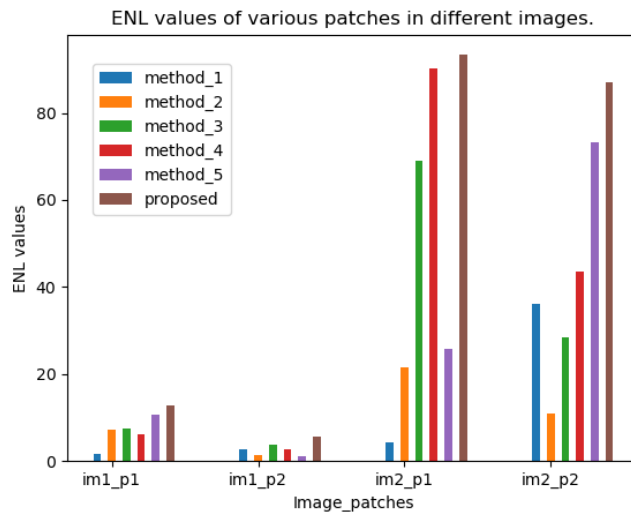


Figure 3.21: Average ENL values for the input images shown in Figure 3.16.

To begin with, the Chapter presents a qualitative comparison of CFP and OCT images available in the public domain. Further, a statistical method involving hypothesis testing and distribution fitting is explained, that can be applied to assess the probability of data distribution in OCT images. The retinex based non-local total generalized framework for contrast enhancement is proposed to address the illumination inhomogeneity of fundus images. The results of the proposed approach is quantitatively analyzed using metrics such as contrast and entropy. Though the proposed NLGVR method is computationally expensive with patch based analysis, the processing time is

boosted using SB approach and GPU. From the result analysis, it is evident that the proposed NLTGVR outperforms CLAHE in terms of detail preservation and contrast enhancement. The GCF has improved due to the proposed non-local image restoration methods. The limitation of the proposed NLTGVR is that it involves numerous regularization parameters that are decided empirically for every input image. Therefore, as a scope of future work, an algorithm has to be developed that can automatically estimate the regularization parameters. As an extension of the proposed NLTGVR method, the OCT images are despeckled and enhanced by adding an appropriate noise suppressing term related to Gamma distribution. The details of this method is published in a journal<sup>1</sup>.

Despeckling OCT images is a daunting task as the speckle is multiplicative in nature. In this Chapter, a novel deep image prior based non-local image restoration is proposed to restore degraded OCT images. Qualitative and quantitative analysis confirms that the proposed model effectively removes speckle from the input images. The major benefit of the proposed method is that it does not require ground-truth clean images. Though this is a deep learning based method, the model is trained using single input image. A skip connection based U-net is employed as the deep learning architecture. The method is tested on OCT images of different datasets and also on synthetic images. The proposed method is compared with other despeckling methods. The results of this work is published as a research article<sup>2</sup>. It is evident that the PSNR and SSIM significantly improved when the proposed method is applied on input synthetic images. As the proposed work performs a sole task of despeckling, future works include extension of the deep image prior models for contrast enhancement using retinex framework. The limitation of the work is that the number of iterations needed to obtain visually pleasing results is intuitively decided using empirical analysis. Nevertheless, certain techniques to identify an automatic stopping criteria is proposed in recent years. It is based on the assessment of the image quality metric such as PSNR in every training iteration. The

---

<sup>1</sup>A. Smitha, I.P. Febin, P. Jidesh, *A retinex based non-local total generalized variation framework for OCT image restoration*, Biomedical Signal Processing and Control, Volume 71, Part B, 2022, 103234, ISSN 1746-8094, <https://doi.org/10.1016/j.bspc.2021.103234>.

<sup>2</sup>A. Smitha & P. Jidesh (2021) *A nonlocal deep image prior model to restore optical coherence tomographic images from gamma distributed speckle noise*, Journal of Modern Optics, DOI: 10.1080/09500340.2021.1968052.

fundamental idea is to train the model until no further improvement is observed in the image quality. Integrating this aspect to the proposed approach is taken up as a future work.

To summarize, the first research objective is mapped to this Chapter. The proposed method answers partial research questions as follows. The retinex based framework effectively enhances the quality of CFP images. The speckle in OCT images follow Gamma distribution. The deep image prior model can be successfully employed to despeckle OCT images. The related works in this direction are reviewed in the previous Chapter. The effect of applying the proposed preprocessing models for automatic retinal disorder classification, will be discussed in the next Chapter.

## Chapter 4

# GENERATIVE ADVERSARIAL NETWORKS FOR AUTOMATIC RETINAL DISORDER DETECTION

This chapter presents novel GAN models to detect the retinal disorders from pre-processed CFP and OCT images. Images from public repository are used to train the GAN model. Some of the relevant works described in Chapter 2 are chosen for comparative study. The rest of the Chapter is organized as follows. The list of datasets used in the proposed work is described in section 4.1. The preliminary details of GAN and its variants is provided in section 4.2. The proposed deep learning architecture for CFP and OCT image analysis is illustrated in sections 4.3 and 4.4, respectively. Section 4.5 is reserved for experimental results and discussion. Finally, a summary of the Chapter is presented in section 4.6. Refer Appendix C.3 for a detailed explanation of various "*layers*" and "*hyperparameters*" used in the deep learning architecture.

### 4.1 DATASETS

Multiple publicly available fundus image datasets are acquired from various sources as shown in Table 4.1. Based on the available ground truth data, the entire repository is divided into three groups as three different tasks are performed. Images from Retinal Fundus Glaucoma Challenge (REFUGE) (Orlando et al. (2020)) and Indian Diabetic Retinopathy Image Dataset (IDRiD) (Porwal et al. (2018a,b, 2020)) grand challenges are considered as first group. These are labeled datasets and have groundtruth data for fovea localization and optic region segmentation. The second group of images in-

cluded Mendeley-2020 (Akram et al. (2020)), IOSTAR-2016 (Zhang et al. (2016)), High-Resolution Fundus (HRF)-2013 (Budai et al. (2013)), CHASE-DB-2012 (Fraz et al. (2012)), Digital Retinal Images for Vessel Extraction (DRIVE)-2004 (Staal et al. (2004)), and STructured Analysis of the Retina (STARE)-2000 (Hoover et al. (2000)). These datasets purely provide the gold standard manual segmentation of retinal blood vessels. Therefore, a combined group of images are used for the blood vessel segmentation task. The final group of images from the ODIR-2019 dataset is used for classification. Data augmentation on the fly is incorporated in every phase to increase the number of samples. Random rotation by 2 degrees, horizontal and vertical flip, and brightness variations are the augmentation considered. All images are normalized and mean-centered before feeding it to the model. In order to train a deep learning model, it is necessary to maintain uniform image resolution for every batch of input training images. After removing the outliers, the average resolution of images is found to be  $1538 \times 1394$ . However, resizing all images to this size and training the model requires massive resources. To ensure optimal memory usage, the images are resized into  $512 \times 512$ .

The OCT datasets used are listed in Table 4.2. It must be noted that, compared to the CFP images, the number of OCT datasets available in public domain is smaller. Moreover, the groundtruth segmentation of OCT images is unavailable in public domain. Therefore, the OCT images are not subdivided for multiple tasks. Training the GANs using heterogeneous datasets enables the model to learn multiple features from the retina, irrespective of the source.

## 4.2 GENERATIVE ADVERSARIAL NETWORKS-PRELIMINARIES

A GAN (Goodfellow et al. (2014)) consists of a generator ' $G$ ' and a discriminator ' $D$ ' modules as shown in Figure 4.1. The generator tries to generate the fake images while the discriminator tries to distinguish between real and fake. The generator maps the noise vector drawn randomly from a probability distribution (Gaussian), to a distribution similar that of the original input. The discriminator accepts samples of real input

Table 4.1: List of CFP datasets in public domain.

Sl. No.	Challenge-Year	No. images	Description	Size	Format
1	REFUGE-2018 & 2020	2000	Grand challenge that involves three tasks: Classification of Glaucoma and normal, segmentation of Optic Disc region and Fovea localization. images are acquired from 2 types of devices.	2056 × 2124	jpg
2	IDRiD-2018	400	Images of diabetic retinopathy grading were acquired from Eye Clinic in Maharashtra, India, using a Kowa VX-10 alpha digital fundus camera with 50-degree field of view. The challenge tasks are similar to REFUGE.	1204 × 1500	jpg
3	Mendeley-2020	100	Images are taken from Armed Forces Institute of Ophthalmology (AFIO), Rawalpindi, Pakistan. Annotated arteries, veins and combined vessels are provided along with fundus images	1504 × 1000	jpg
4	IOSTAR-2016	30	The images in the IOSTAR vessel segmentation dataset are acquired with an EasyScan camera (i-Optics Inc., the Netherlands), which is based on a SLO technique with a 45 degree Field of View (FOV)	1024 × 1024	jpg, tif
5	HRF-2013	45	Fundus images of 15 healthy patients, 15 diabetic retinopathy and 15 glaucoma patients are acquired using a Canon CR-1 fundus camera with a field of view of 45 degrees. Binary gold standard vessel segmentation images are available for each image.	3504 × 2336	jpg
6	CHASE-DB-2012	28	Retinal image analysis is the major focus of this project. Blood vessel manuals for 14 left and 14 right eye fundus images are provided in this repository	1280 × 960	jpg
7	DRIVE-2004	40	33 normal and 7 mild diabetic fundus images and corresponding manual segmentation is available as ground truth. Images are acquired from The Netherlands. Masks to obtain the region of interest is also provided.	768 × 584	tif
8	STARE-2000	20	Images acquired from Shiley Eye Center at the University of California, for blood vessel segmentation from retinal fundus images. Manually annotated blood vessels are provided.	700 × 605	tiff
9	ODIR-2019	7000	Images are collected from various hospitals by Peking University in China. At present, this is the largest diverse repository of fundus images available publicly for classification. However, only 1000 images, belonging to glaucoma and normal condition are used in our work.	1800 × 2400	jpg

Table 4.2: Details of OCT repositories in public domain.

Dataset	Description	Year
Dataset 1, (Reza et al. (2018))	OCT-B scans belonging to three categories, namely: AMD, DME, and NORMAL. Images are acquired using Heidelberg spectral domain-OCT This dataset consists of about 3980 OCT-B scan images.	2017
Dataset 2, (Kermany et al. (2018a))	This dataset has 83,605 OCT images with 4 categories, namely: CNV, DME, NORMAL, and DRUSEN. Images are acquired from various hospitals	2017
OCTID - Dataset 3, (Gholami et al. (2020))	The OCT images are acquired from Carl Zeiss, High Definition OCT device. It consists of 59 images corresponding to Age-related Macular Degeneration and 205 RGB images belonging to Normal Categories.	2019

and the generator output, thereby working as a binary classifier to predict whether the input maps to real data or fake data. Since the generator learns the pattern in input data

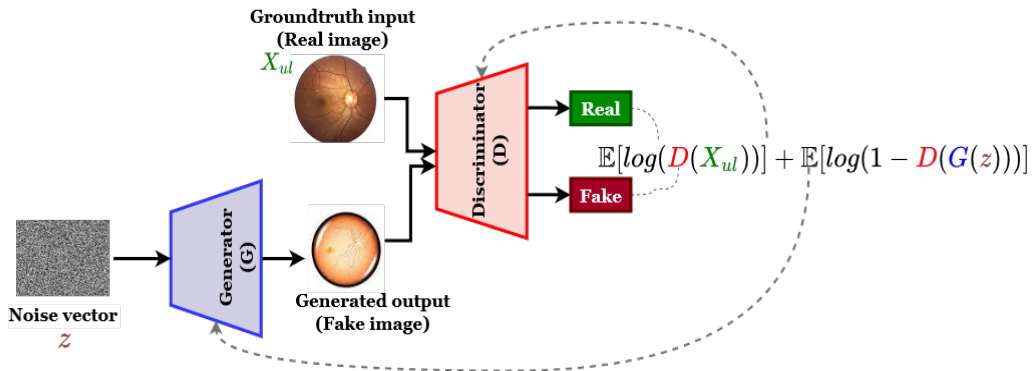


Figure 4.1: Basic architecture of GAN.

without any other information, the generative modeling is also referred as an unsupervised learning. During the training process, both  $G$  and  $D$  are trained simultaneously. In other words, for every batch of input data, the generator generates certain samples. The generated samples are fed to the discriminator. The error in prediction is then computed using the real data. Initially, the discriminator will be able to clearly identify the fake data. Hence, the discriminator weights will be rewarded. On the contrary, the generator weights are penalized to ensure the generator generates realistic samples in the next batch. Gradually, the discriminator fails to distinguish the fake samples and it will then be penalized to improvise its prediction. This scenario of a zero-sum game is referred as adversarial learning. Finally, when the training is complete, the  $D$  is discarded and only trained  $G$  is used for further analysis. The trained features of  $G$  can be used for



applications such as transfer learning. As illustrated by Goodfellow et al. (2014), the adversarial loss of a GAN architecture is given as:

$$\mathbb{E}[\log(D(X_{ul}))] + \mathbb{E}[\log(1 - D(G(z)))]. \quad (4.1)$$

where ' $X_{ul}$ ' represents the set of input images from sample dataset, ' $\mathbb{E}$ ' is the expected value over all real data instances,  $G(z)$  is generator's output with noise vector  $z$  as input, and  $D(X_{ul})$  is the prediction of the discriminator when real images are fed as input. The  $\log$  term refers to the standard log operation with base 10 throughout the rest of the document unless explicitly mentioned with a different base. However, replacing the  $\log$  term with natural log ( $\ln$ ) does not alter the performance of the model. The generator tries to minimize the component  $\log(1 - D(G(z)))$  in order to fool the discriminator. The discriminator tries to maximize the  $\log(D(X_{ul}))$  term. Hence, this loss function is termed as a minimax loss function.

Training a GAN is challenging compared to the traditional CNNs, since the weights associated with  $G$  and  $D$  are updated using the same batch of images sequentially. Moreover, the performance depends on activation function used, the depth of the network and the number of training samples. Some of the definitions of activation functions is given in Appendix C.3. To ease the effort involved in training a GAN model and to boost the performance, specific best practices are proposed in the recent literature (Salimans et al. (2016)). Several variants of GANs are introduced in recent years namely, Auxiliary Classifier (AC-GAN) (Odena et al. (2017)), Pix-to-pix GAN (Isola et al. (2017)), Wasserstein Generative Adversarial Network (WGAN) (Arjovsky et al. (2017)), semi-supervised GAN (Odena (2016)), and so on (Jason (2019)). A semi-supervised GAN works as a classifier when partially labelled dataset is available. The model learns the features from unlabeled dataset and it learns the labels to predict from limited labelled dataset. The discriminator in a semi-supervised GAN is transformed to a multi-class classifier as it classifies the input image into ' $K$ ' different categories (Odena (2016)) in addition to the discrimination between real and fake samples. This is achieved by extending the number of classes from  $K$  to  $K + 1$ , where the extra class denotes the fake category. The ' $D$ ' of a traditional GAN accepts two inputs - real image and the generator output. Unlike these traditional GAN, the ' $D$ ' in a semi-supervised GAN ac-

cepts three inputs - labeled images, unlabeled images, and the generator output. This aids the discriminator to learn features from broader perspective and therefore, it need not be discarded after training. The discriminator now has two branches: one branch is called supervised since it has to compute the loss for multiple class label prediction, and the other is an unsupervised branch that discriminates real and fake using sigmoid activation function. The combined objective function, motivated by Odena (2016), for the semi-supervised GAN is:

$$\mathbb{E}[\log(p(Y_l|X_l, Y_l < K + 1))] + \mathbb{E}[\log(1 - p(Y_l = K + 1|X_l))] + \mathbb{E}[\log(p(Y_{ul} = K + 1|X_{ul}))]. \quad (4.2)$$

where the first term contributes for the supervised classification into ' $K$ ' categories, second and third term monitors the distinction between real and fake samples of labeled  $(X_l, Y_l)$  and unlabeled  $(X_{ul})$  data samples, respectively. It is recommended by Salimans et al. (2016) that the dropout layers can be used extensively in a semi-supervised GAN as regularization to avoid over-fitting scenario. Considering the fact that limited groundtruth segmentation images are available from CFP modality, while relatively a larger number of images are available as labelled dataset for classification, the obvious choice to implement the dual task in the proposed work is semi-supervised GAN. However, the unavailability of groundtruth segmented images for OCT datasets motivates the adaption of WGAN for the proposed work.

WGAN is an extension of traditional GANs, that conceptually shifts the discriminator from discriminating between real and fake to grade the realness of the input image. The discriminator of WGAN is therefore called as a critic model. It works by measuring the distance between the generated distribution and the real distribution and improves the stability of training a GAN with limited data. The objective function of WGAN as recommended by Arjovsky et al. (2017) is expressed as follows:

$$\sup_{\|f\| \leq 1} \mathbb{E}[f(X_l)] - \mathbb{E}[f(z)], \quad (4.3)$$

where  $\mathbb{E}$  is standard expectation function,  $\sup$  is the least upper bound and  $f$  is 1-Lipschitz function satisfying the constraint:

$$|f(a) - f(b)| \leq |a - b|. \quad (4.4)$$

where  $(\cdot)$  represents the discriminator or the critic that predicts the output on real images

and the images generated by the generator. The stability of WGAN, lesser sensitivity to the model architecture, and hyperparameter configurations are the major benefits for choosing WGANs in proposed work of OCT image analysis.

### 4.3 GAN MODEL FOR FUNDUS IMAGES

Multiple feature extraction from retinal fundus images, like fovea, optic disc, and blood vessels, along with classification of retinal disorders is the significant contribution of the proposed work. The appropriate deep learning model to perform multiple tasks is GAN due to its adversarial nature. The training stage includes two GAN models, namely, VGAN for segmentation and semi-supervised GAN for classification as depicted in Figure 4.2. VGAN consists of a generator and a discriminator. The generator accepts

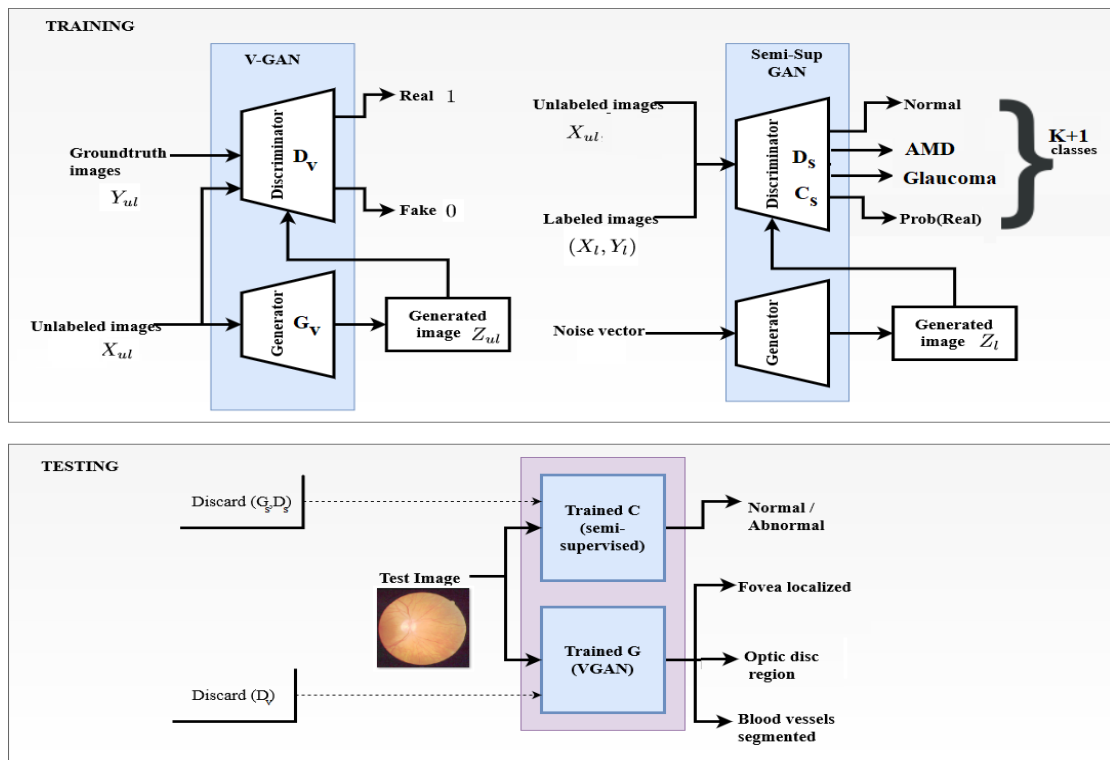


Figure 4.2: Block diagram for fundus image analysis.

input images and generates fake segmented images. Discriminator tries to distinguish between the ground truth segmentation images and fake images generated by the generator. The detailed architecture of the generator and discriminator used for segmentation is shown in Figure 4.3. In the proposed model, the number of filters begins from 16, and

doubles in every layer. It was observed from multiple trials that the number of filters less than or equal to 256 gives a good performance. Considering that  $G_V, D_V$  are generator and discriminator models of VGAN respectively, the overall objective function of VGAN is:

$$\mathcal{L}_{VGAN} = \arg \min_{G_V} \left[ \max_{D_V} \mathcal{L}_{GD} \right] + \alpha \mathcal{L}_{Seg}, \quad (4.5)$$

where  $\alpha > 0$  is balancing parameter to monitor the segmentation ( $\mathcal{L}_{Seg}$ ) and discrimination process. The first term corresponds to the minimax game of the VGAN, similar to 4.1, and is given as:

$$\mathcal{L}_{GD} = -\mathbb{E}[\log(D_V(X_{ul}))] - \mathbb{E}[\log(1 - D_V(G_V(X_{ul})))], \quad (4.6)$$

Considering the segmentation problem to be a mapping of input to groundtruth images, it can be associated with binary cross entropy loss as given as:

$$\mathcal{L}_{Seg} = \mathbb{E}[-Y_{ul} \log(G_V(X_{ul})) - (1 - Y_{ul}) \log(1 - G_V(X_{ul}))]. \quad (4.7)$$

In order to perform segmentation, only the trained generator of VGAN is used and the final discriminator is discarded after training.

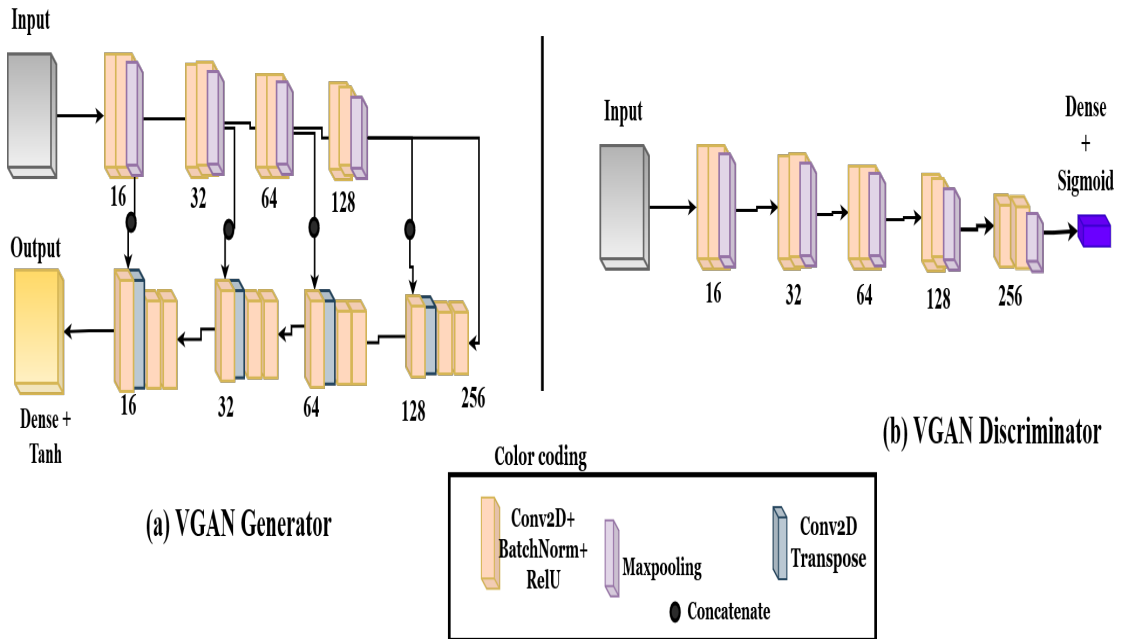


Figure 4.3: Architecture of VGAN.

**Fovea localization:** Consider a fundus image,  $U$  of size ( $width \times height$ ) picked from the REFUGE and IDRiD grand challenge datasets. Let the pixel location of fovea

be denoted as  $U(x, y)$ . A distance map is created by computing the Euclidean Distance of every pixel to this point, given as:

$$\sqrt{(x_i - x)^2 + (y_j - y)^2}; \quad (4.8)$$

$\forall(i, j)$  such that  $1 \leq i \leq width$  &  $1 \leq j \leq height$ . Since the distance of a point to itself is zero, it is the brightest pixel marked in white color, and as the pixel around this point is equidistant, the color gradually turns gray. The image is then normalized using:

$$U(x, y) = \left[ 1 - \frac{U(x, y)}{\max(U(x, y))} \right]^\alpha, \quad (4.9)$$

where  $0 < \alpha \leq 1$ . Such distance maps (feature maps) are used as ground truth data. One such feature map and its corresponding input fundus image are shown in Figures 4.4a and 4.4b, respectively. Now the GAN is trained to generate such fovea localized images for any given input image. Once the model is trained, for any given test fundus image, it generates a fovea distance map. The centroid of the blob region, in the generated feature map, is computed using:

$$\psi_{x,y} = \left( \frac{1}{width} \sum_{k=1}^{width} x_k, \frac{1}{height} \sum_{k=1}^{height} y_k \right). \quad (4.10)$$

**Optic disc segmentation:** For the optic disc, the ground truth images are provided by REFUGE and IDRiD grand challenge. Hence, the GAN model is trained using these ground truth images. For a sample given input image (Figure 4.4c), it generates the optic disc, as shown in Figure 4.4e. The corresponding ground truth is shown in Figure 4.4d. As a next step, the centroid of the blob region is computed using eqn. (4.10). This central pixel marks the center of the optic disc. A region of  $100 \times 100$  is cropped around this center from the original input image as shown in Figure 4.4f.

**Vessel extraction:** The groundtruth data from various resources are combined. The GAN is trained to segment the blood vessels, for any given fundus image. A sample input image, corresponding ground truth manual segmentation and generated images are shown in Figures 4.4g, 4.4h, and 4.4i, respectively.

**Classification:** Once the image is preprocessed, the entire dataset is distributed into three sets for training, validation, and testing in the ratio of 60:20:20 (heuristic ap-

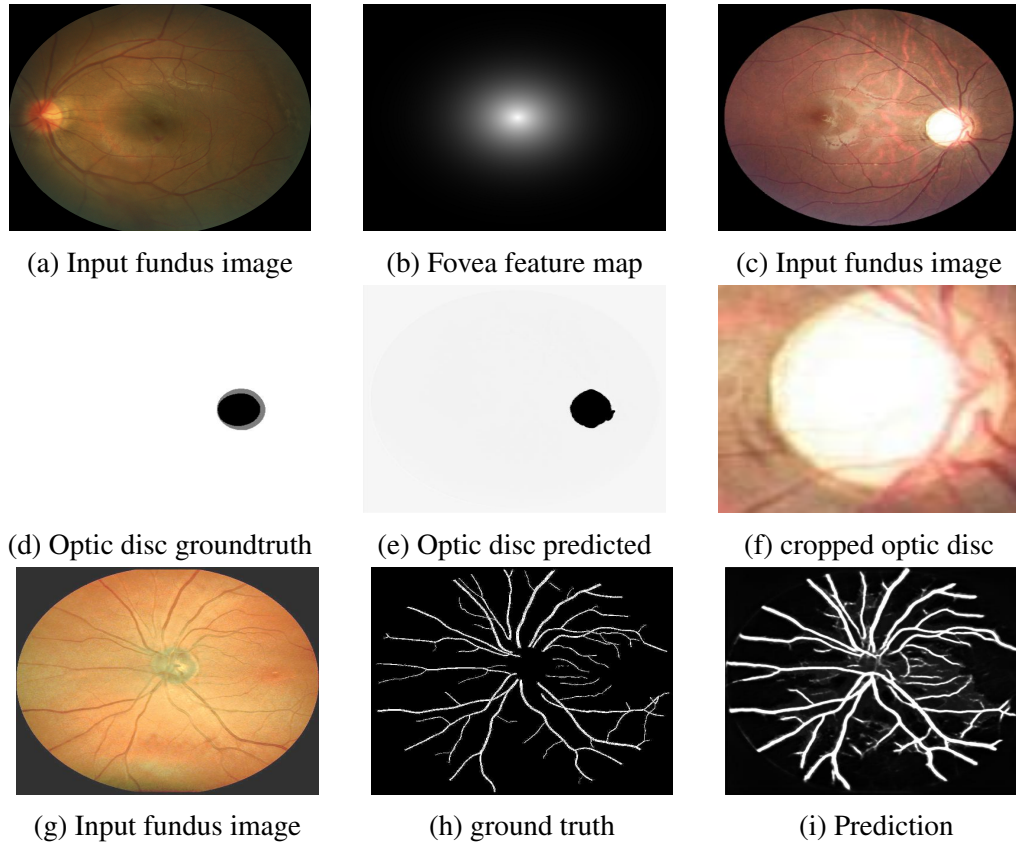


Figure 4.4: Segmentation of optic disc, fovea and blood vessels of randomly selected fundus image using GAN.

proach). The semi-supervised GAN model is trained with these images. It works with partially labeled data. The generator maps the noise vector to fake images that resemble the unlabeled data. The discriminator performs two tasks (as highlighted by Das et al. (2020)). The first task is to classify the labeled images into normal or abnormal ( $C_S$ ), termed as a supervised discriminator. The second task is to discriminate between real and fake images ( $D_S$ ), called as an unsupervised discriminator. This is achieved by extending the K-category classification problem into K+1 level classification. The additional level is obtained by normalizing the outputs of the final layer in the discriminator model. The architecture of the proposed semi-supervised GAN is depicted in Figure 4.5. The loss functions of  $C_S$ ,  $D_S$  and  $G_S$ , adapted from 4.2, are given as:

$$\mathcal{L}(C_S) = -\mathbb{E}[\log(p(Y_l|X_l, Y_l < K + 1))], \quad (4.11)$$

where,  $K$  = Number of classes.

$$\mathcal{L}(D_S) = -\mathbb{E}[\log(1 - p(Y_l = K + 1|X_l))] - \mathbb{E}[\log(p(Y_{ul} = K + 1|X_{ul}))], \quad (4.12)$$

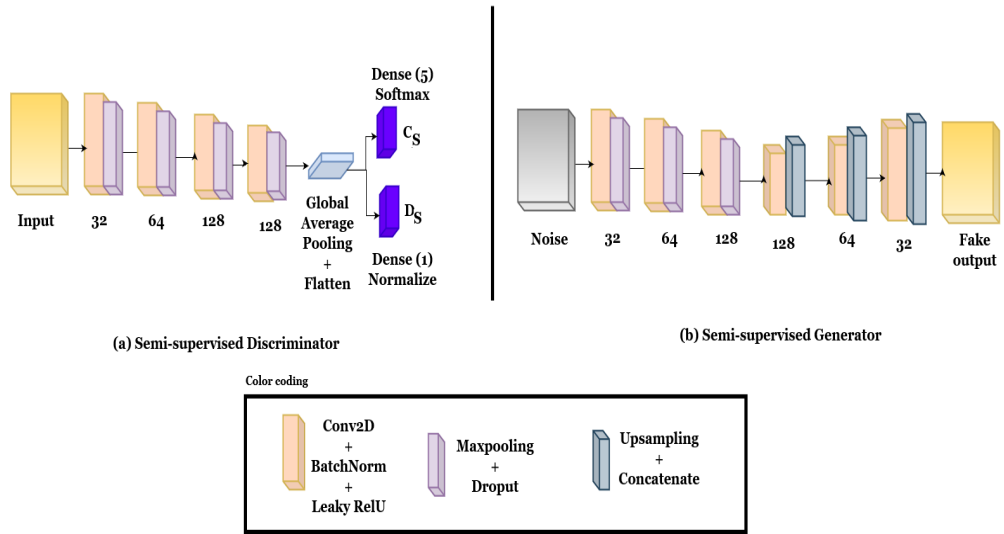


Figure 4.5: Semi-supervised GAN architecture.

where,  $p$  denotes the conditional probability. Once the semi-supervised GAN is trained only the supervised discriminator ( $C_S$ ) is retained to perform classification on different test datasets and the rest of the model is discarded. Throughout the GAN architecture, a stride of 2 and kernel size of  $5 \times 5$  is used. The batch size is set to 64. Adam optimizer and RMS-Prop are used as optimizer for discriminator and generator network, respectively.

#### 4.4 GAN MODEL FOR OCT IMAGES

As mentioned earlier in this Chapter, among the variants of GANs, WGANs is less sensitive to modification in hyperparameters and the number of layers giving a stable performance (Arjovsky et al. (2017); Gulrajani et al. (2017)). So, WGAN is used in the proposed method. Let ' $G_w$ ' and ' $D_w$ ' denote the generator and discriminator of a WGAN. As mentioned earlier, the optimization function of a WGAN is given in 4.3, where the constraint is enforced by 1-Lipschitz function. To enforce this, the weight values are clipped between a particular range of values  $(-c, c)$ . This is termed as clipping constraint. Additionally, a gradient penalty is introduced to ensure stable GAN training as proposed by Gulrajani et al. (2017). Further, as a modification to this WGAN cost function, an additional sparse categorical cross-entropy loss ( $\mathbb{L}_{\text{sparse}}$ ) is added, which is

defined by:

$$\mathbb{L}_{\text{sparse}} = \frac{-1}{S} \sum_{i=1}^S \log p(Y_i \in K_{y_i}), \quad (4.13)$$

where  $S$  denotes total number of samples,  $Y_i$  represents a particular class predicted among the total number of classes,  $K$  and  $p(\cdot)$  corresponds to the probability of obtaining a particular class. Combining 4.3, 4.13 and the gradient penalty (recommended by Gulrajani et al. (2017)), the final objective function of the proposed WGAN is given by:

$$\min_{G_w} \max_{D_w} \mathbb{E}[D_w(X_l)] - \mathbb{E}[D_w(G_w(X_l))] + \lambda \mathbb{E}[(\|\nabla D_w(r)\|_2 - 1)^2] + \mathbb{L}_{\text{sparse}}. \quad (4.14)$$

where ' $r$ ' denotes random samples picked from  $(x, \hat{x})$ . The first two terms in equation (4.14) represent Wasserstein loss for distinguishing real and fake images, the third term denotes the gradient penalty term added for stability and the last term represents the sparse categorical cross entropy. The block diagram representing the proposed methodology is given in Figure 4.6 which is self-explanatory. The number of classes are set as  $K = 3$  for Dataset 1,  $K = 4$  for Dataset 2, respectively. Since the number of images available in Dataset 3 is relatively smaller, these images are combined with Dataset 1 for AMD and Normal categories.

The retinal layers are marked on denoised OCT images using caserel tool (Computer-

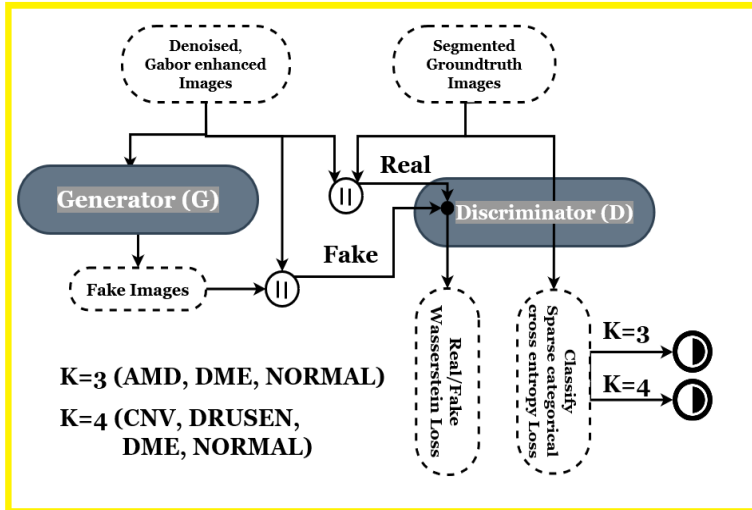


Figure 4.6: Block diagram for OCT image analysis.

Aided SEgmentation of RETinal Layers in OCT images) (Chiu et al. (2010); Teng (2013)). Further, in retinal disorders like AMD and DME, it is important to observe



the RPE layer where the drusen could be present and the inner limiting membrane layer where the edema could be monitored. Hence, the Gabor features (Weldon et al. (1996)) are extracted to enhance the input images before feeding them to the generator. Integrating handcraft features was attempted first by Li et al. (2019). Applying Gabor features greatly enhances the retina layers' visibility and improves the quality of fake images generated by the generator. The Gabor features aids to retain the texture changes on the retinal images and hence distinguishes the edges. Two dimensional Gabor filter is used to extract texture features. The denoised images are enhanced using Gabor filters and fed to the WGAN. The architecture of the generator and discriminator models are shown in Figure 4.7. The discriminator performs two tasks: distinguishing real and fake images and the classification of input images into normal or abnormal categories (K classes). The generator generates the OCT images. The abnormal categories include AMD, DME, and Normal, based on the availability of the public repositories. The input images are resized into  $256 \times 256 \times 3$ . The generator ' $G_w$ ' accepts the

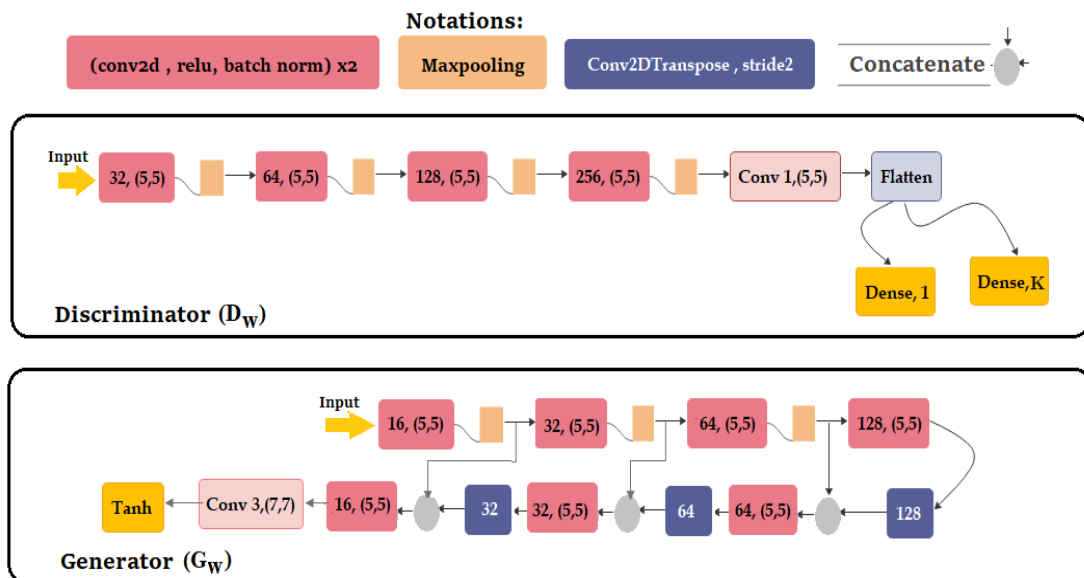


Figure 4.7: Architecture of Wasserstein GAN.

denoised, feature enhanced input images. It follows a U-net module with skip connections. The U-net is constructed by two consecutive convolution operations of  $5 \times 5$  kernel size, stride 1, and "relu" activation function, followed by batch normalization. After two convolution blocks, max-pooling operation is performed to reduce the size

of the feature map. This is repeated for 4 times, where each time the number of filters is doubled (16, 32, 64, and 128). Then the concatenation operation is performed after upsampling the image. A transpose of convolution operation is performed to achieve upsampling. The final activation function is "tanh" which follows a single convolution layer with kernel size  $(7 \times 7)$  to produce a 3 channel image. The discriminator ' $D_w$ ' accepts the original input images along with segmented data as shown in Figure 4.6). In other words, the concatenate operation is performed on input image and groundtruth segmentation images while training for real output. The input image is concatenated with the generator's fake image output while training the discriminator for fake image detection. It performs repeated convolution, batch normalization, and "relu" operation, with intermediate "maxpool" layers like the generator module. The final layer includes a flattening operation followed by a linear activation function to discriminate between real and fake images. Another output is a Dense connection of ' $K$ ' classes with a "softmax" activation function. The hyperparameters are chosen based on the recommendations given by Arjovsky et al. (2017); Gulrajani et al. (2017) and through empirical analysis. It was observed that a deeper discriminator model boosts the performance of the classifier while a stronger generator model improves the quality of segmented images. Algorithm 2 summarizes the training steps in the proposed method. For the sake of simplicity, after the dataset is divided into three parts for training, testing, and validation, the following notations are used to denote the set of training and validation images.  $I_{tr}$ : Training images,  $S_{tr}, C_{tr}$ : Segmented images and labels, corresponding to  $I_{tr}$ , respectively,  $I_{val}, C_{val}$ : Validation images and its corresponding labels,  $B$ : Batch Size,  $E$ : Epochs,  $G_w, D_w$ : Generator and Discriminator of WGAN respectively.

---

**Algorithm 2:** Training and validation steps for the WGAN model.

---

**Input:**  $I_{tr}$ ,  $S_{tr}$ ,  $C_{tr}$ ,  $I_{val}$ ,  $C_{val}$ ,  $B$ , and  $E$ .**Output:** Trained models ' $G_w$ ' and ' $D_w$ '.

```
1  $BPE \leftarrow \frac{\text{Total number of training images}}{B}$ 
2 for  $i=1$  to  $E * BPE$  do
3   Load batch of images- $(I_{tr}, S_{tr}, C_{tr})$ 
4   Train  $D_w$  on batch of  $(I_{tr}, C_{tr})$  for classification
5   Compute sparse categorical cross entropy loss as in 4.13
6   for  $j=1$  to 3 do
7     Generate batch of fake images using  $G_w$ 
8     Train  $D_w$  on batch of  $(I_{tr} || S_{tr})$  with label as 1 (Real images)
9     Train  $D_w$  on batch of  $(I_{tr} || G_{im})$  with label as -1 (Fake images)
10    Compute gradient penalty with random samples of  $S_{tr}$  &  $G_{im}$ 
11    Compute total loss using 4.14 and update its weights.
12  end
13  Freeze the weights and bias of  $D_w$ , Train  $G_w$  on batch of  $I_{tr}$ 
14  Compute loss of  $G_w$  and update weights.
15  if  $i=BPE$  then
16    Validate  $D_w$  on batch of  $I_{val}, C_{val}$ 
17    Save Validation and training loss for plotting
18  end
19 end
```

---

## 4.5 EXPERIMENTAL RESULTS

Implementation is done using Keras (Tensowflow-2 backend) and python on Tesla V100 DGX station with 4 GPUs, each having a memory capacity of 32478MiB, clock @1.53GHz, and compute capability of 7.0. The CPU configuration employed is Intel Xeon E5-2698 v4 @2.2GHz, with 51.2MB cache size, 20 cores, and has 265GB RAM.

Table 4.3: Performance metrics for segmentation.

Task	Method	Performance parameter	Values
Fovea Localization	1. Distance regression Meyer et al. (2018)	Average Euclidean Distance	23
	2. Proposed method		<b>49</b>
Optic Disc Segmentation	1.U-net with CNN Singh et al. (2018)	Dice Coefficient	0.9340
	2. 57 layered CNN Agrawal et al. (2018)		0.88
	3. Proposed method		<b>0.9978</b>
Vessel Extraction	1. U-net model Son et al. (2017)	Dice Coefficient	0.834
	2. Proposed method		<b>0.872</b>

### 4.5.1 Fundus image classifier

Training the GAN model for every stage took around 14 hours when the task was scheduled on a single GPU. The segmentation output images obtained for a randomly picked fundus image from ODIR dataset is shown in Figure 4.8. Table 4.3 summarizes the

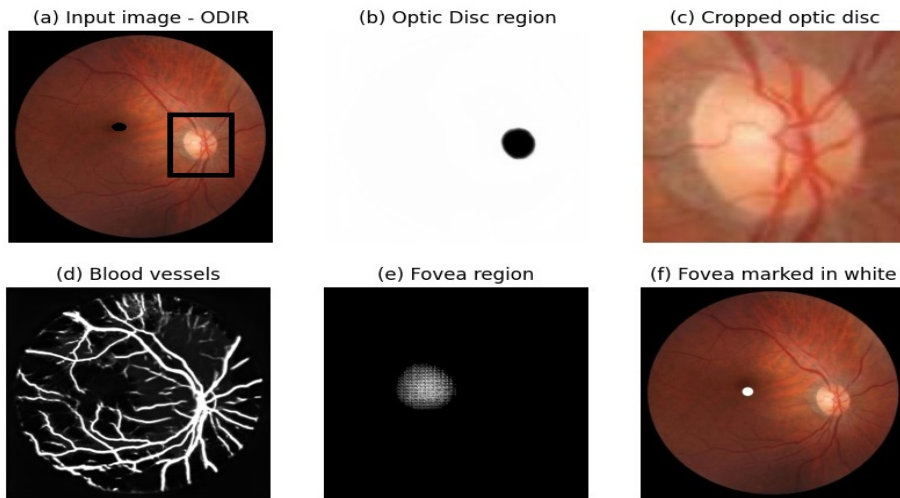


Figure 4.8: Segmented output for ODIR dataset.

segmentation results obtained using the proposed combined GAN model (marked in bold) and compares it with related works. Since REFUGE and IDRiD grand challenges provide the fovea locations, the Euclidean distance between predicted and actual fovea location is measured. To compare the results, the pretrained weights provided by Meyer et al. (2018) was used. Dice Coefficient is measured for optic disc and vessel segmentation. For the semi-supervised classification, accuracy, precision, Kappa score,

AUC-ROC, sensitivity, and F1-scores are measured. Refer Appendix D.1.9 for further details. The test set was grouped into two parts. First part, consists of images randomly selected from REFUGE and IDRiD datasets. The second had fundus images from the ODIR dataset. The results for the first part is given in Table 4.4. The classification met-

Table 4.4: Performance metrics of the classifier for IDRiD dataset.

<b>Parameter</b>	<b>Values</b>
Accuracy	90.63
Precision	0.911
Kappa Score	0.717
AUC-ROC	0.832
Sensitivity	0.83
F1 score	0.94

rics obtained for ODIR dataset is presented in Table 4.5. The proposed semi-supervised GAN model is compared with the transfer learning methods as illustrated in Gour and Khanna (2021); Wang et al. (2020a); Li et al. (2021) and it is compared with InceptionResnet model. Considering the classification accuracy of about 87%, the proposed GAN outperforms the other existing models. The least classification accuracy is obtained using the method in Wang et al. (2020a). The Kappa score is a metric used to measure the inter-rater reliability. A higher value of Kappa score implies that there is higher similarity between expected outcome and actual outcome. This measure is observed to be the highest for the method proposed in Gour and Khanna (2021). However, the proposed GAN model also gives a good level of agreement compared to the other methods Li et al. (2021); Wang et al. (2020a). The F1-score is considered as a weighted average of precision and recall terms. Therefore, the higher value of F1-score and AUC are desirable properties of a good classifier. The proposed GAN model gives an AUC of 0.84 and F1-score of 0.85 which marginally differs from the values obtained using other methods. A combined metric of evaluation is used to compare these models, termed as a final score. It is evaluated as the average of Kappa score, F1-score, and the AUC. The final score of the proposed semi-supervised GAN is 0.833, which is higher than the transfer learning using combined Inception-Resnet models. There is significant difference in final score between the proposed GAN model and the models in Wang et al.

(2020a); Gour and Khanna (2021). The model in Li et al. (2021) is primarily based on transfer learning. Transfer learning models are considered to be appropriate when small dataset is available. However, transfer learning models possess complex architecture and expect the distribution of the originally trained dataset and the test dataset to be the same. In such situations, the proposed GAN model performs well for both segmentation and classification tasks. Overall, from the Table 4.5, we can infer that the proposed semi-supervised GAN is able to achieve an acceptable range of accuracy in classification.

### 4.5.2 Analysis of classifier model

Further, the classifier ( $C_S$ ) performance is analyzed by breaking down the five classes into two categories at a time. The results are recorded and plotted in the form of horizontal column chart as shown in Figure 4.9. Accordingly, the groups considered for various trials are Normal and AMD (N\_A), Normal and DR (N\_D), and Normal with Glaucoma (N\_G). The three classes considered for further trials include Normal, Glaucoma, and AMD (N\_G\_A), as well as Normal, DR, and Glaucoma (N\_D\_G). From Figure 4.9, it is evident that, the highest accuracy of about 90% is obtained for two class classification of Normal and AMD. The lowest accuracy of around 72% is achieved for classification of Normal and DR. The Kappa score is highest for three class classification (N\_D\_G). Following this, the classification categories N\_A, N\_G\_A, and N\_D project higher scores compared to the N\_G classification. The F1-score and AUC are in similar range for classification of N\_A, N\_G\_A, and N\_D\_G. It is found to be least for N\_D classification. Overall, the final score is found to be around 0.89 for N\_A and N\_G\_A classification. The N\_D classification has least final score. Overall, we can infer from the results that the classifier is under-performing with Diabetic Retinopathy classes. This is probably because the majority of the images represents the early stages of Diabetic Retinopathy which might mislead the classifier. On the other hand, the classifier is able to distinguish between Normal and AMD categories well. The AMD fundus images are distinguishable compared to the Glaucoma fundus images as the F1-score is relatively higher for AMD class. Originally, the number of AMD category fundus images available in the dataset is much lesser compared to the number of DR

images. Thus, it can be noted that augmentation has positively influenced the classifier to distinguish AMD from Normal category. As Glaucoma is a condition mostly associated with optic cup and disc ratio, training the classifier using segmented optic disc region can boost the performance of this category. A small subset of images is chosen

Table 4.5: Performance of proposed fundus image classifier on ODIR dataset.

Method	Accuracy	Kappa	F1-score	AUC	Final score
Gour and Khanna (2021)	0.82	0.89	0.82	0.5	0.7367
Wang et al. (2020a)	0.66	0.43	0.85	0.67	0.65
Li et al. (2021)	0.85	0.76	0.94	0.78	0.8267
InceptionResnet	0.723	0.503	0.8793	0.86	0.7474
<b>Proposed</b>	<b>0.87</b>	<b>0.81</b>	<b>0.85</b>	<b>0.84</b>	<b>0.8333</b>

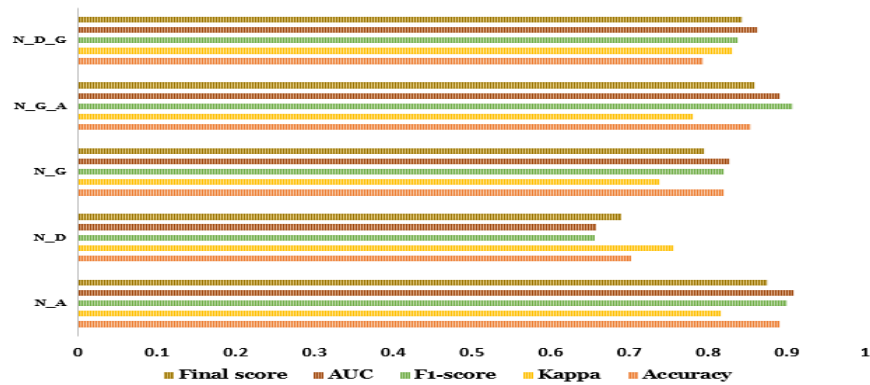


Figure 4.9: Comparison of semi-supervised GAN performance to classify multiple categories of retinal disorders.

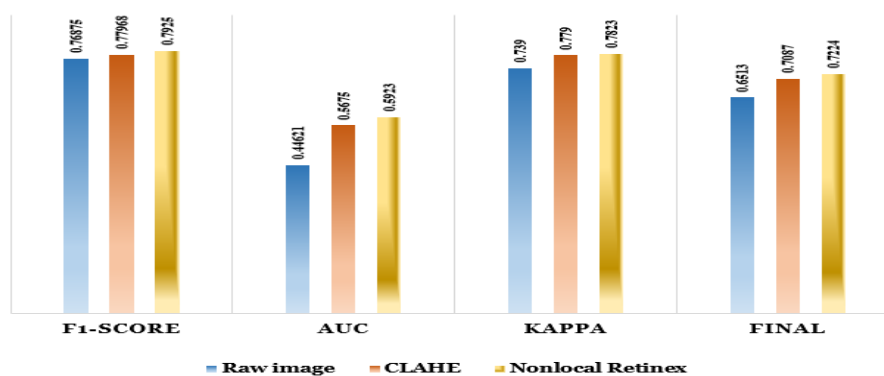


Figure 4.10: The performance metrics recorded for a subset of images to compare classifier performance on the preprocessed images and raw images from ODIR dataset.

randomly from Normal and Other disorders categories to verify whether the classifier

performance is improved with preprocessing. These images are fed to the trained semi-supervised GAN discriminator model in three different ways. The first method is to feed the raw images directly (unprocessed images), the second method is to enhance the images using CLAHE as proposed in other literature, and the third is using nonlocal retinex framework. The results are depicted in Figure 4.10. The nonlocal retinex method boosts the performance of the classifier to some extent compared to the raw images. Particularly, from Figure 4.10, it can be noted that the final score improved from 0.65 to 0.72, for a small subset of images.

### **4.5.3 OCT image classifier**

The proposed GAN model takes about 4 hours (100 epochs) to complete the training on Dataset 1 and it takes about 6 hours to complete the training on Dataset 2 (60 epochs) and attain stable accuracy. Data augmentation was included for Dataset 1 with random angle rotation. The hyperparameters of the deep learning model are empirically set by monitoring the total average accuracy. The proposed model is compared with multiple datasets and with multiple other deep learning models. Figure 4.11 depicts the sample input image, its corresponding groundtruth segmentation image, and the fake image generated by the generator of WGAN when Dataset 1 was used as input. The groundtruth segmentation includes 7 retinal layers. The predicted images favorably mark 2 retinal layers (ILM and RPE) consistently well. The other 5 retinal layers are faint. In general, a good classification result is obtained when the generator generates bad images, as explained in the paper Dai et al. (2017). Therefore, to obtain well segmented output, the generator model has to be trained with better quality images. Since the proposed model uses the research tool to extract the ground truth images for segmentation and considering the fact that the classification is the major aim of the model, quantification of segmentation is not highlighted in this work. Transfer learning models are used to compare the performance of the proposed WGAN for classification on Dataset 1. The state-of-the-art CNNs, like Resnet-50 He et al. (2016), Densenet Huang et al. (2017), and Inception-v3 Szegedy et al. (2016) are tested. The last three layers of these models are removed, and dense connections are added. The classification metrics such as accuracy, sensitivity, specificity, precision, and recall are used to compare the



various models. The performances of all models are in a comparable range. The training loss plots and the corresponding confusion matrices are given in Figure 4.12, and Figure 4.13, respectively.

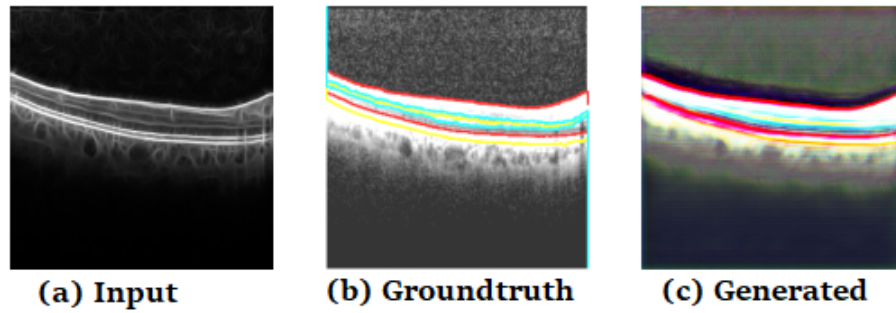


Figure 4.11: Segmentation output of Wasserstein GAN.

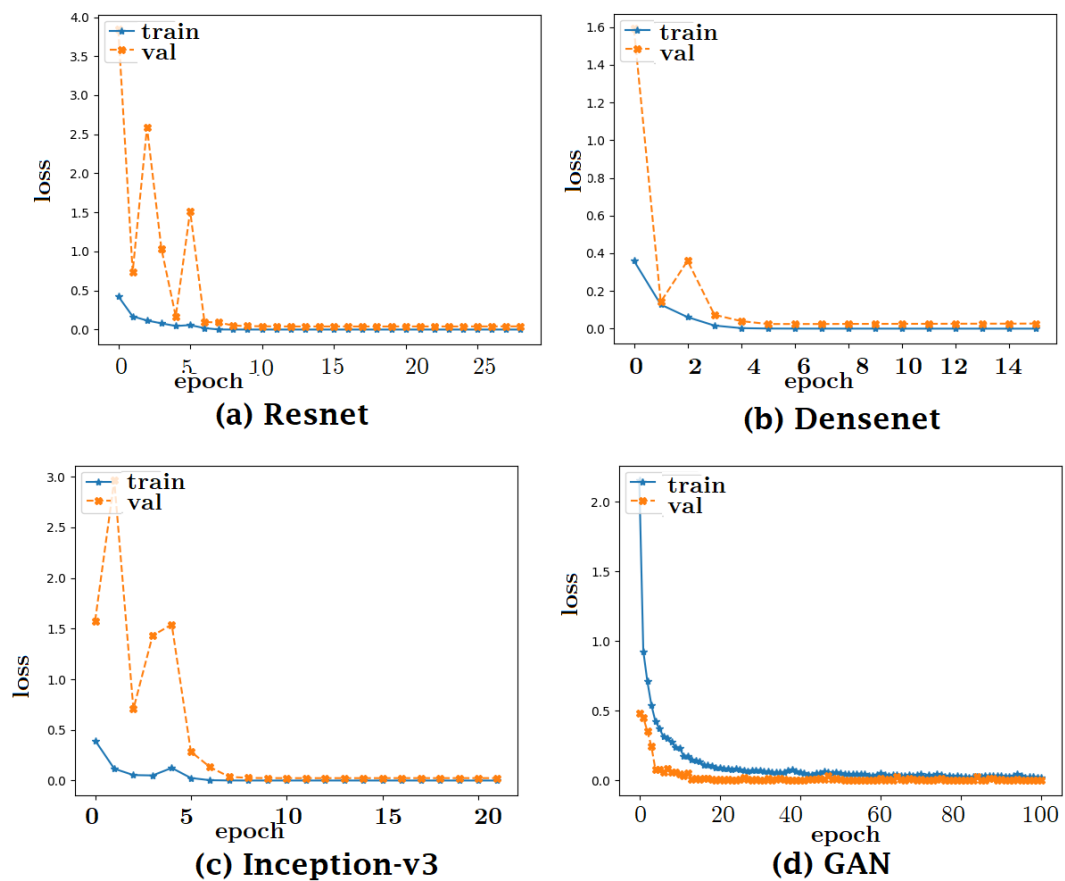


Figure 4.12: Training and validation loss plot with epochs.

**Analysis of Table 4.6:** The first three columns project the result of using Resnet-50, Densenet, and Inception trained on Dataset 1. The last column highlights the classifi-

Table 4.6: Classification metrics on Dataset 1 (Reza et al. (2018)).

Metric		Resnet-50	Densenet	Inception-v3	Proposed GAN
Classwise Sensitivity	AMD	76.65	57.58	76.26	<b>86.38</b>
	DME	40.63	51.14	42.26	<b>63.01</b>
	NORMAL	61.06	62.27	57.54	<b>73.27</b>
Classwise Specificity	AMD	75.23	83.05	65.92	<b>75.23</b>
	DME	84.69	84.86	94.08	<b>94.26</b>
	NORMAL	80.46	66.80	77.94	<b>92.64</b>
Classwise Accuracy	AMD	75.69	74.81	69.26	<b>78.84</b>
	DME	72.54	75.56	79.84	<b>85.64</b>
	NORMAL	72.67	64.98	69.77	<b>84.88</b>
Overall Sensitivity		59.43	56.99	58.75	<b>74.22</b>
Overall Specificity		80.13	78.24	79.31	<b>87.22</b>
Average Accuracy		73.63	71.78	72.95	<b>83.12</b>
Recall		0.594	0.57	0.58	<b>0.742</b>
Precision		0.591	0.59	0.628	<b>0.767</b>
F1-score		0.593	0.58	0.607	<b>0.754</b>

Table 4.7: Classification metrics on Dataset 2 (Kermany et al. (2018a)).

Metric		Method in Alqudah (2020)	Method in Sunija et al. (2021)	Proposed GAN
Classwise Sensitivity	CNV	84.71	93.01	<b>86.29</b>
	DME	60.27	22.13	<b>81.97</b>
	DRUSEN	42.57	21.57	<b>63.22</b>
	NORMAL	79.43	46.43	<b>91.16</b>
Classwise Specificity	CNV	88.34	53.02	<b>92.39</b>
	DME	92.86	97.1	<b>95.29</b>
	DRUSEN	92.58	93.86	<b>95.18</b>
	NORMAL	92.17	93.22	<b>96.29</b>
Classwise Accuracy	CNV	86.72	70.85	<b>89.68</b>
	DME	88.41	86.85	<b>93.47</b>
	DRUSEN	47.43	86.41	<b>91.89</b>
	NORMAL	88.16	78.47	<b>94.67</b>
Overall Sensitivity		66.75	45.79	<b>80.66</b>
Overall Specificity		91.49	84.3	<b>94.79</b>
Average Accuracy		77.68	80.645	<b>92.42</b>
Recall		0.667	0.457	<b>0.806</b>
Precision		0.661	0.552	<b>0.788</b>
F1-score		0.664	0.5	<b>0.797</b>

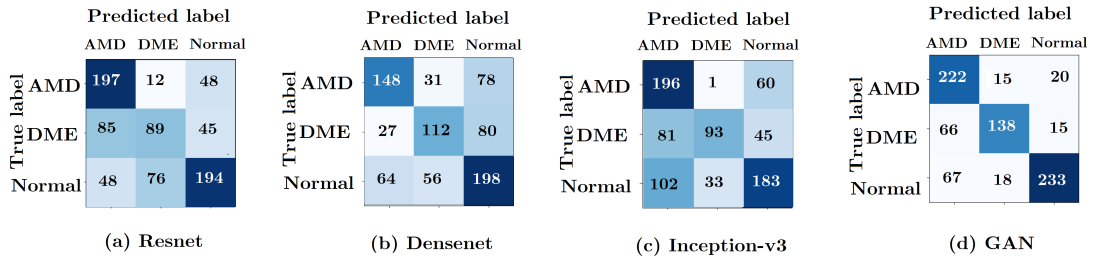


Figure 4.13: Confusion matrices of Wasserstein GAN.

cation metrics of the proposed GAN. From this table, we can observe that the overall average accuracy of the proposed GAN is higher than the transfer learning models. Among the transfer learning models, Resnet-50 performs better than the Inception-v3, which in turn is better than the Densenet models in terms of accuracy. A similar pattern is observed for overall sensitivity, specificity, and recall. Observing the per-class accuracy, it is found that the Densenet and Inception models fail to classify Normal category images precisely. A deeper insight into the confusion matrix given in Figure 4.13 conveys that the Normal category images are misclassified into AMD category. However, the proposed GAN model can distinguish these classes in a better way. The Densenet architecture is more complex compared to Resnet-50, and it consumes more memory. The dense blocks fail to identify the subtle changes between the three classes (AMD, DME, and NORMAL). The precision and recall values are in a comparable range among the transfer learning models. As F1-score depends on the precision and recall scores, it is evident that the least F1-score is exhibited by Densenet model and the highest is obtained by the proposed GAN model.

Overall, the proposed GAN-based model performs better than the other transfer learning methods. Transfer learning could be an optimal choice when the number of training samples (input image, label pairs) are available in small numbers. Furthermore, the pre-trained model must be trained using a similar training dataset when compared to the dataset used while applying the pre-trained model (on a new problem). However, they do not perform the segmentation of retinal layers. Moreover, the transfer learning models are complex in architecture and have more trainable parameters. Designing a model that can efficiently perform both segmentation and classification will facilitate over-the-air programming. Such a model embraces the property to improve the perfor-

mance by learning dynamically from new samples.

**Analysis of Table 4.7:** As a further stratum of evaluation, the performance of the proposed GAN model is compared with methods in Alqudah (2020) and Sunija et al. (2021) using Dataset 2. These results are tabulated in Table 4.7. Accordingly, it can be seen that the proposed model performed well overall compared to others with an average accuracy of 92.42% and a F1-score of 0.797. It is observed that the sensitivity for DRUSEN class is relatively low for all the models. This is because the number of training images available in this class is much smaller than the other classes. Hence the diversity of the images is not sufficiently predicted by the deep learning models. Moreover, drusen represent the early stages of AMD. This stage has marginal differences between the normal category and therefore, these subtle variations within the retinal layers are not captured by the model. The architecture of Alqudah (2020) and Sunija et al. (2021) are similar in terms of a number of layers and the activation function. However, the difference in performance is due to the kernel size, optimizer and the learning rate used in their methods. Compared to Dataset 1, the number of images in Dataset 2 is large. Additionally, the images provided in Dataset 2 are less affected by speckle noise compared to Dataset 1. Therefore, the proposed GAN performs better on Dataset 2 when compared to Dataset 1.

#### 4.5.4 Design of Wasserstein GAN model

This section elaborates on the empirical analysis conducted to design the proposed WGAN model. Though several trials were conducted, some of the relevant results are illustrated here. Since the discriminator is mainly responsible for the classification of the OCT images, the discussion is limited to the architecture of the discriminator. For simplicity, the different architectures experimented with are represented as A, B, and C. This notation denotes how the number of filters are varied in each layer of the discriminator and is given as:

- A: 16-32-64-128.
- B: 32-64-128-256.
- C: 128-256-512-1024.

The average accuracy obtained using different architecture layouts of the discriminator when tested on Dataset 2 is shown in Figure 4.14. However, a similar pattern was observed when Dataset 1 was utilized for testing. In Figure 4.14, the X-axis represents various model architectures, and the Y-axis denotes the average accuracy considering 4 classes. The kernel size (KS) and the learning rate (LR) are varied for each model. This is depicted as a grouped vertical bar plot in Figure 4.14. Accordingly, the highest accuracy is achieved with a learning rate of 0.005, kernel size of 5, employed on the discriminator architecture layout following the 32-64-128-256 number of filters in each layer (B). The number of epoch is selected based on early stopping for each trial. On average, it is seen that modifying the number of filters, kernel size, and the learning rate does not significantly influence the performance of the discriminator. This is implied from Figure 4.14, where the average accuracy is found to vary between 85% and 93%. Therefore, considering the number of trainable parameters, the discriminator and generator architectures are designed as shown in Figure 4.7. To highlight the significance of incorporating handcrafted Gabor features, the proposed WGAN model was tested on raw images and enhanced (using Gabor features) images. The accuracy and F1-scores corresponding to this trial is depicted as a bar graph in Figure 4.15. Accordingly, training on raw images exhibited the lowest accuracy and F1-scores. Incorporating Gabor feature extraction further boosts the performance of the classifier. A similar trend was observed for both datasets.

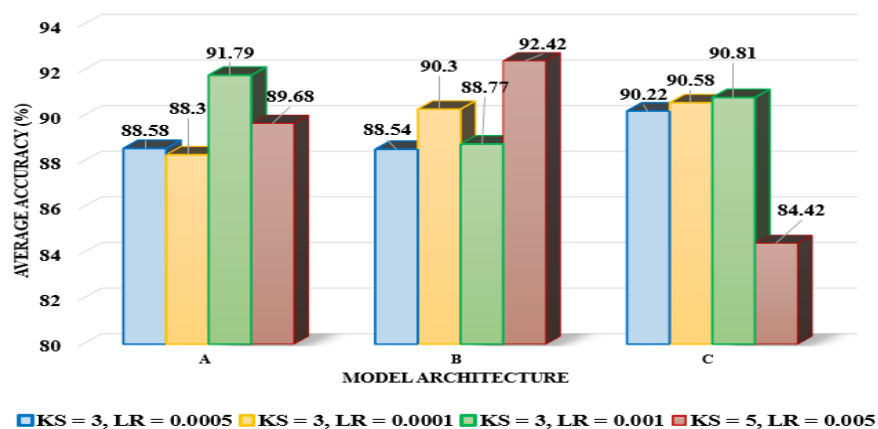


Figure 4.14: Comparison of Wasserstein GAN with different parameters, \*KS = Kernel Size, LR = Learning Rate.

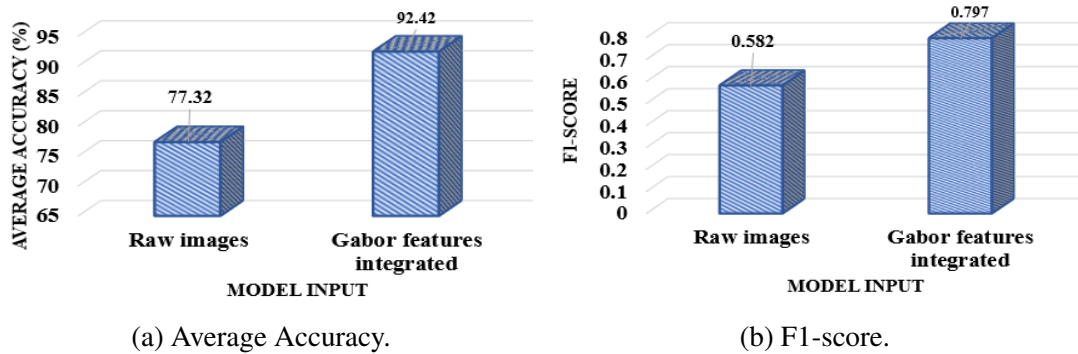


Figure 4.15: Effect of including Gabor feature.

## 4.6 SUMMARY

Automatic retinal disorder detection has been a prime focus of research with the advent of deep learning in the recent years. Majority of the works are oriented towards single task such as segmentation or classification. Furthermore, the deep learning models are trained and tested on a single dataset. Therefore, the proposed research work aims to train the GAN using images from multiple repository. Some of the existing works perform segmentation on retinal images such as the optic disc region and then perform classification such as Normal and Glaucoma condition. In such cases, the segmentation performance severely affects the classification results. In the proposed work, GANs generate the segmentation output that can be used by ophthalmologists to perform other diagnosis such as estimation of choroidal thickness and optic cup to disc ratio.

This Chapter discusses various datasets available in public domain that are accessed in the proposed work. As discussed in Chapter 1, a fundus provides information about optic cup and disc ratio, blood vessels, and the fovea region. Since limited groundtruth images are available in public domain to perform segmentation of these regions, a semi-supervised GAN is proposed in this Chapter. It uses partially labeled data in addition to the unlabeled dataset. The hyperparameters are chosen based on the empirical analysis. The performance of the proposed GAN is assessed using standard classification metrics such as accuracy, F-1 score and Kappa score. The results of the fundus image analysis is published as research articles <sup>1,2</sup>.

<sup>1</sup>Smitha, A., & Jidesh, P. (2021). "A Semi-supervised Generative Adversarial Network for Retinal Analysis from Fundus Images". *Computer Vision and Image Processing*, 351–362. doi:10.1007/978-981-16-1086-8\_31.

<sup>2</sup>Smitha, A., Jidesh, P. *Classification of Multiple Retinal Disorders from Enhanced Fundus Im-*

Similar to fundus images, the OCT images are also acquired from public domain. Apart from the public repositories, OCT images were acquired from a private Indian hospital - Shankara Nethralaya, Chennai, India. As the number of images are less than 100, to validate the effect of preprocessing, choroidal thickness was estimated and compared with annotated values<sup>3</sup>. The performance of the classifier is compared with transfer learning approaches and certain other relevant methods. Furthermore, the design of the WGAN architecture and the effect of preprocessing are discussed in detail in this Chapter. It is evident from the results that the GAN models can achieve reliable results to classify retinal disorders from fundus and OCT images, favourably addressing the research question about the GAN models. The results presented about the analysis of OCT images using WGAN is submitted to a journal<sup>4</sup>. Overall upto 90% of an average classification accuracy is obtained on heterogeneous datasets using semi-supervised GAN and WGAN for fundus and OCT images, respectively.

Deep learning methods have revolutionized the approach of solving complex contemporary problems. Nevertheless, it is like a black box method of solving certain challenging classification problems. In other words, it is unclear what leads the deep learning model to arrive at a particular output. Thus, the explainability of the model is still in juvenile stage. Certain methods like Grad-CAM are efficient to some extent in visualizing the features identified by the deep learning model. This is easily applicable to a transfer learning model. GANs on the other hand have an adversarial nature and therefore, methods such as Grad-CAM are not easily adaptable. Some of the metrics such as Inception score and FID often grade the GAN in comparison to the transfer learning models. Computing these and exploring the explainability of the proposed model will thus be taken up as a future work.

---

*ages Using Semi-supervised GAN*. SN COMPUT. SCI. 3, 59 (2022).<https://doi.org/10.1007/s42979-021-00945-6>

<sup>3</sup>Smitha, A., Jidesh, P., Jothi Balaji, J., & Lakshminarayanan, V. *Development and validation of a novel automated method for quantification of choroidal thickness in age-related macular degeneration*. In Proc. of SPIE Vol (Vol. 11842, pp. 118422E-1).

<sup>4</sup>Smitha, A., Jidesh, P. *Detection of retinal disorders from OCT images using generative adversarial networks*. Multimedia Tools Applications (2022).





## Chapter 5

# CONCLUSION AND FUTURE WORKS

Ophthalmologists often rely on multiple imaging modalities to analyze the abnormalities. Motivated by this, an automatic multi-modal retinal disorder detection has been a novel domain of interest in the recent years. So, a conceptual study in this direction is presented here.

### 5.1 MULTI-MODAL ANALYSIS

Multi-modal analysis is the technique of using more than one imaging modality to classify a particular retinal disorder. As discussed in Chapter 1, various modalities capture details from different perspective. Therefore, combining the modalities can strengthen the identification of distinguishing features for various retinal disorders. Some of the recent works in this direction are summarized in Table 5.1. From this table, it can be inferred that the transfer learning approaches give promising results in terms of classification accuracy for multi-modal analysis. The existing works are oriented towards classification of AMD and normal categories using the images acquired from private hospitals. A multi-modality repository is released in the public domain in the month of July 2021, to detect the various stages of Glaucoma (GAMMA-2021 available at <https://aistudio.baidu.com/aistudio/competition/detail/90>). This repository has 100 CFP images. For each CFP, 255 corresponding OCT images are provided. The images are categorized as normal, early stage, and advanced stage Glaucoma. As project "Macula" offers a small number of images in the public domain for AMD and normal categories, a novel transfer learning based approach is proposed for multi-modal analysis that combines Macula database with GAMMA-2021 dataset. The number of classes

can be extended to three - AMD, Glaucoma, and Normal categories.

Table 5.1: Summary of existing works on multi-modal analysis.

Sl. No.	Paper details	Datasets	Task	Description	Results	Advantages	Disadvantages
1	Vaghefi et al. (2020)	Private hospital	Classify AMD and healthy	Inception with Resnet v2 architecture is applied for OCT, OCT-A, and CFP.	Accuracy - 0.92 to 0.99	Transfer learning is applied. Most recent imaging modalities are considered. Output performance is measured using attention maps.	Very small dataset of 75 patients was used. Clinical findings like drusen are not segmented.
2	Wang et al. (2019b)	Private hospital	Classify dry AMD, wet AMD and normal	Loose pair training method (label based) is adopted with pretrained Resnet 18 architecture	Accuracy: 0.97	Multimode class activation map is generated for visual interpretation	Performing segmentation can enhance the reliability of the model.
3	Yoo et al. (2018)	Project Macula database	Classify AMD and normal	VGG-19 transfer learning on single and multiple modality	Accuracy: 0.95	First experimental analysis on multiple mode analysis.	Small dataset without performing segmentation of drusen.

Unavailability of images in large number is a huge setback to train the GAN in this direction. The choice of transfer learning model over the GAN model allows the visualization of features recognized by the deep learning model using heat maps. The block diagram for the conceptualization of multi-modal analysis is shown in Figure 5.1. Accordingly, there are three blocks: feature extraction, feature fusion, and classification. The first step is to enhance the quality of the input images using the preprocessing

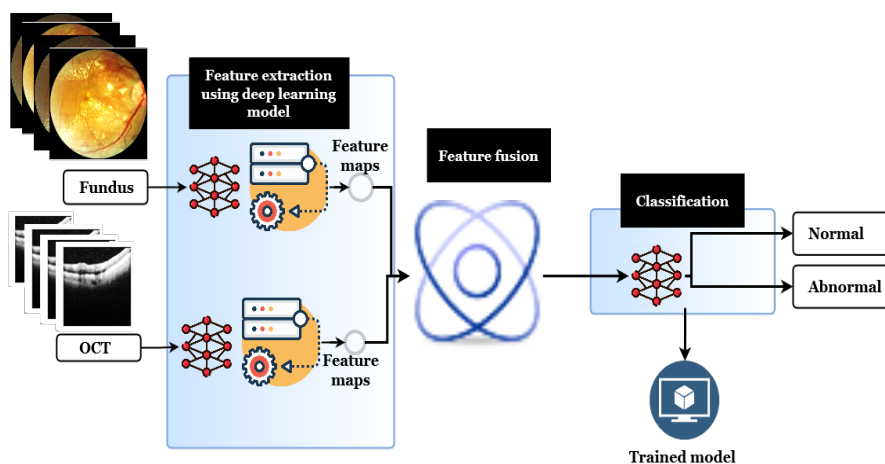


Figure 5.1: Conceptual diagram for multi-modal analysis.

algorithms illustrated in Chapter 3 of the thesis. The enhanced images are used for further processing. The state-of-the-art deep learning models are employed to extract the feature maps. The attention mechanism based deep learning architectures can independently learn the features to map the input images to their corresponding output labels. The output of the first stage will be the feature maps. This is obtained by removing the final activation layer of the deep learning architecture. The feature fusion stage involves merging the feature maps derived from the two imaging modalities. The fusion is achieved using concatenation operation. Finally the concatenated feature maps are fed to a pre-trained classifier. The classification module is similar to the feature extraction module in terms of deep learning architecture. The output labels are normal and abnormal. Further the abnormal category includes AMD and Glaucoma. Once the model is trained, it can be tested on unseen images. The performance of the model can be assessed using standard classification metrics as discussed in Chapter 4 of the thesis.

## **5.2 CONCLUSION**

The retinal disorders are automatically detected using deep learning models from two different modalities - CFP and OCT as illustrated in Chapter 4. Specifically, the GAN models are explored to classify the input images while segmenting the prominent features considered by the ophthalmologists. The semi-supervised and WGAN models demonstrated in Chapter 4 has lesser depth (in terms of number of layers), simple architecture compared to the complex transfer learning models, thereby leading to lesser number of trainable parameters.

Fundus images are colored (RGB) while OCT images are acquired in grayscale. The image acquisition principle differs in these two modalities and hence it is necessary to incorporate appropriate processing modules. The objective of the research work was to propose an adaptive image restoration model that can be tuned to handle multiplicative speckle in OCT and illumination inhomogeneity in CFP images. Further, the performance of GAN towards segmentation of prominent features, and classification of disorders such as AMD, DME, and Glaucoma, from OCT and CFP images, had to be compared with other relevant existing models. The proposed model can be fine-tuned to

test on larger set of images acquired from the Indian hospitals. It contributes to expedite the diagnosis procedure of retinal disorders. Prior to the research work, we were aware from the existing literature survey that:

- The fundus images suffer contrast degradation. The illumination inhomogeneity masks certain prominent structures in the fundus.
- OCT images are severely degraded with speckles, which are inherent according to the principle of formation of OCT images.
- The transfer learning models perform well in terms of classification accuracy provided the tested images are similar to the training dataset.

The proposed research work augments to the existing literature by favourably supporting the following facts.

- The statistical analysis ascertains that the multiplicative speckle can be formulated to follow a Gamma distribution.
- The Bayesian maximum a priori in addition to deep image prior concept aids to obtain the despeckling framework that can be tuned to alleviate the effect of speckles in OCT images.
- The variational models assist to enhance the features in retinal images without leading to an over-smoothing paradigm.
- The semi-supervised GAN achieves a stable performance in classifying the retinal images into normal or abnormal even when tested using heterogeneous datasets.
- The multi-modal retinal image analysis, being the limelight of research, is a desirable approach to strengthen the performance of the proposed deep learning models.

The quantitative and qualitative results of the proposed restoration and classification models are elaborated in Chapters 3 and 4 respectively. The image quality metrics such as CNR, GCF, entropy, and ENL values are in favour of the proposed preprocessing model. Further the enhanced images are used to train the GAN model.

The computational complexity of the deep learning models is assessed in terms of execution time for each module, which is highlighted in Chapter 4. When fundus image is considered, the non-local retinex model uses the fast numerical approach called Split-Bregman and produces the enhanced image in order of seconds depending on the image size. The time taken by non-local deep image prior model greatly depends on early stopping criteria and the image resolution. As the non-local computation is not executed in parallel mode, compared to the proposed NLTGVR method, non-local deep image prior model takes relatively more time to restore the image. It must be noted that the Wasserstein and semi-supervised GAN models are trained on machine specifications as described in Chapter 4, while the non-local deep image prior model is trained on a different machine, as described in Chapter 3. The proposed preprocessing algorithm (Chapter 3) enhances the retinal layers, prominent features such as optic disc and macula region in fundus and OCT images. It can be integrated as an in-built preprocessing algorithm in image acquiring devices. The processed images can be used to measure choroidal thickness and choroidal vascularity index. As the era of digital world is emerging, smartphone enabled applications are available in market to grade the diabetic retinopathy. The application of the proposed work can also be diverted in a similar direction, where the disorders such as AMD, DME, and Glaucoma are classified. The proposed work can be tested on latest fundus images acquired from Indian hospitals to automatically classify the images as normal or abnormal.

Overall, retinal disorders are automatically detected from fundus and OCT images using GANs and the performance is analysed qualitatively and quantitatively. Serving as an end-to-end solution for ophthalmologists, the proposed work acts as a baseline to extend the automatic grading of disorders such as glaucoma and AMD, that are primary concerns of elderly population. Such a model embraces the property of over-the-air programming and can be useful in many such medical applications.

### **5.3 FUTURE WORKS**

The proposed work has certain limitations that paves way for future works. The potential extension of the proposed work is presented in the summary of the Chapters 3 and

4. Primarily, the unavailability of large number of retinal images of various modality is a potential cause for the marginal improvement in performance of the GAN model. Therefore, it is essential to acquire images from Indian hospitals and validate the performance of the proposed GAN model. As explained in Chapter 3, the proposed non-local deep image prior model is a novel way of despeckling the OCT images. This deep learning based approach can be further extended to incorporate perceptually inspired retinex model such that the restoration framework performs enhancement of OCT images in addition to despeckling. The generative model could be used to produce the retinex and illuminance components of an image. Further by modifying the loss function to integrate the contrast enhancement of the image, non-local deep image prior model can be extended in the variational domain (by incorporating  $L_2$  norm and total variational norm terminologies). The proposed semi-supervised GAN provides the segmented fundus images. The segmented images can be used to extract clinical information such as optic cup to disc ratio, cyst dimension, and discontinuities in blood vessels. Particularly, active contour based methods can be utilized to extract the optic cup and disc region from the segmented images and the cup to disc ratio can be combined with the image to classify Glaucoma or normal conditions. However, availability of ground truth data for this quantification is a challenge. The multi-modal retina image analysis is yet another scope of future work as discussed earlier in this Chapter.

## APPENDIX I

This section highlights the necessary preliminaries, definitions and examples related to the non-local variational restoration models.

### A.1 INVERSE PROBLEMS AND ILL-POSED PROBLEMS

Mathematical minimization problems could be categorized as (Refer Yagola (2010)):

- Well posed problems.
- Ill-posed regularizable problems.
- Ill-posed nonregularizable problems.

According to Hadamard (1953), a problem is well-posed if the following conditions are satisfied:

- Existence: it has atleast one solution.
- Uniqueness: it has atmost one solution.
- Continuity: the solution depends continuously on data and parameters. In other words, the unique solution is stable in all conditions.

Any problem that fails to satisfy any one of the above mentioned conditions is termed as an ill-posed problem. An example for ill-posed problem is:

$$x = 1 + y, \tag{A.1}$$

$$x = 3 + y.$$

The above mentioned system of linear equations has no solution. However, if eqn. (A.1) can be modified as:

$$x = 1 + y, \tag{A.2}$$

$$x = 3 + y(1 + \varepsilon).$$

where  $\varepsilon \neq 0$ , then an unique approximate solution can be sought. Thus, this category of ill-posed problem is referred as ill-posed and regularizable. Another illustration of this

category is an inverse problem. Due to the ill-posed nature of these problems, the inverse may or may not exist. Even if the solution exist, small perturbations in the data can lead to large impact in the solution (unstable solution). Penalization techniques seeks solution to these inverse problems. Consider the system of linear equations represented as  $Ax = B$ . The solution is sought as  $x = A^{-1}B$ . To obtain the solution from  $B$ ,  $A^{-1}$  must exist. Therefore, it is an inverse problem. Using penalization theory, the ill-posed problem is stated as a constraint minimization problem with a penalty function ( $\mathcal{P}$ ) such that, addition of  $\mathcal{P}$  compels to satisfy the condition of well-posed problem. The penalization theory can be either regularization approach or a sparsity model. Using the regularization approach, the solution is given as:

$$\min_x \{ \|Ax - B\|_2 + \lambda \| \mathcal{P}x \|_{L_p} \}. \quad (\text{A.3})$$

where  $\lambda$  is a nonzero regularization parameter,  $\|\cdot\|_2$  is standard  $L_2$  norm, and  $\|\cdot\|_{L_p}$  is  $L_p$  norm. If the data is sparsely represented, then  $L_p$  norm can be either  $L_0$  or  $L_1$ . When the data of interest is smooth, regularization approach is used where  $L_p$  norm is valid for all  $p > 1$ . Tikhonov Regularization (Tikhonov and Arsenin (1977)) is commonly used iterative technique of regularization that incorporates  $L_2$  norm term. Choosing  $\| \mathcal{P}x \|_{L_2}$  ensures smooth regularized approximations. The above discussed approach can be generalized to non-linear systems as well. Though the penalization theory can find approximate solutions of the ill-posed problems, in degradation process (denoising an image), the penalization theory tends to ignore the statistical properties of the distortion (Mohammad-Djafari (2002)). This can be incorporated using Bayesian approach. Here, the ill-posed problem is transformed to an energy minimization problem using Bayesian Maximum apriori estimation and then the problem is solved using variational approach. However, it penalizes the image edges. Refer Aubert and Aujol (2008) for details.

## A.2 REVIEW OF VARIATIONAL FRAMEWORKS

Consider the image degradation problem represented as:  $U_0 = UN$ , where  $U_0$  is observed image,  $U$  is clean image, and  $N$  is the noise that causes degradation. The degradation is inevitable in observed image due to the acquisition procedure or artifacts in



devices. The objective of these problems is to estimate  $U$  from  $U_0$  using some prior knowledge about the degradation  $N$  and image features. It is therefore an instance of ill-conditioned inverse problem. The distortions in the image can be suppressed by minimizing the total variation. Restoration of the clean image can therefore, be formulated as an energy minimization problem. Formulating a functional ( $J(U)$ ) that defines the image quality, a constraint minimization can be given as:

$$\min_U J(U); \quad \text{such that } \left\| \frac{U_0}{U} - 1 \right\|_2^2 = \sigma_N^2. \quad (\text{A.4})$$

where  $\sigma_N^2$  represents the noise variance, and a smaller value of  $J(U)$  indicates better image quality. Reformulating the constraint minimization problem to an unconstrained problem as:

$$\min_U \left\{ \int_{\Omega} J(U) d\Omega + \frac{\lambda}{2} \int_{\Omega} \left( \frac{U_0}{U} - 1 \right)^2 d\Omega \right\}. \quad (\text{A.5})$$

where  $\lambda$  is a regularization parameter. The constraint imposed ensures that the details of the image is well preserved. In general, the energy minimization problem can be expressed in the form:

$$E_T = E_S + \lambda E_F, \quad (\text{A.6})$$

where,  $E_T$  denotes the total energy,  $E_S$  corresponds to the smoothness component and  $E_F$  represents the energy of fidelity term. The smoothness term involves derivatives of  $L_p$  norm, and it is called the regularization term. The level of smoothness is controlled by fidelity term. In other words, addition of fidelity ensures minimum deviation from the actual solution. The data fidelity term preserves the details and the balance between these two factors is achieved using the regularization parameter  $\lambda$ . Considering the fact that, in order to improve the image quality, the total variation has to be minimized, the term  $\int_{\Omega} J(U) d\Omega$  can now be replaced with a total variation term. This variational framework is proposed by Rudin et al. (1992). Assuming that the image belongs to bounded space ( $\Omega \subseteq \mathbb{R}^2$ ), the total variation is defined as:

$$TV(U_0) = \int_{\Omega} \|\nabla U_0\| d\Omega. \quad (\text{A.7})$$

where  $\nabla U_0 = \left[ \frac{\partial U_0}{\partial x} \quad \frac{\partial U_0}{\partial y} \right]$ . Therefore, eqn. (A.5) can be written as:

$$\min_U \left\{ \int_{\Omega} \|\nabla U_0\| d\Omega + \frac{\lambda}{2} \int_{\Omega} \left( \frac{U_0}{U} - 1 \right)^2 d\Omega \right\}. \quad (\text{A.8})$$

The total variation based approach permits discontinuities in the solution. If the above equation satisfies,  $TV(U_0) < \infty$ , then it is termed as total bounded variation. Refer

Garipey (2001) for more details. Total bounded variation restoration model preserves the edge and fine details on the image. The solution to eqn. (A.7) can be obtained using Euler-Lagrange method. The downside of using variational framework is the piecewise effect or staircase effect in the solution with slow convergence. Moreover, it is local in nature. To compensate this, non-local variational models are introduced by Aubert and Aujol (2008). According to this, the variation minimization given by eqn. (A.8) can be expressed as:

$$\min_U \left\{ \int_{\Omega \times \Omega} \|\nabla_{NL} U_0\| d\Omega + \frac{\lambda}{2} \int_{\Omega} \left( \frac{U_0}{U} - 1 \right)^2 d\Omega \right\}. \quad (\text{A.9})$$

Refer Vese and Guyader (2015) for fundamentals of non-local total variational restoration models. Further, the non-local restoration models are formulated to address the noise distributions that follow Gamma and Poisson distributions using Bayesian maximum apriori technique (Jidesh and Banothu (2018); Balaji and Jidesh (2017); Holla Kayyar and Jidesh (2018)). It is also investigated for enhancement purposes using perceptually inspired retinex models (Febin et al. (2020); Jidesh and Febin (2020)).

### A.3 EULER-LAGRANGE METHOD

Consider minimization of the Lagrange functional that aims to find a function that minimizes the energy given as:

$$E(u) = \int_a^b g(x, u, u') dx. \quad (\text{A.10})$$

The solution can then be obtained using Euler-Lagrange equation given as:

$$\partial g / \partial u - \frac{d}{dx} \left( \frac{\partial g}{\partial u'} \right) = 0. \quad (\text{A.11})$$

### A.4 SPLIT-BREGMAN APPROACH

Split-Bregman is the fastest numerical iterative approach to solve minimization problems (convex), in particular, regularization problems. It is based on the Bregman distance metric. It is a symmetric measure. Consider the minimization problem of the form:

$$\min_U \{ \|S(U)\| + H(U) \}, \quad (\text{A.12})$$

that follows eqn. (A.6), Here S(U) and H(U) are convex in nature. Applying Split-Bregman scheme, the problem given in eqn. (A.12) is rewritten as a constrained opti-

mization problem:

$$\min_{U:c_1=S(U)} \{\|c_1\| + H(U)\}, \quad \text{such that } c_1 - S(U) = 0. \quad (\text{A.13})$$

Further, it is converted to an unconstrained problem as:

$$\min_{U:c_1=S(U)} \{\|c_1\| + H(U) + \frac{\lambda}{2} \|c_1 - S(U)\|_2^2\}. \quad (\text{A.14})$$

where  $\lambda$  is a constant. As the domain is now split into sub problems, it can be solved using Bregman algorithm. Refer Goldstein et al. (2009); Liu and Huang (2010); Goldstein and Osher (2009) for further details.

## APPENDIX II

This section recalls the concept of deep image prior highlighting the advantages and how it differs from traditional deep learning mechanism.

### B.1 Deep Image Prior

The concept of deep learning based blind image denoising was proposed by Ulyanov et al. (2020). Accordingly, a randomly initialized convolutional neural network such as U-net can be used as a prior to solve inverse problems. Consider denoising (an example of an inverse problem), where the image is degraded due to additive white gaussian noise, represented as:  $U_0 = U + N$ . As mentioned in Appendix A.1, the inverse problems can be formulated as a energy minimization problem expressed in the form:

$$\min_U \{R(U) + E(U; U_0)\}, \quad (\text{B.1})$$

where  $R(U)$  is a regularization term and  $E(U; U_0)$  is task-dependent term (for example, denoising) responsible for the data fidelity. Here, the regularizer can be implicit or explicit. Deep image prior used the deep learning model itself as an implicit regularizer or a prior term. Therefore, setting the regularizer as zero ( $R(U) = 0$ ), eqn. (B.1) accounts to minimizing:

$$\min_U \|U - U_0\|^2, \quad (\text{B.2})$$

where  $E(U; U_0) = \|U - U_0\|^2$  is the data fidelity term. A comparison of traditional denoising approach to the deep image prior method is given in Figure B.1. Accordingly, in traditional U-net models designed for denoising, for each degraded input image, a corresponding groundtruth image is fed to the model. Initially, the training begins

with an initial estimate ( $U^0$ ). The error is computed between the reconstructed image and the expected image. For every iteration, the gradient of minimization function is computed with respect to the input ( $U$ ) and then a new estimate is obtained. Then the model parameters are updated accordingly after seeing multiple training images. This process continues until some point of convergence is attained. The updated outputs in each iteration is denoted as  $U^1, U^2, U^3$  and so on, as shown in Figure B.1(a). Deep image prior on the other hand, uses a single input image. It searches the space of model parameters that maps to the expected output. Further, instead of computing the gradient of the function with respect to the input, it is computed with reference to the model parameters as depicted in Figure B.1(b). Here, the minimization problem is

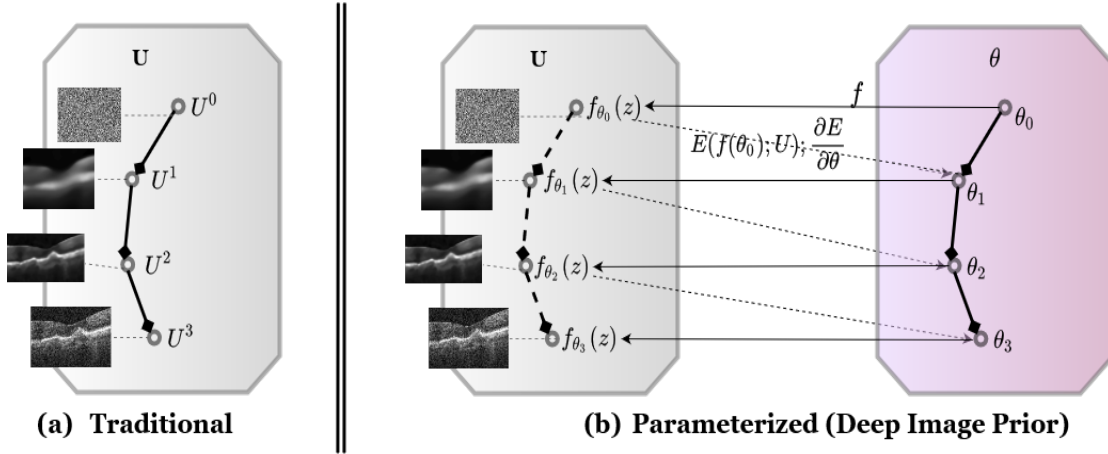


Figure B.1: Comparison of traditional deep learning method with parameterized method of denoising.

solved using optimizers, where the aim is to estimate the model parameters (weights and biases) that maps input to its corresponding image domain  $f_{\theta}(z)$ . In other words, minimizing eqn. (B.2) is equivalent to minimizing:

$$\min_{\theta} \|f_{\theta}(z) - U_0\|^2. \quad (\text{B.3})$$

For any noisy image,  $U_0$ , the U-net is optimized to generate the prior of noisy image, which is the denoised image or the reconstructed image ( $\hat{U} = f_{\theta}(z)$ ) from a random noise in latent space  $z$ . Obtaining minimum value of the optimization problem results in the situation where the reconstructed image is similar to the noisy image. However, to achieve denoising, it is assumed a cleaner denoised image is obtained by earlystopping the model training. This is depicted in Figure B.2, recreated as given by Ulyanov et al.

(2020). Here the denoising problem is depicted as transforming from one domain to another. Initially, the deep image prior model begins from noise in latent space. As the training progresses (instantiated from  $t_0 - t_4$  in Figure B.2), the model parameters are updated based on the loss computed ( $E(f_\theta(z), U_0)$ ). When the global minima is attained, the restored image is in the domain of degraded image ( $E(U; U_0) = 0$ ), that is within a gray shaded area depicted in Figure B.2. However, at time  $t_2$  depicted in this Figure, the reconstructed image is much closer to the cleaner groundtruth image. Therefore, it becomes the optimal stopping point to pause the training of deep image prior.

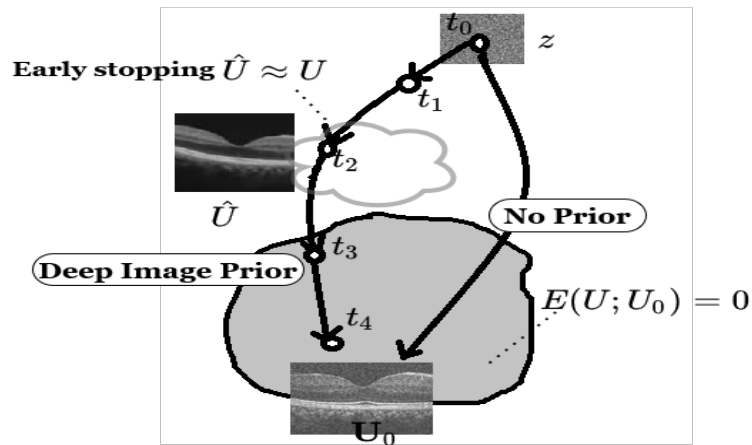


Figure B.2: Denoising strategy in deep image prior.

It is assumed that since the model is used as a prior, it always follows the path of convergence such that, there always crosses the point of closest approximation to the clean image. The splendid success of deep image prior led to recent advancements such as incorporating explicit regularizer as explained by Fan et al. (2020). Perceptually inspired retinex models are also integrated to the deep image prior model Zhao et al. (2021). It is extended to enhance images using vector bundled variants by Batard et al. (2021). The deep image prior model does not need massive number of training images or the groundtruth images. These advantages meet the requirements of restoring OCT images. Therefore, this method is adapted and further tweaked appropriately to address speckle as discussed in the thesis.

## APPENDIX III

### C.1 LAYERS OF DEEP LEARNING ARCHITECTURE

The GAN architecture proposed in Chapter 4 of the thesis has U-net model and a convolutional neural network layout. The details of each of these layers are given below.

- **Convolution layer:** Convolution is an elementary operation involved in obtaining the features of a convolutional neural network. It is a linear operation, where a filter or a kernel slides through the input image and a weighted sum (dot product) is computed. The sliding operation happens from left to right direction and from top to bottom. The output of this operation is termed as a feature map. This operation decreases the size of the input image. The size of the filter is much smaller than the input image. Generally  $3 \times 3$ ,  $5 \times 5$ , or  $7 \times 7$  are used to extract the features (Simard et al. (2003)). Some examples of the kernels include edge detection filters and sharpening filters. The convolution operation is depicted in Figure C.1. The feature map of the convolution operation is fed to an activation

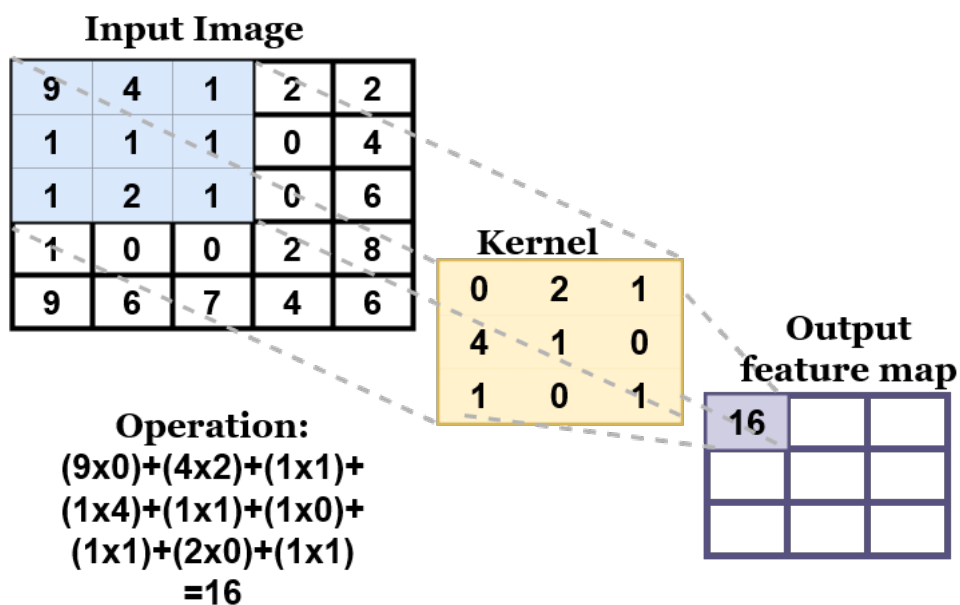


Figure C.1: Convolution operation.

function. The role of activation function is to detect whether the feature is present in a given location of the image. Padding is usually incorporated to ensure the output feature map retains the size of the input image. Stride operation refers to

the amount of pixels that can be skipped before sliding to neighboring pixel. This reduces the computation and also downsamples the output feature map. When the input image is of 3 channels (RGB), the kernel is also of 3 channel, to ensure the output will be 3-channel. The filter size, stride, and the number of filters are hyper parameters in convolutional layer.

- **Maxpooling:** As the name suggests, the maxpooling operation selects the maximum element. It is used to reduce the dimension of the feature map. It further reduces the computations and number of learning parameters of the model. The outcome is similar to the stride operation. The output size after applying maxpooling layer is calculated as:  $\frac{h_f - k + 1}{s} \times \frac{w_f - k + 1}{s} \times n_c$ , where  $h_f, w_f$  represents the height and width of the feature map,  $k, s$  denotes the kernel size and stride length, respectively, and  $n_c$  corresponds to number of channels. By selecting the maximum values from the region of feature map, prominent information is passed on to the next layer thus preserving the necessary image details. Maxpooling also aids to make the model as insensitive to the location of features in the input.
- **Batch normalization:** This layer ensures that the convolutional layers in the model learn independently without overfitting. It normalized the output obtained in previous layer. It is a method of regularization. It is applied on every batch of input. Normalization aids to avoid the influence of biased learning and speeds up the training process. Let  $X$  be the input ( $W, b$ ) be the weights associated hidden neuron. The output of neural network without batch normalization will be computed as:  $\phi(f(W, X) + b)$  where  $\phi$  corresponds to the activation function and  $f$  is the linear transformation function. When batch normalization is incorporated, this is modified as:  $\phi\left(\frac{(f(W, X) + b) - \mu}{\sigma}\right)\alpha + \beta$  where,  $\mu, \sigma$  denotes the mean and standard deviation of the neuron's output,  $\alpha, \beta$  are learning parameters.
- **Global Average pooling:** Similar to maxpooling, global average pooling is a downsampling technique. This layer replaces fully connected layers, by considering the average of feature maps which is vectorized and then fed to the activation layer. It reduces the number of learning parameters. Computing the average

aids to preserve the correspondence between feature maps and the outputs. Max-pooling captures the edge details or variations in the image. Average pooling on the other hand, extracts the features smoothly. In a nutshell, it avoids overfitting to a certain extent. This layer does not introduce any learnable parameters to the model.

- **Dropout:** Dropout is similar to a  $L_2$  regularization method adapted to reduce overfitting. Certain number of neurons are randomly selected and deactivated to make the model more robust in terms of features. Dropout can be applied to any of the hidden layers in the neural network. Dropout layer does not have any learnable parameters.
- **Upsampling or transpose of convolution:** In tasks such as semantic segmentation, it is necessary to ensure that the output feature map of the U-net must be same as that of the input image resolution. However, adding layers such as strided convolution and maxpooling reduces the image resolution. To restore the size, upsampling or transpose of convolution operation is performed. In case of upsampling, as the pixel values are lost during maxpool operation, it uniformly replicates the intensity values to a higher resolution. It cannot restore any lost details from the feature map. Upsampling can be applied to ensure that the concatenated layers have equal resolution, specifically in skip connection based network architecture like Unet model.
- **Dense connection:** Dense layer of a convolutional neural network connects all outputs of a certain stage (preceding) to every other input of the next stage. Dense layer preserves the feed-forward nature of the neural network. As the number of connections increases, the associated learnable parameters such as weights of the link also increases. This layer is generally used as the last layer of convolutional neural network. It is useful when there is association between any feature maps and outputs.



## C.2 ACTIVATION FUNCTIONS

Activation function are the mathematical functions that determine the output of a neural network. It also helps to normalize the input to some extent. For instance, using activation functions like simoid, softmax, or tanh, buonds the output values between 0 to 1 or -1 to +1. Any non-linear (or linear) function that continuously differentiable, monotonic, and has a range of values can be used as an activation function. The non-linearity introduced by activation functions aids to learn the features through backpropagation. The most common activation functions that are used in this thesis work are defined in Table C.1

Table C.1: Activation functions.

Name	Definition	Remarks
Linear	$\phi(x) = mx$ where $m$ is slope, and $x$ is input.	It ranges from $-\infty$ to $+\infty$ . It is used in the final/ output layer, establishing a linear relationship between the feature maps and the output classes.
Sigmoid	$\phi(x) = \frac{1}{1+e^{-x}}$ .	It is non-linear activation function that bounds the output values between 0 to 1. It is used in binary classification, where output is the prediction probability for a particular class.
Softmax	$\phi(x) = \frac{e^x}{\sum_{k=1}^I e^{x_k}}$ , where $I$ is total number of inputs.	It can be used at the output layer when the classes are mutually exclusive.
Relu	$\phi(x) = \max(0, x)$ .	It is rectified linear unit. it is computationally less expensive compared to tanh and softmax. It activates positive valued neurons.
Leaky Relu	$\phi(y) = \begin{cases} \alpha y & y \leq 0 \\ y & y > 0 \end{cases}$ , where $\alpha$ is a constant.	The problem of large negative bias of relu is avoided by using leaky relu. This adds a small negative slope and prevents vanishing gradient problem.
Tanh	$\phi(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ .	It helps to center the data around zero, making the learning easier.

### **C.3 HYPERPARAMETERS**

Model parameters like weights and biases are derived during training process. However, hyperparameters are those influential parameters which controls the convergence rate and the learning process. It needs to be initialized before the training process. A slight modification in the hyperparameters can lead to a different result. Therefore, to identify the values of hyperparameters that gives best model performance, techniques such as grid search, random search and Bayesian optimization is used. Some of the hyperparameters are epochs, batch size, number of neurons, optimizers, and learning rate. Optimizers are used to iteratively find the global minima of a minimization problem. Gradient descent is most commonly used optimizer. It uses gradient of the objective function to navigate the search space. However it leads to vanishing gradient problem. Therefore, to accelerate the learning, momentum is provided to optimizers. Adam, Adagrad, and RMSProp (Root Mean Square) are examples of fast optimizers. RMSProp is an adaptive optimization technique. Further details can be obtained from Goodfellow et al. (2016). In Chapter 4 of the thesis, the hyperparameters are set based on empirical study and recommendations, and therefore, identifying precise set of hyperparameters using Bayesian learning, to further improve the performance of the model is left for future work.

## **APPENDIX IV**

The standard image quality metrics, classification metrics, and certain other statistical measures are defined in this section along with an insight into the Gabor features.

### **D.1 STANDARD METRICS FOR QUANTITATIVE ANALYSIS**

#### **D.1.1 Jensen–Shannon divergence (JSD)**

The Jensen–Shannon divergence (JSD) is a statistical measure that is generally used to compute the similarity between two distributions. Given two distributions in discrete

domain,  $F_1$  and  $F_2$ , the JSD is calculated as:

$$JSD(F_1||F_2) = \frac{Q(F_1||F_2)}{2} + \frac{Q(F_2||F_1)}{2},$$

$$\text{where } Q(A||B) = \sum_i A(i) * \log_{10} \left( \frac{A(i)}{B(i)} \right) \quad (\text{D.1})$$

$$\text{and } Z = \frac{(F_1 + F_2)}{2}.$$

The square root of the divergence is denoted as a distance metric. JSD is a symmetric measure and it gives a finite value. Refer Fuglede and Topsoe (2004) for further details.

### D.1.2 Global Contrast Factor (GCF)

Global contrast Factor (GCF) corresponds to the human perception of contrast. A local contrast is the average difference between neighboring pixels. The GCF is measured using:

$$GCF = \sum_{i=1}^T \delta_i \mathcal{C}_i, \quad (\text{D.2})$$

$$\text{where } \mathcal{C} = \frac{1}{width \times height} \sum_{j=1}^{width \times height} lc_j.$$

where  $\delta$  is weight associated with the local average contrast,  $lc_j$  is the average local contrast of all pixels in the image of size  $width \times height$ , and  $T$  is the number of iterations for which the local average is measured. In each iteration, the image pixels are combined to vary the image resolution.

### D.1.3 Entropy

Entropy is defined by:

$$ENTROPY = - \sum_{j=0}^{MAX} p_j \log_{10} p_j. \quad (\text{D.3})$$

where  $MAX$  corresponds to the maximum possible intensity value in the image,  $p_j$  is the probability of each intensity value in the image. Entropy is smaller if the image is degraded due to low contrast.

### D.1.4 Peak Signal to Noise Ratio (PSNR)

The Peak Signal to Noise Ratio (PSNR) quantifies the extent of denoising an image. It is a full-reference metric. Let  $U_0$  be the observed noisy image, and  $U$  be the clean

image, then PSNR is defined by:

$$PSNR = 10 \log_{10} \left( \frac{\mathcal{M}_U^2}{MSE} \right), \quad (D.4)$$

where  $\mathcal{M}_U^2$  represents maximum possible intensity (pixel) value in the image,  $MSE$  corresponds to mean square error given as:

$$MSE = \frac{1}{width \times height} \sum_{i=0}^{width-1} \sum_{j=0}^{height-1} (U(i, j) - U_0(i, j))^2. \quad (D.5)$$

If the noisy image or the restored image significantly differ from the clean image, the mean square error increases, leading to a lower PSNR value. As the amount of noise decreases in the image, the PSNR value increases, conveying a better image quality and a better restoration.

### D.1.5 Structural Similarity (SSIM)

The Structural similarity is recommended by Wang et al. (2004). It is used to analyze the noise reduction, structure and contrast preservation. SSIM is evaluated using:

$$SSIM(u_0, u) = \frac{(2\mu_{u_0}\mu_u + \delta_1)(2\sigma_u + \delta_2)}{(\mu_{u_0}^2 + \mu_u^2 + \delta_1)(\sigma_{u_0}^2 + \sigma_u^2 + \delta_2)}. \quad (D.6)$$

where  $\delta_1, \delta_2$  are constants added to stabilize the values,  $\mu_{u_0}, \mu_u$  denotes the mean,  $\sigma_{u_0}^2, \sigma_u^2$  denotes variances of  $u_0$  &  $u$ , respectively. It signifies the closeness or similarity of the two images or patches under evaluation.

### D.1.6 Contrast Noise Ratio (CNR)

Contrast Noise Ratio (CNR), introduced by Timischl (2015), is a metric that quantifies the contrast level of the image similar to PSNR. It is computed as:

$$CNR_{dB} = 10 \log_{10} \frac{|M_f - M_b|}{\sqrt{0.5(V_f + V_b)}}. \quad (D.7)$$

Here,  $M_f, M_b, V_f, V_b$  represents the mean and variance of small homogeneous patches considered in foreground and background region, respectively. These patches are selected intuitively such that there is considerable change in the intensity levels of the patches. Similar to the PSNR metric, a better image quality or better contrast is conveyed with higher CNR value.

### D.1.7 Equivalent Number of Looks (ENL)

Equivalent Number of Looks (ENL) is a blind or referenceless image quality metric (Refer Gomez et al. (2017)). It is calculated on homogeneous regions or patches of the

images, and is given by:

$$ENL = \frac{\mu_r^2}{\sigma_r^2}. \quad (D.8)$$

where  $\mu_r, \sigma_r^2$  are the mean and variance measured in the patch of an image. Greater the variance, smaller is the ENL value, indicating that the image is noisy.

### D.1.8 Wasserstein Distance

Wasserstein GAN is based on the Wasserstein distance metric (Vaserstein (1969)). This distance is also called as Earth Mover's distance. Traditional GANs are trained on the basis of JSD. However, since Wasserstein distance is continuous and differentiable everywhere, it allows the stable training of the GANs using optimizers. Given two distributions, this metric approaches to zero when the two distributions are similar. In discrete domain, it is measured as :

$$W_d = \inf_{i,j} \|i - j\|_p, \quad (D.9)$$

where (i,j) corresponds to the pair of random variables in corresponding cumulative distribution function  $F_1, F_2$ .

### D.1.9 Classification metrics

The classification metrics such as accuracy, sensitivity, specificity, and F-1 score are defined from a confusion matrix (Grandini et al. (2020)). Let  $TP, TN, FP, FN$  denote the total true positives, true negatives, false positives, and false negatives respectively in a binary class confusion matrix.

1. **Accuracy:** Accuracy conveys how often the classifier predicts correctly. It includes correct prediction of true positives and true negatives. A higher value of accuracy indicates better classifier performance.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}.$$

2. **Sensitivity:** It represents the number of correct predictions. It is also referred as Recall measure. A value closer to 1 is considered to be ideal for a classifier.

$$\text{Sensitivity} = \frac{TP}{TP+FN}.$$

3. **Specificity:** The measure of correct negative predictions is obtained from specificity. The sensitivity and specificity are important to know about the type I and

type II errors.

$$\text{Specificity} = \frac{TN}{TN+FP}.$$

4. **Precision:** Precision is also referred as positive predictive value. A higher value of precision indicates better (reliable) classifier.

$$\text{Precision} = \frac{TP}{TP+FP}.$$

5. **F1-Score:** Often, precision and recall are not sufficient to understand the reliability of the classifier. In other words, a low precision and a high recall (or vice-versa) may not guarantee a better classifier. F1-score denotes the harmonic mean between precision and recall measures. A desirable classifier exhibits higher F1-score with high precision and high recall values.

$$\text{F1-score} = \frac{2*TP}{2*TP+FP+TN}.$$

6. **Kappa:** This measure is used as interrater reliability (McHugh (2012)). Its value can range from -1 to +1, signifying the amount of agreement between graders. The value closer to +1 indicates stronger agreement and negative values indicate no agreement. The Choen's kappa statistic is measured as  $k = \frac{p_o - p_e}{1 - p_e}$  where  $p_o$  is relative observed agreement among raters and  $p_e$  is hypothetical probability of chance agreement.

7. **Dice coefficient:** It is the ratio of twice the area of overlap between the groundtruth and predicted image to the total number of pixels. When the segmented image is much similar to the groundtruth image, the overlap is more and hence the dice coefficient increases. It is expressed as F1-score when it is derived from the confusion matrix.

8. **Area under the curve of Receiver Characteristic Operator (AUC):** Receiver Characteristic Operator (ROC) is a plot of true positive rate against the false positive rate. AUC is a measure of ability of a classifier to distinguish between classes. It is the area measured using ROC plot. A higher value (closer to 1) represents that the classifier model is able to distinguish between positive and negative classes precisely.

## D.2 GABOR FILTER

The 2-D Gabor filter Weldon et al. (1996) is given as:

$$G = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left[ \frac{-1}{2} \left( \left( \frac{x}{\sigma_x} \right)^2 + \left( \frac{y}{\sigma_y} \right)^2 \right) + j\omega(x\cos\theta + y\sin\theta) \right]. \quad (\text{D.10})$$

where  $\sigma$  denotes spatial spread,  $\theta$  represents orientation, and  $\omega$  is the frequency. To enhance the OCT images, as in Chapter 4 of the thesis,  $\theta$  is considered as  $0^\circ$ ,  $45^\circ$ , and  $90^\circ$ ,  $\sigma$  is 2, and  $\omega$  is empirically set to 0.25. The OCT scans have prominent features in the horizontal direction. Hence, the Gabor features are extracted at angles  $0^\circ$ ,  $45^\circ$ , and  $90^\circ$ .





## Bibliography

- Abràmoff, M. D., Garvin, M. K., and Sonka, M. (2010). Retinal imaging and image analysis. *IEEE reviews in biomedical engineering*, 3(1), 169–208.
- Agrawal, V., Kori, A., Alex, V., and Krishnamurthi, G. (2018). Enhanced optic disk and cup segmentation with glaucoma screening from fundus images using position encoded cnns. *arXiv preprint arXiv:1809.05216*, 1–8.
- Akram, M. U., Akbar, S., Hassan, T., et al. (2020). Data on fundus images for vessels segmentation, detection of hypertensive retinopathy, diabetic retinopathy and papilledema. *Data in Brief*, 29(1), 105282.
- Alqudah, A. (2020). Aoct-net: a convolutional network automated classification of multiclass retinal diseases using spectral-domain optical coherence tomography images. *Med Biol Eng Comput*, 58(1), 41–53.
- Alwazzan, M. J., Ismael, M. A., and Ahmed, A. N. (2021). A hybrid algorithm to enhance colour retinal fundus images using a wiener filter and CLAHE. *Journal of Digital Imaging*, 34(3), 750–759.
- Anoop, B., Girish, G., Sudeep, P., and Rajan, J. (2019). Despeckling algorithms for optical coherence tomography images. In *Advanced Classification Techniques for Healthcare Analysis*, 286–310. IGIGlobal.
- Apostolopoulos, S., Salas, J., Ordóñez, J. L. P., et al. (2020). Automatically enhanced OCT scans of the retina: A proof of concept study. *Scientific Reports*, 10(1), 1–8.
- Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein generative adversarial networks. In *International conference on machine learning*, 214–223. PMLR.

- Asgari, R., Waldstein, S., Schlanitz, F., et al. (2019). U-net with spatial pyramid pooling for drusen segmentation in optical coherence tomography. In Fu, H., Garvin, M., MacGillivray, T., et al., editors, *International Workshop on Ophthalmic Medical Image Analysis*, 11855, 77–85. Springer, Shenzhen, China.
- Aubert, G. and Aujol, J.-F. (2008). A variational approach to removing multiplicative noise. *SIAM Journal on Applied Mathematics*, 68(4), 925–946.
- Balaji, B. and Jidesh, P. (2017). Non-local gradient fidelity model for multiplicative gamma noise removal. In *2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR)*, 1–6. IEEE.
- Batard, T., Haro, G., and Ballester, C. (2021). Dip-vbtv: A color image restoration model combining a deep image prior and a vector bundle total variation. *SIAM Journal on Imaging Sciences*, 14(4), 1816–1847.
- Bellemo, V., Burlina, P., Yong, L., et al. (2018). Generative adversarial networks (gans) for retinal fundus image synthesis. In Carneiro and adn Shaodi, G., editors, *Asian Conference on Computer Vision*, 289–302, Perth, Australia. Springer International Publishing.
- Bertalmio, M., Caselles, V., and Provenzi, E. (2009). Issues about retinex theory and contrast enhancement. *Int. J. Comput. Vis.*, 83(1), 101–119.
- Bourne, R., Flaxman, S., Braithwaite, T., et al. (2017). Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *The Lancet Global Health*, 5(9), 1–10.
- Buades, A., Coll, B., and Morel, J.-M. (2005). A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2 of 1, 60–65, San Diego, CA, USA. IEEE.
- Budai, A., Bock, R., Maier, A., et al. (2013). Robust vessel segmentation in fundus images. *International Journal of Biomedical Imaging*, 2013(1), 1–11.

- Chhablani, J., Krishnan, T., Sethi, V., and Kozak, I. (2014). Artifacts in optical coherence tomography. *Saudi Journal of Ophthalmology*, 28(2), 81–87.
- Chiu, S. J., Li, X. T., Nicholas, P., et al. (2010). Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation. *Optics Express*, 18(18), 19413.
- Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1251–1258. IEEE.
- Dai, Z., Yang, Z., Yang, F., et al. (2017). Good semi-supervised learning that requires a bad GAN. *CoRR*, abs/1705.09783.
- Das, V., Dandapat, S., and Bora, P. K. (2020). A data-efficient approach for automated classification of OCT images using generative adversarial network. *IEEE Sensors Letters*, 4(1), 1–4.
- Duan, J., Lu, W., Tench, C., et al. (2016). Denoising optical coherence tomography using second order total generalized variation decomposition. *Biomedical Signal Processing and Control*, 24, 120–127.
- Ehrlich, R., Harris, A., Wentz, S. M., et al. (2011). Anatomy and regulation of the optic nerve blood flow. In Dean, B. and Joseph, B., editors, *The Retina and Its Disorders*, 1 of 1, chapter 5, 28–38. Elsevier Science, San Diego, USA, 1 edition.
- El-Baz, A. (2018). Classification of retinal diseases based on OCT images. *Frontiers in Bioscience*, 23(1), 247–264.
- Fan, W., Yu, H., Chen, T., and Ji, S. (2020). Oct image restoration using non-local deep image prior. *Electronics*, 9(5), 784.
- Fang, L., Shutao, L., Qing, N., et al. (2012). Sparsity based denoising of spectral domain optical coherence tomography images. *Biomed. Opt. Express*, 3(5), 927–942.

- Fauw, J. D., Ledsam, J. R., Romera-Paredes, B., et al. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, 24(9), 1342–1350.
- Febin, I. P. and Jidesh, P. (2021). Despeckling and enhancement of ultrasound images using non-local variational framework. *The Visual Computer*, 2021(1), 1–14.
- Febin, I. P., Jidesh, P., and Bini, A. A. (2018). Noise classification and automatic restoration system using non-local regularization frameworks. *The Imaging Science Journal*, 66(8), 479–491.
- Febin, I. P., Jidesh, P., and Bini, A. A. (2020). A retinex-based variational model for enhancement and restoration of low-contrast remote-sensed images corrupted by shot noise. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 941–949.
- Franceschetti, G., Pascazio, V., and Schirinzi, G. (1995). Iterative homomorphic technique for speckle reduction in synthetic-aperture radar imaging. *Journal of the Optical Society of America A*, 12(4), 686.
- Fraz, M. M., Remagnino, P., Hoppe, A., et al. (2012). An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Transactions on Biomedical Engineering*, 59(9), 2538–2548.
- Fuglede, B. and Topsoe, F. (2004). Jensen-shannon divergence and hilbert space embedding. In *International Symposium on Information Theory, 2004. ISIT 2004. Proceedings.*, page 31, Chicago, IL, USA. IEEE.
- Fujimoto, J. G., Pitris, C., Boppart, S. A., and Brezinski, M. E. (2000). Optical coherence tomography: An emerging technology for biomedical imaging and optical biopsy. *Neoplasia*, 2(1-2), 9–25.
- Garcia-Layana, A., Ciuffo, G., Zarranz-Ventura, J., and Alvarez-Vidal, A. (2017). Optical coherence tomography in age-related macular degeneration. <https://amdbook.org/content/>

optical-coherence-tomography-age-related-macular-degeneration.

(Accessed:20-01-2021).

Gariepy, R. F. (2001). FUNCTIONS OF BOUNDED VARIATION AND FREE DISCONTINUITY PROBLEMS (oxford mathematical monographs) by LUIGI AMBROSIO, NICOLO FUCSO and DIEGO PALLARA: 434 pp., £55.00, ISBN 0-19-850254-1 (clarendon press, oxford, 2000). *Bulletin of the London Mathematical Society*, 33(4), 492–512.

George, T. T. and Michael, K. (2005). The direct ophthalmoscope how it works and how to use it. The MIT Media Lab, University of Kansas. <https://web.media.mit.edu/~raskar/Eye/TheDirectOphthalmoscope.pdf> (Accessed on 13-06-2020).

Gholami, P., Roy, P., Parthasarathy, M. K., and Lakshminarayanan, V. (2020). OCTID: Optical coherence tomography image database. *Computers & Electrical Engineering*, 81(1), 106532.

Gilboa, G. and Osher, S. (2009). Nonlocal operators with applications to image processing. *Multiscale Modeling & Simulation*, 7(3), 1005–1028.

Girish, G., Thakur, B., Chowdhury, S. R., et al. (2018). Segmentation of intra-retinal cysts from optical coherence tomography images using a fully convolutional neural network model. *IEEE journal of biomedical and health informatics*, 23(1), 296–304.

Glen, S. (2014). P-value in statistical hypothesis tests: What is it?". StatisticsHowTo.com: Elementary Statistics for the rest of us! <https://www.statisticshowto.com/p-value/> (Accessed:20-01-2021).

Goldstein, T., Bresson, X., and Osher, S. (2009). Geometric applications of the split bregman method: Segmentation and surface reconstruction. *Journal of Scientific Computing*, 45(1-3), 272–293.

Goldstein, T. and Osher, S. (2009). The split bregman method for  $l_1$ -regularized problems. *SIAM Journal on Imaging Sciences*, 2(2), 323–343.

- Gomez, L., Ospina, R., and Frery, A. C. (2017). Unassisted quantitative evaluation of despeckling filters. *Remote Sensing*, 9(4), 389.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). Convolutional networks. In *Deep learning*, Adaptive Computation and Machine Learning series, chapter 9. MIT press, London, England. <https://books.google.co.in/books?id=omivDQAAQBAJ>.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27(1), 1–9.
- Gour, N. and Khanna, P. (2021). Multi-class multi-label ophthalmological disease detection using transfer learning based convolutional neural network. *Biomedical Signal Processing and Control*, 66, 102329.
- Grandini, M., Bagli, E., and Visani, G. (2020). Metrics for multi-class classification: an overview. *arXiv,stat.ML*, 2008.05756. Accessed: 23-06-2021.
- Gulrajani, I., Ahmed, F., Arjovsky, M., et al. (2017). Improved training of wasserstein gans. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS’17, page 5769–5779, Red Hook, NY, USA. Curran Associates Inc.
- Guo, A., Fang, L., Qi, M., and Li, S. (2021). Unsupervised denoising of optical coherence tomography images with nonlocal-generative adversarial network. *IEEE Transactions on Instrumentation and Measurement*, 70(1), 1–12.
- Hadamard, J. (1953). Cauchy’s fundamental theorem. In Morse, P., editor, *Lectures on Cauchy’s problem in linear partial differential equations*, 8–18. Dover publications, New York, 6 edition.
- Haloi, M. (2018). Towards ophthalmologist level accurate deep learning system for OCT screening and diagnosis. *CoRR*, abs/1812.07105.
- Haque, A. (2020). EC-GAN: low-sample classification using semi-supervised algorithms and gans. *CoRR*, abs/2012.15864.

- Hassan, B., Hassan, T., Li, B., et al. (2019). Deep ensemble learning based objective grading of macular edema by extracting clinically significant findings from fused retinal imaging modalities. *Sensors*, 19(13), 2970.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- Helga, K., Ralph, N., Peter, A., et al. (2007). The architecture of the human fovea. In Kolb, H., Fernandez, E., Nelson, R., and for Biotechnology Information (États-Unis), N. C., editors, *Webvision: The Organization of the Retina and Visual System*, NCBI Bookshelf. National Library of Medicine (US).
- Holla Kayyar, S. and Jidesh, P. (2018). Non-local total variation regularization approach for image restoration under a poisson degradation. *Journal of Modern Optics*, 65(19), 2231–2242.
- Hoover, A., Kouznetsova, V., and Goldbaum, M. (2000). Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging*, 19(3), 203–210.
- Hossein, R., Raheleh, K., and Zahra, A. (2016). Optical coherence tomography image analysis. *Wiley Encyclopedia of Electrical and Electronics Engineering*, 1(1), 1–16.
- Howard, A. G., Zhu, M., Chen, B., et al. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *CoRR*, abs/1704.04861.
- Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708.
- Huang, Z., Tang, C., Xu, M., et al. (2020). Both speckle reduction and contrast enhancement for optical coherence tomography via sequential optimization in the logarithmic domain based on a refined retinex model. *Appl. Opt.*, 59(35), 11087–11097.

- Islam, M. T., Imran, S. A., Arefeen, A., et al. (2019). Source and camera independent ophthalmic disease recognition from fundus image using neural network. In *2019 IEEE International Conference on Signal Processing, Information, Communication Systems (SPICSCON)*, 59–63.
- Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1 of 1, 1125–1134.
- Jason, B. (2019). What are generative adversarial networks. In *Generative Adversarial Networks with Python: Deep learning generative models for image synthesis and image translation*, 3–14. Machine Learning Mastery, v1.5 edition. <https://books.google.co.in/books?id=YBimDwAAQBAJ>.
- Jidesh, P. and Banothu, B. (2018). Image despeckling with non-local total bounded variation regularization. *Computers & Electrical Engineering*, 70(1), 631–646.
- Jidesh, P. and Febin, I. (2020). A perceptually inspired variational model for enhancing and restoring remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 18(2), 251–255.
- Jitendra, K. and Pradeep, T. (2019). Economical blindness in bundelkhand region: Analytical study. *IOSR Journal of Dental and Medical Sciences IOSR-JDMS*, 18(6), 23–27.
- Jonathan, A. and Micieli, M. (2019). Optical coherence tomography. In *Case-Based Neuro-Ophthalmology*, chapter 2. Independently published. <https://casebasedneuroophthalmology.pressbooks.com/chapter/optical-coherence-tomography/>.
- Jordi, C., Joan, M. N., and Carles, V. (2019). Ocular disease intelligent recognition through deep learning architectures. *Universitat Oberta de Catalunya*, page 1–114. <http://openaccess.uoc.edu/webapps/o2/handle/10609/113126> (Accessed on: 18-8-2020).



- Jorjandi, S., Rabbani, H., Kafieh, R., and Amini, Z. (2017). Statistical modeling of optical coherence tomography images by asymmetric normal laplace mixture model. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 4399–4402.
- Kermany, D., Zhang, K., Goldbaum, M., et al. (2018a). Labeled optical coherence tomography (oct) and chest x-ray images for classification. *Mendeley Data*, 2(2). <https://www.kaggle.com/paultimothymooney/kermany2018>.
- Kermany, D. S., Goldbaum, M., Cai, W., et al. (2018b). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5), 1122–1131.
- Khurana, A. (2015). Physiology of eye and vision. In *Comprehensive Ophthalmology*, 1 of 1, chapter 2. Jaypee Brothers Medical Publishers, India, 7 edition.
- Kim, Y. and Kudo, H. (2020). Nonlocal total variation using the first and second order derivatives and its application to ct image reconstruction. *Sensors*, 20(12), 3494.
- Kimmel, R., Elad, M., Shaked, D., et al. (2003). A variational framework for retinex. *International Journal of computer vision*, 52(1), 7–23.
- Kuan, D. T., Sawchuk, A. A., Strand, T. C., and Chavel, P. (1985). Adaptive noise smoothing filter for images with signal-dependent noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7(2), 165–177.
- Lam, C., Yu, C., Huang, L., and Rubin, D. (2018). Retinal lesion detection with deep learning using image patches. *Investigative Ophthalmology & Visual Science*, 59(1), 590.
- Lecouat, B., Chang, K., Foo, C.-S., et al. (2018). Semi-supervised deep learning for abnormality classification in retinal images. *ArXiv*, abs/1812.07832.
- Lee, J.-S. (1981). Speckle analysis and smoothing of synthetic aperture radar images. *Computer Graphics and Image Processing*, 17(1), 24–32.

- Leopold, H. A., Orchard, J., Zelek, J. S., and Lakshminarayanan, V. (2019). Pixelbnn: Augmenting the pixelcnn with batch normalization and the presentation of a fast architecture for retinal vessel segmentation. *Journal of Imaging*, 5(2), 26.
- Li, C., Ye, J., He, J., et al. (2020a). Dense correlation network for automated multi-label ocular disease detection with paired color fundus photographs. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 1–4.
- Li, H., Zhang, L., and Shen, H. (2012). Perceptually inspired variational method for the uneven intensity correction of remote sensing images. *IEEE Transactions on Geosci. and Remote Sens.*, 50(8), 3053–3065.
- Li, N., Li, T., Hu, C., et al. (2021). A benchmark of ocular disease intelligent recognition: One shot for multi-disease detection. In Wolf, F. and Gao, W., editors, *Benchmarking, Measuring, and Optimizing*, 177–193. Springer International Publishing.
- Li, Q., Li, S., He, Z., et al. (2020b). DeepRetina: Layer segmentation of retina in OCT images using deep learning. *Translational Vision Science & Technology*, 9(2), 61.
- Li, X., Shen, L., Shen, M., and Qiu, C. S. (2019). Integrating handcrafted and deep features for optical coherence tomography based retinal disease classification. *IEEE Access*, 7(1), 33771–33777.
- Liu, G., Wang, Z., Mu, G., and Li, P. (2018). Efficient OCT image enhancement based on collaborative shock filtering. *Journal of Healthcare Engineering*, 2018(1), 1–7.
- Liu, L., Pang, Z. F., and Duan, Y. (2017). A novel variational model for retinex in presence of severe noises. In *IEEE International Conference on Image Processing (ICIP)*, 3490–3494. IEEE.
- Liu, X. and Huang, L. (2010). Split bregman iteration algorithm for total bounded variation regularization based image deblurring. *Journal of Mathematical Analysis and Applications*, 372(2), 486–495.
- Lou, S., Chen, X., Liu, J., et al. (2021). Fast OCT image enhancement method based on the sigmoid-energy conservation equation. *Biomedical Optics Express*, 12(4), 1792.

- Malik, S., Kanwal, N., Asghar, M. N., et al. (2019). Data driven approach for eye disease classification with machine learning. *Applied Sciences*, 9(14), 2789.
- Malvika, A. and Nadia, K. W. (2019). Optical coherence tomography angiography. In Ehlers, J., editor, *The Retina Illustrated*, chapter 4, 40–45. Thieme, Ohio, 1 edition. <https://books.google.co.in/books?id=6UqyDwAAQBAJ>.
- Matković, K., Neumann, L., Neumann, A., et al. (2005). Global contrast factor - a new approach to image contrast. In *Proceedings of the First Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging*, Computational Aesthetics'05, page 159–167, Goslar, DEU. Eurographics Association.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3), 276–282.
- Mehdizadeh, M., MacNish, C., Xiao, D., et al. (2021). Deep feature loss to denoise OCT images using deep neural networks. *Journal of Biomedical Optics*, 26(04).
- Meller, G. (2020). Ocular disease recognition using convolutional neural networks. *University of Milan*. <https://towardsdatascience.com/ocular-disease-recognition-using-convolutional-neural-networks-c04d63a7a2da> (Accessed on 24-10-2020).
- Meyer, M., Galdran, A., and Mendonca, A.M. and Campilho, A. (2018). A pixel-wise distance regression approach for joint retinal optical disc and fovea detection. In Frangi, A. and Schnabel, J., Davatzikos, C., Alberola-López, C., and Fichtinger, G., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, 11071, 39–47. Lecture Notes in Computer Science, Springer, Cham.
- Mir, N., Khan, M. A., and Hussain, M. G. (2020). Portable smart healthcare solution to eye examination for diabetic retinopathy detection at an earlier stage. In Lytras, M. D. and Sarirete, A., editors, *Innovation in Health Informatics*, 1 of 1, chapter 13, 305–322. Elsevier, Academic Press, Saudi Arabia, 1 edition.
- Mohammad-Djafari, A. (2002). Bayesian inference for inverse problems. In *AIP Conference Proceedings*, 617 of 1, 477–496. American Institute of Physics.

- Muxingzi, L., Ramzi, I., Biswarup, C., and Wolfgang, H. (2017). Statistical model for oct image denoising. *Biomed. Opt. Express*, 8(9), 3903–3917.
- Negi, A., Chauhan, P., Kumar, K., and Rajput, R. (2020). Face mask detection classifier and model pruning with keras-surgeon. In *2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, 1–6, Jaipur, India.
- Ng, M. K. and Wang, W. (2011). A total variation model for retinex. *SIAM Journal on Imaging Sciences*, 4(1), 345–365.
- Odena, A. (2016). Semi-Supervised Learning with Generative Adversarial Networks. *arXiv e-prints*, page arXiv:1606.01583.
- Odena, A., Olah, C., and Shlens, J. (2017). Conditional image synthesis with auxiliary classifier gans. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70, ICML'17*, page 2642–2651, Sydney, NSW, Australia. JMLR.org.
- Orlando, J. I., Fu, H., Breda, J. B., et al. (2020). Refuge-challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Medical Image Analysis*, 59(101570), 1–21.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *PLOS Medicine*, 18(3), e1003583.
- Pan, X., Zhan, X., Dai, B., et al. (2020). *Exploiting deep generative prior for versatile image restoration and manipulation*, 262–277. Springer.
- Panwar, N., Huang, P., Lee, J., et al. (2016). Fundus photography in the 21st century—a review of recent technological advances and their implications for worldwide health-care. *Telemedicine and e-Health*, 22(3), 198–208.
- Paul, D., Tewari, A., Ghosh, S., and Santosh, K. (2020). Octx: Ensembled deep learning model to detect retinal disorders. In *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, 526–531, Rochester, MN, USA. IEEE.

- Paul, K. and Kumar, V. (2015). Fundus imaging based affordable eye care. In *Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 5*, BIOSTEC 2015, page 634–641, Setubal, PRT. SCITEPRESS - Science and Technology Publications, Lda.
- Porwal, P., Pachade, S., Kamble, R., et al. (2018a). Indian diabetic retinopathy image dataset (idrid): a database for diabetic retinopathy screening research. *Data*, 3(3), 25.
- Porwal, P., Pachade, S., Kamble, R., et al. (2018b). Indian diabetic retinopathy image dataset (IDRiD): A database for diabetic retinopathy screening research. *Data*, 3(3), 25.
- Porwal, P., Pachade, S., Kokare, M., et al. (2020). Idrid: Diabetic retinopathy–segmentation and grading challenge. *Medical image analysis*, 59, 101561.
- Raeba, M. and Sobha, S. (2014). Age-related macular degeneration. In Amresh, C. and Tin, A., editors, *Multimodal Retinal Imaging*, 1 of 1, chapter 9, 85–101. JP medical publishers, India, 1 edition.
- Rahimizadeh, N., Hasanzadeh, R. P., and Janabi-Sharifi, F. (2020). An optimized non-local LMMSE approach for speckle noise reduction of medical ultrasound images. *Multimedia Tools and Applications*, 80(6), 9231–9253.
- Raj, A., Shah, N. A., and Tiwari, A. K. (2022). A novel approach for fundus image enhancement. *Biomedical Signal Processing and Control*, 71, 103208.
- Raj, A., Tiwari, A. K., and Martini, M. G. (2019). Fundus image quality assessment: survey, challenges, and future scope. *IET Image Processing*, 13(8), 1211–1224.
- Ram, A. and Reyes-Aldasoro, C. C. (2020). The relationship between fully connected layers and number of classes for the analysis of retinal images. *CoRR*, abs/2004.03624, 1–10.
- Ranftl, R., Bredies, K., and Pock, T. (2014). Non-local total generalized variation for optical flow estimation. In Fleet, D., Pajdla, T., Schiele, B., and Tuytelaars,

- T., editors, *Computer Vision – ECCV 2014*, 439–454, Lecture Notes in Computer Science. Springer, Cham.
- Resnikoff, S., Lansingh, V. C., Washburn, L., et al. (2019). Estimated number of ophthalmologists worldwide (International Council of Ophthalmology update): will we meet the needs? *British Journal of Ophthalmology*, 104(4), 588–592.
- Reza, R., Hossein, R., Alireza, M., and Fedra, H. (2018). Macular oct classification using a multi-scale convolutional neural network ensemble. *IEEE Transactions on Medical Imaging*, 37(4), 1024–1034.
- Roy, A. G., Conjeti, S., Karri, S. P. K., et al. (2017). Relaynet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. *Biomedical optics express*, 8(8), 3627–3642.
- Rudin, L., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1), 259–268.
- Ryan, S., Schachat, A., Wilkinson, C., et al. (2013). Optical coherence tomography. In Srinivas, R. S., editor, *Retina E-Book*, 1, chapter 3, 82–111. Elsevier Health Sciences, China, 5 edition.
- Salimans, T., Goodfellow, I., Zaremba, W., et al. (2016). Improved techniques for training gans. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS’16, page 2234–2242, Red Hook, NY, USA. Curran Associates Inc.
- Sandeep, B. (2019). Challenges and opportunities of ophthalmic care in india. *Express Healthcare, EH News Bureau*. <https://www.expresshealthcare.in/blogs/guest-blogs-healthcare/challenges-and-opportunities-of-ophthalmic-care-in-india/411165/> (Accessed on: 05-09-2021).
- Sarki, R., Ahmed, K., Wang, H., and Zhang, Y. (2020). Automatic detection of diabetic eye disease through deep learning using fundus images: A survey. *IEEE Access*, 8, 151133–151149.

- Schaal, S., ElTanboly, A., Ismail, M., et al. (2016). A novel automated method for the objective quantification of retinal layers reveals sequential changes that occur in the normal retina with age. *Front Nanosci Nanotech*, 2(5), 1–6.
- Schmitt, J. M., Xiang, S., and Yung, K. (1999). Speckle in optical coherence tomography. *Journal of biomedical optics*, 4(1), 95–105.
- Sedai, S., Antony, B., Rai, R., et al. (2019). Uncertainty guided semi-supervised segmentation of retinal layers in oct images. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, page 282–290, Berlin, Heidelberg. Springer-Verlag.
- Sengupta, S., Singh, A., Leopold, H. A., and Lakshminarayanan, V. (2018). Ophthalmic diagnosis and deep learning - a survey. *CoRR*, abs/1812.07101. <http://arxiv.org/abs/1812.07101>.
- Serener, A. and Serte, S. (2019). Dry and wet age-related macular degeneration classification using oct images and deep learning. In *2019 Scientific Meeting on Electrical-Electronics Biomedical Engineering and Computer Science (EBBT)*, 1–4. IEEE.
- Sharma, S., Shivhare, S. N., Singh, N., and Kumar, K. (2018). Computationally efficient ann model for small-scale problems. In *Machine Intelligence and Signal Analysis*, 423–435. Springer.
- Shen, Z., Fu, H., Shen, J., and Shao, L. (2021). Modeling and enhancing low-quality retinal fundus images. *IEEE Transactions on Medical Imaging*, 40(3), 996–1006.
- Shengjian, H., Chen, T., Min, X., et al. (2019). Bm3d-based total variation algorithm for speckle removal with structure-preserving in oct images. *Appl. Opt.*, 58(23), 6233–6243.
- Sidorenko, S. (2018). Progressive gan for image denoising. *Github Repository*. [https://github.com/ssidorenko/pg\\_dip/blob/master/report.pdf](https://github.com/ssidorenko/pg_dip/blob/master/report.pdf) (Accessed on 17-09-2019).

- Simard, P. Y., Steinkraus, D., Platt, J. C., et al. (2003). Best practices for convolutional neural networks applied to visual document analysis. In *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.*, 3 No. 2003, Edinburgh, UK.
- Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In Bengio, Y. and LeCun, Y., editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings.*
- Singh, V. K., Rashwan, H. A., Saleh, A., et al. (2018). REFUGE challenge 2018-task 2: Deep optic disc and cup segmentation in fundus images using u-net and multi-scale feature matching networks. *CoRR*, abs/1807.11433.
- Son, J., Park, S. J., and Jung, K. (2017). Retinal vessel segmentation in fundoscopic images with generative adversarial networks. *CoRR*, abs/1706.09318.
- Sonali, Sahu, S., Singh, A. K., et al. (2019). An approach for de-noising and contrast enhancement of retinal fundus image using CLAHE. *Optics & Laser Technology*, 110, 87–98.
- Song, D., Liu, Y., Lin, X., et al. (2019). Research and comparison of oct image speckle denoising algorithm. In *2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, 1 of 1, 1554–1558. IEEE.
- Sourya, S., Amitojdeep, S., and Vasudevan, L. (2021). Edgewavenet: edge aware residual wavelet gan for oct image denoising. In Deserno, T. M. and Park, B. J., editors, *Medical Imaging 2021: Imaging Informatics for Healthcare, Research, and Applications*, 11601, 110 – 115. International Society for Optics and Photonics, SPIE.
- Spaide, R. F., Fujimoto, J. G., and Waheed, N. K. (2015). Image artifacts in optical coherence tomography angiography. *Retina*, 35(11), 2163–2180.
- Staal, J., Abramoff, M., Niemeijer, M., et al. (2004). Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4), 501–509.



- Sukesh, A. V. and Sivaswamy, J. (2018). Shared encoder based denoising of optical coherence tomography images. In *ICVGIP 2018: Proceedings of the 11th Indian Conference on Computer Vision, Graphics and Image Processing*, ICVGIP 2018, 1–8, New York, NY, USA. Association for Computing Machinery.
- Sunija, A., Saikat, K., Gayathri, S., et al. (2021). Octnet: A lightweight cnn for retinal disease classification from optical coherence tomography images. *Computer Methods and Programs in Biomedicine*, 200, 105877.
- Szegedy, C., Vanhoucke, V., Ioffe, S., et al. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826.
- Tan, M. and Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *CoRR*, abs/1905.11946.
- Tao, Z., Zhang, W., Yao, M., et al. (2021). A joint model for macular edema analysis in optical coherence tomography images based on image enhancement and segmentation. *BioMed Research International*, 2021, 1–9.
- Teng, P.-y. (2013). Caserel - an open source software for computer-aided segmentation of retinal layers in optical coherence tomography images. *Zenodo, Public repository*, 1, 1. <https://doi.org/10.5281/zenodo.17893>.
- Tikhonov, A. N. and Arsenin, V. Y. (1977). Solutions of ill-posed problems. *New York*, 1(30), 487.
- Timischl, F. (2015). The contrast-to-noise ratio for image quality evaluation in scanning electron microscopy. *Scanning*, 37(1), 54–62.
- Ting, D. S. W., Pasquale, L. R., Peng, L., et al. (2019). Artificial intelligence and deep learning in ophthalmology. *British Journal of Ophthalmology*, 103(2), 167–175.
- Trafton, A. (2014). In the blink of an eye. *MIT News on Campus and around the world*. <https://news.mit.edu/2014/in-the-blink-of-an-eye-0116> (Accessed on: October 3 2021).

- Ulyanov, D., Vedaldi, A., and Lempitsky, V. (2020). Deep image prior. *International Journal of Computer Vision*, 128(7), 1867–1888.
- Vaghefi, E., Hill, S., Kersten, H. M., and Squirrell, D. (2020). Multimodal retinal image analysis via deep learning for the diagnosis of intermediate dry age-related macular degeneration: A feasibility study. *Journal of Ophthalmology*, 2020, 1–7.
- Van Hulle, M., Sladojevic, S., Arsenovic, M., et al. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Computational Intelligence and Neuroscience*, 1(1), 1–12.
- Vaserstein, L. N. (1969). Markov processes over denumerable products of spaces, describing large systems of automata. *Problemy Peredachi Informatsii*, 5(3), 64–72.
- Vese, L. and Guyader, C. (2015). Variational image restoration models. In *Variational Methods in Image Processing*, Chapman & Hall/CRC Mathematical and Computational Imaging Sciences Series, chapter 3, 51–80. CRC Press, Boca Raton, 1 edition. <https://books.google.co.in/books?id=nIvwCgAAQBAJ>.
- Vikram, B. (2020). Clean data is the foundation of effective machine learning. *The Newstack*. <https://thenewstack.io/clean-data-is-the-foundation-of-effective-machine-learning/> (Accessed on 18-05-2021).
- Wang, D. and Wang, L. (2019). On oct image classification via deep learning. *IEEE Photonics Journal*, 11(5), 1–14.
- Wang, J., Wang, Z., Li, F., et al. (2019a). Joint retina segmentation and classification for early glaucoma diagnosis. *Biomedical Optics Express*, 10(5), 2639.
- Wang, J., Yang, L., Huo, Z., et al. (2020a). Multi-label classification of fundus images with EfficientNet. *IEEE Access*, 8, 212499–212508.
- Wang, W., Xu, Z., Yu, W., et al. (2019b). Two-stream CNN with loose pair training for multi-modal AMD categorization. In *Lecture Notes in Computer Science*, 1 of 1, 156–164. Springer International Publishing.

- Wang, W., Zhang, C., and Ng, M. K. (2020b). Variational model for simultaneously image denoising and contrast enhancement. *Opt. Express*, 28(13), 18751–18777.
- Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4), 600–612.
- Website (2017). National programme for control of blindness and visual impairment NPCB&VI. Directorate General of Health Services, Government of India. [https://dghs.gov.in/content/1354\\_3\\_NationalProgrammeforControlofBlindnessVisual.aspx](https://dghs.gov.in/content/1354_3_NationalProgrammeforControlofBlindnessVisual.aspx) (Accessed: 08-02-2020).
- Weickert, J. (1998). Image smoothing and restoration by pdes. In Hlavac, V. and Sara, R., editors, *Anisotropic diffusion in image processing*, 1, chapter 1. ECMI Series, Teubner Stuttgart, Denmark.
- Weldon, T. P., Higgins, W. E., and Dunn, D. F. (1996). Efficient gabor filter design for texture segmentation. *Pattern Recognition*, 29(12), 2005–2015.
- Xu, M., Tang, C., Hao, F., et al. (2020). Texture preservation and speckle reduction in poor optical coherence tomography using the convolutional neural network. *Medical Image Analysis*, 64, 101727.
- Xuehua, W., Xiangcong, X., Yaguang, Z., and Dingan, H. (2021). A new method with seu-net model for automatic segmentation of retinal layers in optical coherence tomography images. In *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, 260–263. IEEE.
- Yagola, A. G. (2010). Ill-posed problems and methods for their numerical solution. In Yanfei, W., Changchun, Y., and Anatoly, G. Y., editors, *Optimization and Regularization for Computational Inverse Problems and Applications*, 17–34. Springer, Berlin, Heidelberg, 1 edition.

- Yan, Z., Yang, X., and Cheng, K.-T. (2018). A three-stage deep learning model for accurate retinal vessel segmentation. *IEEE journal of Biomedical and Health Informatics*, 23(4), 1427–1436.
- Yanagihara, R. T., Lee, C. S., Ting, D. S. W., and Lee, A. Y. (2020). Methodological Challenges of Deep Learning in Optical Coherence Tomography for Retinal Diseases: A Review. *Translational Vision Science & Technology*, 9(2), 11–11.
- Yi, X., Walia, E., and Babyn, P. (2019). Generative adversarial network in medical imaging: A review. *Medical image analysis*, 58(101552), 1–20.
- Yoo, T. K., Choi, J. Y., Seo, J. G., et al. (2018). The possibility of the combination of OCT and fundus images for improving the diagnostic accuracy of deep learning for age-related macular degeneration: a preliminary experiment. *Medical & Biological Engineering & Computing*, 57(3), 677–687.
- Yu, H., Ding, M., Zhang, X., and Wu, J. (2018). PCANet based nonlocal means method for speckle noise removal in ultrasound images. *PLOS ONE*, 13(10), e0205390.
- Yung, K. M., Lee, S. L., and Schmitt, J. M. (1999). Phase-domain processing of optical coherence tomography images. *Journal of Biomedical Optics*, 4(1), 125 – 136.
- Zahra, A. and Hossein, R. (2017). Optical coherence tomography image denoising using Gaussianization transform. *Journal of Biomedical Optics*, 22(8), 1–12.
- Zhang, J., Dashtbozorg, B., Bekkers, E., et al. (2016). Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores. *IEEE Transactions on Medical Imaging*, 35(12), 2631–2644.
- Zhao, D., Guoyan, L., Guangming, N., et al. (2020). Optical coherence tomography image denoising using a generative adversarial network with speckle modulation. *Journal of Biophotonics*, 13(4), e201960135.
- Zhao, Z., Xiong, B., Wang, L., et al. (2021). Retinexdip: A unified deep framework for low-light image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*, 1(1), 1–14.

Zosso, D., Tran, G., and Osher, S. J. (2015). Non-local retinex—a unifying framework and beyond. *SIAM Journal on Imaging Sciences*, 8(2), 787–826.



# LIST OF PUBLICATIONS

## Journal Publications

- [1] Smitha, A., & Jidesh, P. (2021). A nonlocal deep image prior model to restore optical coherence tomographic images from gamma distributed speckle noise. *Journal of Modern Optics*, Taylor and Francis, Informa UK Limited, 68(18), 1002–1017. <https://doi.org/10.1080/09500340.2021.1968052>, (SCI and Scopus indexed, IF - 1.46, Q2).
- [2] Smitha, A., Febin, I. P., & Jidesh, P. (2022). A retinex based non-local total generalized variation framework for OCT image restoration. *Biomedical Signal Processing and Control*, Elsevier BV, 71(B), 103234, 1-16, ISSN 1746-8094. <https://doi.org/10.1016/j.bspc.2021.103234>, (SCIE, Scopus, and web of Science, IF- 3.88, Q2).
- [3] Smitha, A., Jidesh, P. Detection of retinal disorders from OCT images using generative adversarial networks. *Multimedia Tools Applications* (2022), Springer Science and Business Media LLC, <https://doi.org/10.1007/s11042-022-12475-1>, (SCIE, Scopus, IF- 2.757, Q1).
- [4] Smitha, A., & Jidesh, P. (2022). Classification of Multiple Retinal Disorders from enhanced Fundus Images Using Semi-supervised GAN. *Springer Nature Computer Science*, Springer Science and Business Media LLC, 3(1), 1–12, <https://doi.org/10.1007/s42979-021-00945-6>.

## Conference Publications

- [1] Smitha, A., Jidesh, P., & Febin, I. P. (2020). Retinal vessel classification using the non-local Retinex method. *Lecture Notes in Computer Science*. Springer International Publishing, 163–174, International Conference on Intelligent Human Computer Interaction-2019, IIIT Allahabad. [http://doi.org/10.1007/978-3-030-44689-5\\_15](http://doi.org/10.1007/978-3-030-44689-5_15).
  
- [2] Smitha, A., & Jidesh, P. (2021). A Semi-supervised Generative Adversarial Network for retinal Analysis from Fundus Images. *Communications in Computer and Information Science*. Springer, 351–362, International Conference on Computer Vision and Image Processing-2020, IIIT Allahabad, [http://doi.org/10.1007/978-981-16-1086-8\\_31](http://doi.org/10.1007/978-981-16-1086-8_31).
  
- [3] Smitha, A., Jidesh, P., Balaji, J. J., & Lakshminarayanan, V. (2021). Development and validation of a novel automated method for quantification of choroidal thickness in age-related macular degeneration. In *Applications of digital image processing XLIV*, 11842. International Society for Optics and Photonics., SPIE-2021, San Diego, USA. <https://doi.org/10.1117%2F12.2594557>



## BIO-DATA

- Name** : Smitha A.
- Email Id** : smitha.ict@gmail.com
- Contact No** : +91-9480242428
- Date of Birth** : 13<sup>th</sup> July , 1989
- Permanent address** : D/o. N. Anil Kumar,  
#1-530, Shree Guru Krupa Nilaya,  
Hirebandady Post and village, Palethady, Uppinangady  
Dakshina Kannada,Karnataka, India. Pin code - 574241.
- Qualifications** : 1. M.tech (Network Engineering), Manipal Institute of  
Technology, Manipal. (2011-2013).  
2. B.E. (Electronics and Communication), St. Joseph Engi-  
neering College, Mangalore. (2007-2011).
- Academic Experience** : Assistant Professor (2013 to Present), Department of Infor-  
mation and Communication Technology, Manipal Institute  
of Technology, Manipal, Karnataka. (On Study leave from  
Dec 2018).
- Publications** : 4 International Journals (2 SCI), 7 International Confer-  
ence.