AUTOMATIC DETECTION OF RED LESIONS IN RETINAL FUNDUS IMAGES

Goutam Kumar Ghorai

AUTOMATIC DETECTION OF RED LESIONS IN RETINAL FUNDUS IMAGES

Thesis submitted by

Goutam Kumar Ghorai

Doctor of Philosophy (Engineering)

Electrical Engineering Department
Faculty Council of Engineering & Technology
Jadavpur University
Kolkata, India

2024

Index No.52/16/E

1. Title of the Thesis:

AUTOMATIC DETECTION OF RED LESIONS IN RETINAL FUNDUS IMAGES.

2. Name, Designation, and Institution of the Supervisors:

(a) Dr. Gautam Sarkar
Professor,
Department of Electrical Engineering
Jadavpur University, Kolkata-700032, India

(b) Dr. Ashis Kumar Dhara
Assistant Professor,
Department of Electrical Engineering
National Institute of Technology, Durgapur-713209

3. List of Publications:

Journal Published

- **Goutam kumar Ghorai**, Swagata Kundu, Gautam Sarkar, Ashis kumar Dhara, "Multi-Scale Feature Pyramid for Red Lesions in Fundus Images". International Journal of Recent Technology and Engineering (IJRTE). Volume-12 Issue-4, November 2023.
- Swagata Kundu, Vikrant Karale, **Goutam Ghorai**, Gautam Sarkar, Sambuddha Ghosh, Ashis Kumar Dhara, "Nested U Net for Segmentation of Red Lesions in Retinal Fundus Images and Sub Image Classification for removal of False Positives". Journal of Digital Imaging 35, no. 5(2022): 1111-1119.
- Sandip Sadhukhan, Arpita Sarkar, Debprasad Sinha, **Goutam Kumar Ghorai**, Gautam Sarkar, Ashis Kumar Dhara, "Attention Based Fully Convolutional Neural Network for Simultaneous Detection and Segmentation of Optic Disc in Retinal Fundus Images". International Journal of Medical and Health Science Vol. 14No: 8, pp. 200-204, 2020.
- Sandip Sadhukhan, **Goutam Kumar Ghorai**, Debprasad Sinha, Souvik Maiti, Gautam Sarkar and Ashis Kumar Dhara "Segmentation of optic disc in retinal images using fully convolutional network" Journal of Current Indian Eye Research Vol. 6, Issue 2, December 2019, pp. 40-47.

Conference Communicated

- Sandip Sadhukhan, **Goutam Kumar Ghorai**, Souvik Maiti, Gautam Sarkar, Ashis Kumar Dhara, "Optic disc localization in retinal fundus images using faster R-CNN". Proceedings of Fifth International Conference on Emarging Applications of Information Technology (EAIT 2018), January 12-13, Kolkata,India.
- Sandip Sadhukhan, **Goutam Kumar Ghorai**, Souvik Maiti, Vikrant Anilrao Karale, Gautam Sarkar, and Ashis Kumar Dhara, "Optic Disc segmentation in retinal fundus images using fully convolutional network and removal of false-positives based on shape features" Proceedings of

4th International Workshop, DLMIA 2018 and 8th International Workshop, ML-CDS 2018, September 20, 2018. Granada, Spain.

- Sandip Sadhukhan, **Goutam Kumar Ghorai**, Souvik Maiti, Debprasad Sinha, Gautam Sarkar, and Ashis Kumar Dhara, "Fully Convolutional Network for Segmentation of Optic Disc in Retinal Fundus Images" Proceeding of IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) April 8-11, 2019.Venice Italy.
- Debasis Maji, Souvik Maiti, Moyenak Biswas, **Goutam Kumar Ghorai**, Sandip Sadhukhan, Debprasad Sinha, Ashis Kumar Dhara, Gautam Sarkar, "Automatic Patch Based Tortuosity Retinal Vessel Classification using VGG16 Network". IEI Impact in Changing Energy Mix in the Power Sector, ISBN:978-81-942561-2-0,2019, Kolkata, India.

4. List of Patents: Nil

Statement of Originality

I, Goutam Kumar Ghorai, registered on 01.12 2016, do hereby declare that this thesis entitled "AUTOMATIC DETECTION OF RED LESIONS IN RETINAL FUNDUS IMAGES", contains literature survey and original research work done by the undersigned candidate as part of Doctoral studies. All information in this thesis have been obtained and presented in accordance with existing academic rules and ethical conduct. I declare that, as required by these rules and conduct, I have fully cited and referred all materials and results that are not original to this work. I also declare that I have checked this thesis as per the "Policy on Anti Plagiarism, Jadavpur University, 2019", and the level of similarity as checked by iThenticate software is 4%.

Signature of Candidate Goutam Kumar Ghorai

Date: 14/02/2024

Certified by Supervisors: (Signature with date and seal)

14/02/2024

Dr. Gautam Sarkar
(Professor)
Electrical Engineering Department
Jadavpur University
Kolkata – 700032, India

Prof. Gautam Sarkar
Electrical Engineering Department
Jadavpur University
Kolkata, INDIA

Ashis 14/02/24

Dr. Ashis Kumar Dhara
(Assistant Professor)
Electrical Engineering Department
National Institute of Technology
Durgapur – 713209, India

Dr. Ashis Kumar Dhara
Assistant Professor
Electrical Engineering Department
National Institute of Technology
DURGAPUR - 713209

CERTIFICATE OF APPROVAL

This is to certify that the thesis entitled "Automatic detection of red lesions in retinal fundus images" submitted by Sri. Goutam Kumar Ghorai, who got his name registered on 1st December 2016 for the award of Ph.D. (Engineering) degree of Jadavpur University, is absolutely based upon his own work under the supervision of Dr. Gautam Sarkar and Dr. Ashis Kumar Dhara and that neither his thesis nor any part of the thesis has been submitted for any degree/diploma or any other academic award anywhere before.

14/02/2024

Dr. Gautam Sarkar
Professor
Electrical Engineering Department
Jadavpur University
Kolkata– 700032, India

Prof. Gautam Sarkar
Electrical Engineering Department
Jadavpur University
Kolkata, INDIA

14/02/24

Dr. Ashis Kumar Dhara
Assistant Professor
Electrical Engineering Department
National Institute of Technology
DURGAPUR - 713209

Dr.Ashis Kumar Dhara
Assistant Professor
Electrical Engineering Department
National Institute of Technology
Durgapur – 713209, India

Dedicated to

My beloved parents and other family members

Acknowledgment

I would like to heartily thank my supervisors Prof. Gautam Sarkar and Dr. Ashis Kumar Dhara for their guidance and support throughout my research work. Their direction and motivation ascertained that I maintain high-standards at my research. I am really grateful to my mentors. I must also acknowledge the technical and administrative support provided by Jadavpur University during the research. I am obliged to Prof. Biswanath Roy, HoD, Department of Electrical Engineering, Jadavpur University, for his valuable guidance during the work. I am also thankful to Prof. Sugata Munshi, Prof. Palash Kumar Kundu, Prof. Amitava Chatterjee, Prof. Arabinda Das and Dr. Debangshu Dey for their constant encouragement during the research.

I would like to thank Dr. S. Ghosh, Professor, Calcutta National Medical College, Kolkata for his constant support. I am thankful to Dr. Biswapati Jana of Computer Science Department, Vidyasagar University for his valuable guidance. I am thankful to Mr. Jayhari Bera, Smt. Benubala Maity, Smt. Durga Ghorai, Prof. Gopal Chandra Bera and Prof. Montu Kumar Das for their support.

I had an enjoyable time working in Electrical Engineering Department. fellow researchers Debprasad Das, Souvik Maiti, and Debasis Maji who have shared their knowledge and experiences contributing to the development of my ideas.

Lastly, heartfelt thanks to my friends and family for their constant support, understanding

Thanking You,

Goutam Kumar Ghorai

Goutam Kumar Ghorai

Contents

Title Page i

Declaration iii

Acknowledgement ix

Contents xi

List of Abbreviation xiv

List of Figures xv

List of Tables xvii

Abstract xix

1 Introduction 1

1.1 Diabetic Retinopathy 1

1.2 Prevalence of diabetes and DR 3

1.3 Pathogenesis of DR 5

1.4 Screening of DR 7

1.5 Scope of the work 8

1.6 Outline of the Thesis 8

1.6.1 **Chapter 1: Introduction** 9

1.6.2 **Chapter 2: Literature review** 9

1.6.3 **Chapter 3: Multi-Scale Feature Pyramid for Detection of Red Lesions** 9

1.6.4 **Chapter 4: Segmentation of Red Lesions using Nested U-Net** 10

1.6.5 **Chapter 5: Conclusion** 10

2 Literature Review **11**
2.1 Object Detection Methods 12
2.1.1 Classical Object Detection Methods 12
2.1.2 Deep Learning Based Object Detection Methods 13
2.2 Segmentation Methods 18

3 Multi-Scale Feature Pyramid for Detection of Red Lesions **21**
3.1 Introduction 22
3.2 State-of-the-art-methods for red lesion detection 24
3.3 Multi-Scale Feature Pyramid for Red Lesion Detection 25
3.3.1 Preprocessing 25
3.3.2 Network Architecture 26
3.4 Results and Discussions 30
3.5 Conclusion 32

4 Segmentation of Red Lesions using Nested U-Net **33**
4.1 Introduction 34
4.2 Methods for Segmentation of Red Lesions 35
4.2.1 Preprocessing 35
4.2.2 Network architecture for red lesions segmentation 36
4.3 Description of Database 39
4.4 Results and Discussions 40
4.4.1 Metric 40
4.5 Conclusion 44

5 Conclusions and Future Directions **47**
5.1 Summary of Studies 48
5.2 Contributions of the Thesis 49
5.3 Future Scope 49

Appendices **51**

A Machine Learning Fundamentals **53**
A.1 Machine Learning Fundamentals 54
A.1.1 Introduction 54
A.1.2 Learning algorithms 56
A.1.3 Understanding how deep learning works 58

References **61**

Publications **65**

Author's Biography **67**

LIST OF ABBREVIATIONS

Abbreviation	Description
AI	Artificial Intelligence
AMD	Age-related Macular Degeneration
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Network
DPM	Deformable Part Based Model
DR	Diabetic Retinopathy
FCN	Fully Convolutional Network
FPN	Feature Pyramid Network
FROC	Free Response Receiver Operating Characteristic
HOG	Histogram of Oriented Gradients
IDF	International Diabetes Federation
IRMA	Intraretinal Micro-vascular Abnormalities
MESSIDOR	Methods to Evaluate Segmentation and Indexing Techniques in the Field of Retinal Ophthalmology
NPDR	Non-Proliferative Diabetic Retinopathy
OD	Optic Disc
PDR	Proliferative Diabetic Retinopathy
RCNN	Region-based Convolutional Neural Networks
ReLU	Rectified Linear Unit
RPN	Region proposal Network
SSD	Single Stage Detection Network
SN	Sankar Nethralaya
DIARETDB0	Standard Diabetic Retinopathy Database Calibration Level 0
DIARETDB1	Standard Diabetic Retinopathy Database Calibration Level 1
SGD	Stochastic Gradient Descent

List of Figures

1.1 Stages of DR: (a) Mild NPDR, (b) Moderate NPDR, (c) Severe NPDR, and (d) PDR. 2

3.1 Hemorrhages are marked in blue, vessel crossover are marked in yellow and microaneurysms are marked in green. 22
3.2 (a) and (b) sample relational fundus images from DIARECTDB1 with non-uniform illumination 23
3.3 Preprocessing step: (a) fundus image and (b) fundus image after preprocessing 25
3.4 Multi-scale feature pyramid with classification and regression head for red lesion detection 26
3.5 FROC of multi-scale feature pyramid network for red lesion detection . 27
3.6 Attention augmented feature pyramid network with classification and regression head for small red lesion detection 28
3.7 FROC of attention augmented feature pyramid network 29
3.8 Visual results for a few test images of DIARECTDB1 database, where true positives, false positives, and false negatives are marked with green, red and blue boxes. 31

4.1 Block diagram of the proposed methodology 34
4.2 (a) Normal image and (b) result after preprocessing. 35
4.3 Network architecture of proposed segmentation method. 37
4.4 The Residual Block 38
4.5 predicted lesions and overlap with ground truth. 41

4.6 ground truths: (a), (d) and (g), segmentation results using proposed network: (b), (e) and (h), result after false-positives removal: (c), (f) and (i). Red corresponds to false negative candidates, green represents the false positive, and true positive candidates are true positive candidates are in blue. 42
4.7 Results of U-Net: (a), (d) and (g), results of attention U-Net (b),(e) and (h), and results of proposed method presented in (c), (f) and (i). Red corresponds to false negative candidates, green represents the false positive, and true positive candidates are true positive candidates are in blue. 43

A.1 Pictorial presentation of Artificial Intelligence, machine learning and deep learning 54
A.2 Pictorial representation of Machine learning methodology. 55
A.3 Pictorial presentation of Supervised and Unsupervised algorithm 58
A.4 A deep neural network for digit classification (Source: Deep Learning with Python by François Chollet) 59
A.5 A deep neural working mechanism 60

List of Tables

4.1 Performance Evaluation for Red lesion Segmentation 44

Abstract

DR is one of the important causes of permanent blindness of the eyes [1] caused by long term diabetes. One-third number of diabetic patients come under DR and one-tenth of them face the severity of permanent vision loss. Early detection and timely management could help to preserve vision in individuals with DR. An automated detection system with computerized analysis of retinal fundus images could help diagnosis of DR and follow-up of DR. The first clinically identifiable sign of DR is microaneurysm and its detection at an early stage is very important.

Deep learning techniques achieved better performance in object detection and segmentation. A novel multi-scale feature pyramid network is developed for automatic detection of red lesions from retinal fundus images. The multi-scale features are extracted using the feature pyramid network. Attention augmented feature pyramid network is developed to detect microaneurysms. The network is end-to-end trainable in image level with several scales and works with acceptable performance.

The proposed segmentation architecture uses informative training images and can accurately segment red lesions present in fundus images with a minimum number of false-negative candidates. The false-posive cadidate also reduces to a large extent in this method.The technique used in this work can successfully segment the mico-aneurysms with a small number of false-positive candidates. With the help of this technique, we can detect DR in the early stage and prevent permanent blindness due to DR.In the case of non-proliferative DR, this computer-based diagnosis technique becomes helpful for an opthalmologist making clinical decisions.

Keywords: Deep learning, Diabetic reinopathy, Referable diabetic reinopathy, Retinal fundus image analysis Red lesions, Screening of diabetic reinopathy

CHAPTER 1

Introduction

1.1 Diabetic Retinopathy

Diabetes is a metabolic disorder due to insufficient or less activity of pancreas secretion's insulin hormone in human physique. In the long term, uncontrolled diabetes affects the eyes, kidneys, heart, and even the nervous system of the human body. Diabetes creates an eye disease called Diabetic Retinopathy (DR). In the future, it will cause permanent blindness if not treated in the early stage of DR. Poor control of glucose levels in the blood enhances the DR, then the well-maintained blood glucose levels. DR is one of the important causes of permanent blindness of the eyes [1]. One-third number of diabetic patients come under DR, and one-tenth of them face the severity of permanent vision loss. Twenty to seventy-year-old people lose their vision mainly due to DR in USA. Therefore, regular, at least annual retina screening is necessary.

The retina screening is started by taking a fundus image of the retina using a mydriatic camera with the help of paramedical personnel. Then, this fundus image is analyzed by the eye's specialist doctors to detect DR. In India, DR is presently taken

care of in many ways, such as through annual eye checkup camps in rural and urban areas. Sometimes diabetic patients come to the eye clinics for other ocular problems, sometimes using telemedicine infrastructure, fundus images are collected for analysis, and DR is detected by the physician.

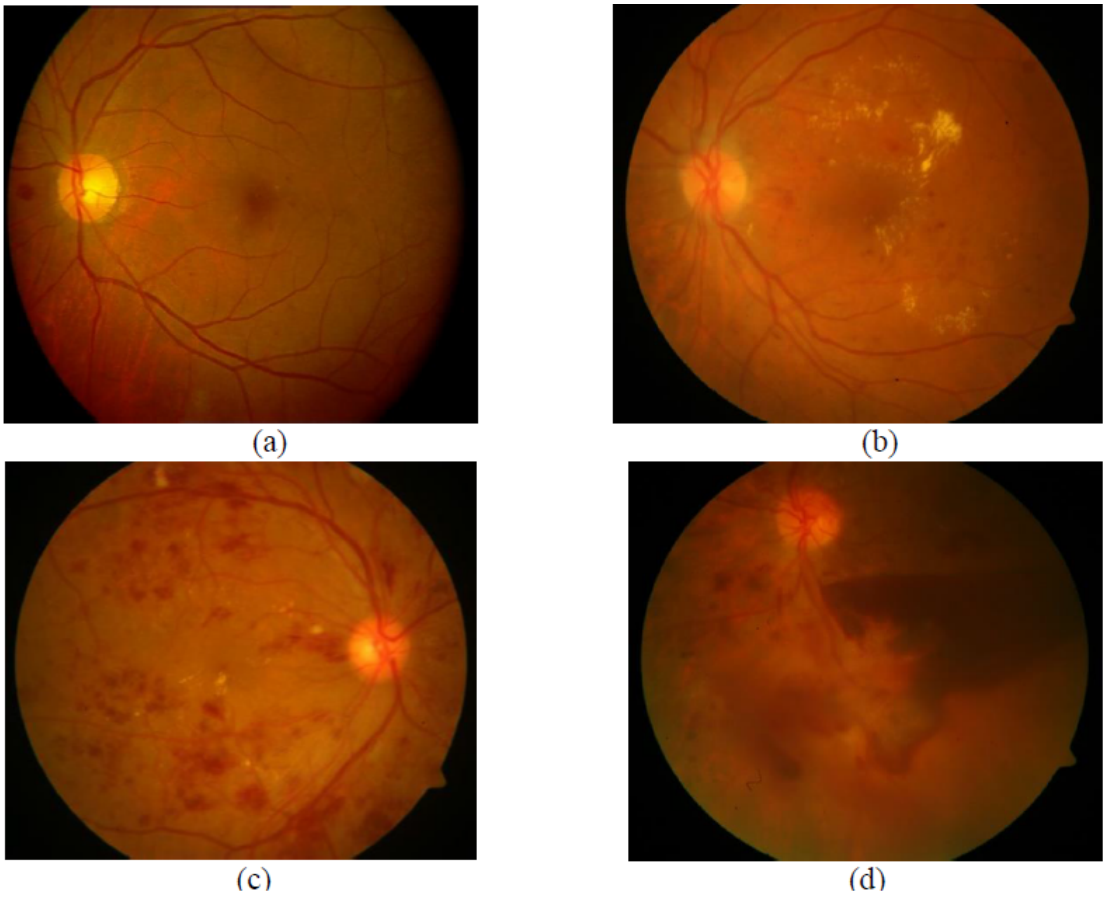

Figure 1.1: Stages of DR: (a) Mild NPDR, (b) Moderate NPDR, (c) Severe NPDR, and (d) PDR.

The major population-based limitations that affect screening for DR are lack of awareness and poor access to specialised clinic. In rural areas, throughout the year, young and elderly villagers are engaged in agriculture and harvesting work. In India, most of the patients come for check-up to the ophthalmologist when they already have established severe non-proliferative DR (NPDR) or proliferative DR (PDR) as shown in

Fig.1.1. Early detection and timely management has help preserve vision in individuals with DR. Therefore, management becomes complex with the day-to-day increased number of diabetic patients and the insufficient number of trained clinical specialists. An automated detection system with computerized analysis of retinal fundus images could reduce the burden of eye screening and diagnosis of DR and follow-up of DR.

1.2 Prevalence of diabetes and DR

Worldwide fast urbanization, a sedentary lifestyle, and processed food consumption increase obesity and insulin resistance, which causes type-2 diabetes. Diabetes is an important issue worldwide among all age groups of the population. The International Diabetes Federation (IDF) reported 537 million diabetes patients globally in 2021 [2]. Diabetes patients increased from 463 million to 537 million within two years. It is expected that by 2030, there will be 643 million, and by 2045, there will be 785 million. In 2010, diabetes patients were 12.1 million out of the 379 million adult population in the African sub-continent, with a diabetes prevalence of 3.8%. It is expected that within 2030, the adult population become 653 million, and diabetes patients will be 23.9 million, with a diabetes prevalence of 4.7%. In the Eastern Mediterranean and the Middle East, 26.6 million diabetes patients out of 344 million had a diabetes prevalence of 9.3% in 2010. It is expected that 51.7 million of the adult population will be diabetes patients out of 533 million of the population with a diabetes prevalence of 10.8% by 2030.

In Europe, 55.4 million diabetes patients out of 646 million, with a diabetes prevalence of 6.9% in 2010. It is expected that 66.5 million of the adult population will be diabetes patients out of 695 million of the population with a diabetes prevalence of 8% by 2030. In North America, 37.4 million diabetes patients out of 320 million, with a diabetes prevalence of 10.2% in 2010. It is expected that 53.2 million of the adult population will be diabetes patients out of 390 million of the population with a diabetes prevalence of 12.1% by 2030. In South and Central America, 18 million diabetes pa-

tients out of 287 million, with a diabetes prevalence of 6.6% in 2010. It is expected that 29.6 million of the adult population will be diabetes patients out of 382 million of the population with a diabetes prevalence of 7.8 by 2030. In South Asia,58.7 million diabetes patients out of 838 million, with a diabetes prevalence of 7.6% in 2010. It is expected that 101 million of the adult population will be diabetes patients out of 1200 million of the population with a diabetes prevalence of 9.1% by 2030.

In India, the Indian Council of Medical Research-India DIABetes(ICMR-INDIAB) selected the states Tamilnadu, Maharastra, Jharkhand, and one union territory Chandigarh to understand the prevalence of diabetics [3]. The phase-1 survey shows an overall weighted prevalence of diabetes in Tamilnadu at 10.4%, Maharastra at 5.3%, Jharkhand at 5.3%, and Chandigharh at 13.6%. The overall weighted prediabetes patients were 8.3% of overall weighted prediabetes patients were found in Tamilnadu, In the case of Maharastra, 12.8%, 8.1%, and 14.6% in the case of Jharkhand and Chandigarh, respectively. This result turns in the case of Tamilnadu out of 67.4 million population, 4.8 million diabetic and 3.9 million prediabetic patients in the case of Maharashtra, 6.0 million diabetic and 9.2 million prediabetic patients out of 112.7 million population. Out of the 31.4 million population in Jharkhand, 0.96 million and 1.5 million people are under diabetes and pre-diabetes, respectively. Chandigarh has 0.12 million and 0.13 million diabetic and pre-diabetic patients, respectively. Therefore, 11.8 million diabetic and 14.6 million prediabetic patients out of 211.6 million of the population are living within these three states. The extrapolated result shows 62.2 million diabetic and 77.2 million pre-diabetic patients in India. The survey shows that prevalence in the age group 25-34 years is more prone than the age group after age 65.

The National Urban Diatebes study reported that in six metropolitan cities, the overall the age-standardized prevalence of diabetes is 12.1% overall, whereas the incidence of impaired glucose tolerance is 14%. According to a study on the incidence of diabetes in India, the prevalence is 2.7 and 5.9 in rural and small-town areas, respec-

tively. The urban area is more prone to prevalence of diabetes then the rural area. This difference is more significant in Jharkhand state, where the socio-economics difference between rural and urbanism. In the future, increased number of middle-class families, urbanization, and and growth of population will lead to diabetes prevalence in India. The estimated DR prevalence is 18% [4] of diabetic patients in India.

1.3 Pathogenesis of DR

DR is a microangiopathy in nature and affects mainly retinal precapillary arterioles and capillaries and lead to microvascular occlusion and spillage of blood. Pathogenesis under microvascular occlusion is capillary changes that occur due to the swelling of the base membrane of capillary. The inner blood-retinal barrier is formed by the close association of endothelial cells. Neo-vascularization occurs due growth factor on the retina and optic nerve head. Endothelial cells and pericytes are the cellular elements of retinal capillaries. The inside blood-retinal bloked is made by the close association of endothelial cells. The pericytes are wrapped around the capillaries and maintain the structural integrity of the vessel wall. Normal people have one pericyte for every endothelial cell but diabetics have fewer pericytes. The vessel wall's strength is preserved by the pericytes, which form a covering inside the capillaries. Local capillaries either bleed or thrombose resulting in the formation of microaneurysms. Because of the distension of the capillary wall caused by this reduction in pericytes, the blood-retinal barrier breaks down, allowing the plasma components to enter the retina. Increased vascular permeability causes diffuse or localized intraretinal haemorrhage. The hard exudates are formed by the chronic localized retinal odema at the junction of the healthy and oedematous retina. The Lipid-filled macrophages and lipoproteins are the primary constitutions of exudates. Clinically, diabetic retinopathy can be identified by the following lesions in the fundus image of the eye. The following lesions in the fundus image of the eye indicate clinically diabetic retinopathy.

- Microaneurysms: The first clinically identifiable sign of DR is microaneurysm,

which looks like a small round dot.

- Hard exudates: Hard exudates look like yellow waxy with a relatively distinct margin, clumps, and ring shape.

- Cotton-wool spots: Cotton-wool spots: Cotton-wool spots are formed due to the capillary obstruction at the retinal nerve fiber area and white colour of these lesions is caused by ischemia.

- Edema: Edima is distinguished by retinal thickness, obstructing the choroid and retinal pigment epithelium underneath. Initially, it is seen in the middle of the outer plexiform and inner nuclear region; then, it plays a role in forming oedematous in the retina by increasing the entire thickness of the retina involving the inner plexiform and nerve fiber layer.

- Haemorrhage: Haemorrhage appears generally two type: dot bolt type and flame shaped type. Dot bolt types come from the capillaries' venous end and are placed in the middle region of the retina. Flame-shaped bleeding comes from more specific pre-capillary arterioles that follow the path of the retinal nerve fiber region.

- Intraretinal microvascular abnormalities (IRMA): IRMAs are mostly found near the capillary bed. Shortage of profuse discharge on fluorescein angiography and the failure of cross-major retinal blood vessels are the primary characteristics of IRMA.

- Neovascularization: It is the most important signature of proliferative diabetes. Neovascularization is observed on or surrounding the optic nerve. It is observed that before new vessels form at the disc, one-fourth of the inside of the retina becomes non-perfused. Endothelial proliferations is the early stages of new vasculature, which most commonly originate from veins. The area covered by new vessels relative to their diameter can be used to determine the severity of neovascularization.

1.4 Screening of DR

Regular screening is an effective means to understand the eye health of all diabetic patients. Patients with diabetes who are over 12 or who are about to start puberty should be screened. An annual review is required for patients when fundus is normal and mild background DR with small haemorrhage or small hard exudates exist. Review in regular basis is required for history of DR with large exudate within the major temporal arcade but not threatening the macula. An urgent review is required for patients for proliferative DR and vitreous haemorrhage. AI and deep neural networks can be very helpful in the detection of diabetic retinopathy by identifying several risk factors of DR in fundus images [5]. Diabetic retinopathy, glaucoma, and age-related macular degeneration(AMD) can be successfully diagnosed by using Deep learning algorithms [6]. Artificial intelligence-assisted algorithms have been employed to bring about correlation of datasets in retinopathy, cataract, corneal disease and oculoplasty. Automated retinal image analysis system developed by different groups uses a number of programs for algorithms to correlate the images to desired parameters [7].

Different Organizations such as iGradingM (Medalytix Group Ltd. UK), RetMarker (RetMarker SA, Portugal), and Eye Ar (Eyenuk, California) have been involved in the creation of AI algorithms and the correlation of fundus images of various stages DR [8]. Singapore Diabetic Retinopathy screening program (SIDRP) employed deep learning algorithms to screen multi-ethnic cohorts of patients [9]. They validated this algorithm using ten additional multi-ethnic datasets from different countries. These algorithms detected reliably different grades of DR; it was also used for vision threatening conditions like glaucoma and AMD. Previous analysis of deep learning to grade diabetic retinopathy severity used about 25,000 gradable images and compared the human graders versus deep learning methods [10]. Gulshan et al. [11] generated a deep learning algorithm to detect DR in Indian population and used about 3049 patients from SN, Chennai and Aravind eye hospital, Madurai. The above study was done with retina images

in the people with diabetes in India. The sensitivity of the algorithm detecting DR ranged from 88.9% to 95.2% in different cohorts of patients. The above two papers have shown that automated algorithm for DR detection generalises the prediction levels and demonstrated the feasibility of using automated DR grading real time.

1.5 Scope of the work

Currently, there is a need to develop an indigenous imaging system that can be set up in primary healthcare centres and large hospitals to enable screening of the general public, even individuals who are not aware of the impact of DR in long duration. The resulting technology can be gainfully utilized as clinical decision support system for cardiovascular disease risk assessment. The technology must be easy to use, reasonably priced, and require little upkeep in such a way that it should be used in rural areas where DR screening.

AI has clear applications in ophthalmology, where there are large numbers of patients to be examined and require large amounts of data to process, but the results are straightforward and well-defined.A computer-based technology is necessary to assist the ophthalmologists' to more accurate detection of the signs of DR in fundus images and undrerstand the severity to initiate the treatment. Deep learning are using as a powerful tool for medical images analysis. Early detection of DR is now possible due to the rapid advancement of this new optical imaging technologies over the years. Ophthalmic imaging has developed rapidly over the past years and scope is there for early diagnosis of DR. This tool could segregate 'Normal' and 'high risk' patients by imaging-based screening and recommending the latter group for immediate treatment.

1.6 Outline of the Thesis

In this thesis, novel methods are developed for detection and segmentation of red lesions in retinal fundus images. The overview of the thesis is provided below:

1.6.1 Chapter 1: Introduction

Diabetes is a metabolic disorder due to insufficient secretion or less activity of insulin hormone in the human body. Worldwide fast urbanization, a sedentary lifestyle, and processed food consumption increase obesity and insulin resistance, which causes type-2 diabetes and is an important issue worldwide among all age groups of the population. Chapter 1 provides the worldwide prevalence of diabetes and DR and scope of research for early diagnosis of DR.

1.6.2 Chapter 2: Literature review

In recent year,computer vision and pattern recognition technique is an emerging research area of deep learning. Deep learning techniques achieved better performance in object detection and segmentation. Chapter 2 provides a comprehensive review on state-of-the-art-methods for object detection and image segmentation.

1.6.3 Chapter 3: Multi-Scale Feature Pyramid for Detection of Red Lesions

In this chapter, a novel multi-scale feature pyramid network is developed for automatic detection of red lesions i.e., both microaneurysms and hemorrhages from color fundus images. Chapter 3, comprises methodology of an accurate and fast red lesions detection. The vital contribution of this work is to develop a deep learning-based framework that can effectively manage severe class imbalance and size imbalance of red lesions present in fundus images. Feature pyramid network is in this method to extract multi-scale features from images. Strong semantic information is easily extracted by a pyramid of feature. The network is end-to-end trainable and can handle several scale variations with wide range variation of red lesions with acceptable results. The proposed method shows sensitivity of 0.6 with six number of false-positive per image on DIRETDB1 data set. This deep learning-based red lesion detection technique could be used in DR screening program.

1.6.4 Chapter 4: Segmentation of Red Lesions using Nested U-Net

Segmentation of red lesions using nested U-Net [12] false positives removal based on the sub-image classification is presented in Chapter 4. The proposed technique has sufficient potential to assist ophthalmologists in DR diagnosis. This technique uses a minimum number of images to train the network. Therefore, with this minimum information, the method can produce accurate outcomes. The suggested technique can successfully segment red lesions from the given fundus image with a few numbers of false-negative candidates.In this technique the number of false-positive candidates also decrease to a large extent. The proposed method was evaluated on DIARETDB1 data set [13] and achieves sensitivity of 88.79%, precision of 71.50%, and F1-Score of 79.21% and outperforms the competing networks.

1.6.5 Chapter 5: Conclusion

Early detection of red lesions in fundus images has immense potential in clinical practice and screening programs. Computer-aided screening of DR is very essential in handling a large number of patients who needs fundus examination. The improved performance of the methods for red lesion detection could improve the screening performance. The segmentation of red lesions could be used for quantitative image analysis. The nested U-Net based architecture is developed for the segmentation of red lesions in fundus images with improved performance. Chapter 5 concludes the work done and future scope of work.

CHAPTER 2

Literature Review

In developing countries, the ratio of people suffering from various kind of eye diseases to that of available ophthalmologists is increasing in an alarming rate. A computer based screening system would reduce the burden on the health care system of these kind of countries. Machine learning techniques are considered for several classification tasks for automated diagnosis of eye disorder. Most of the work focused on feature engineering, which is the expert's computation of explicit feature details [14].

Deep learning technique is a subset of the Machine learning techniques, consists of multiple computation layers, which follow an algorithm to capture the suitable predictive feature considering examples instead of hand-crafted features. The deep learning technique is a technique to adapt correct values [15] of the different parameters by which it can execute a given work. Initially, the values of parameters are assigned by using arbitrarily random value. Then, for every image, the prediction result evaluated by the network is compared with the approved and recognised training data set values. Little modification of the model's parameters is necessary for that image to decrease the error in the result. The model selects the proper parameter by repeating the method

till it's all trained. Recently, the convolutional neural networks [9] [16] have been used and achieved higher accuracy and precision in detecting numerous diseases, like diabetic retinopathy,hypertensive retinopathy,and other disease. The current state-of-the-art methods used in object detection and segmentation are reviewed in this chapter.

2.1 Object Detection Methods

Object detection is a technique for the identification and localization of objects within a image or a video. The classical methods for object detection are Viola Jones detectors [17], histogram of oriented gradients(HOG) [18], deformable part based model(DPM) [19].

2.1.1 Classical Object Detection Methods

Viola Jones detectors [17] selects consider only a few numbers of crucial visual features within a large data set. Adaptive Boosting (AdaBoost) [20] is a boosting algorithm used in the Viola-Jones framework to consider a very small data set of important visual features from a larger pool. The algorithm works by iteratively training weak classifiers on the dataset, assigning higher weights to wrongly classified samples in every iteration process. The final strong classifier is a weighted sum of these weak classifiers, and it focuses on features that are more informative for distinguishing between positive and negative instances. Haar-like wavelets are used in the Viola-Jones framework to define simple rectangular features that capture intensity differences in image regions. These features serve as the basis for the weak classifiers. The framework evaluates the Haar-like features at various positions and scales in the image to create a diverse set of potential rules for object detection. The Viola-Jones framework trains the AdaBoost algorithm on a cascade of classifiers. Every stage in the cascade is made up of a powerful classifier built by AdaBoost using a specific subset of Haar-like features. The cascade is designed in a way that simpler and faster-to-compute classifiers come first. If an image region passes the early stages, it is subjected to progressively more complex classifiers at later stages. The combination of AdaBoost for efficient feature selection and Haar-like

wavelets for region rejection rules allows the Viola-Jones framework to achieve highly accurate detection of objects in real-time. The cascade structure and focus on computationally efficient features make it well-suited for applications where speed is crucial, such as face detection in images and videos.

The gradient orientation procedure is used in HOG [18] to localize the most critical parts of an image. The method depends on assessing the gradient-oriented local histogram of images with a well-normalized dense grid. The gradient magnitude and angle are obtained for the current pixel by exploring the entities in their horizontal and vertical surroundings.

Objects are considered as collection of parts in DPM [21] and each part in a grammar-based model can be represented directly or represented in therms of other objects. In DPMs, objects are represented as collections of parts, and the relationships between these parts are explicitly modeled. Each part is associated with a spatial filter, and the arrangement and relationships between these parts are learned during the training phase. This approach allows the model to handle variations in object appearance, scale, and pose by explicitly modeling the deformations and articulations of object parts.

Grammar-based models often use a hierarchical structure to represent objects. The combination of gradient-based features for capturing image properties and the hierarchical representation of objects in DPMs or grammar-based models enables effective object detection and recognition in computer vision. These models also provide a natural framework for sharing information and computation between different object classes.

2.1.2 Deep Learning Based Object Detection Methods

Deep learning based object detection methods are very popular because of its powerful learning ability and handling occlusion. Deep learning-based object detection method

is categorized into one-stage approach and two-stage approach.

Single Stage Object Detection Methods

The single-stage object detection model works directly on the dense sampling location, whereas the two-stage model initially works on the region proposal layer and then the dense sampling location. The examples of single stage detector are single shot detector(SSD) [22], YOLO [23], and RetinaNet [24]. The SSD provides a speedy method to attain acceptable object detection performance in real time and combines the benefits of a multibox object identification framework for localization and feature extraction by using convolutional neural networks(CNNs). The backbone of SSD typically uses a pretrained CNN like VGG, ResNet, or MobileNet to extract feature maps from the input image. SSD utilizes multiple layers of the feature hierarchy to capture information at different scales. These feature maps are obtained from different convolutional layers in the network. By using feature maps from various layers, SSD is capable of detecting objects of different sizes. SSD uses a set of default bounding boxes (also known as anchor boxes) of different aspect ratios at each spatial location in the feature maps. These default boxes act as templates for potential object locations and help the model predict the offsets for these boxes to fit the actual object positions. SSD's ability to capture multi-scale features, make predictions at different resolutions, and efficiently process images in a single pass contributes to its effectiveness in real-time object detection tasks. The use of anchor boxes and post-processing steps further enhance the accuracy and robustness of the detection system. It is inferior to the Faster RCNN in small-scale object detection and greatly suffers from lowering the resolution of the images to a lower quality and perform not good enough than the Faster RCNN incase of small scale objects.

Object detection in YOLO is considered as a regression problem and provide the class probabilities of the detected image and finds application in agriculture, healthcare, security surveillance and self driving cars. The advantages of YOLO are detection in real time with high accuracy and speed and better generalization for new domain. YOLO V2

was created to make a better version of YOLO v1 algorithm. The main improvement was the use DARKNET-19, batch normalization, higher resolution inputs, convolution layer with anchors, dimensionality clustering and fine grained features. The improvement of YOLO v3 is use of Darknet-53 with better bounding box regression and accurate class predictions. In YOLO V3, predictions at three different scales were performed for each location in the input image. YOLO V4 has an optimal speed and accuracy of object detection for the use of spatial pyramid pool, which significantly increases the receptive field, and does not affect the network speed. YOLO V7 architecture is designed by integrating efficient layer aggregation network(EELAN) which allows the model to learn more adverse feature for better learning. The algorithm scales its architecture by concatenating the architecture of the models which is derived from YOLO v4. .

RetinaNet [24] is a popular model among the other object detection models in terms of object detection by using single-shot object detection algorithms. The architecture of RetinaNet is designed by eliminating the previous issues of other single-shot algorithm-based models; as a result, it can provide a better, more effective, and more efficient result than the others. The architecture of RetinaNet is built in such a way that the previous issues of single shot detector are somewhat balanced out to produce more effective and efficient results. The architecture is a combination of three main entities such as the ResNet model, the feature pyramid network(FPN), and the focal loss. In the Feature pyramid architecture, most of the deficiencies have been eliminate that was found in previous architecture. A semantic-rich feature of low resolution is combined with semantically weak features of high-resolution features of the image in FPN. The final output of FPN can provide both the classification and regression of objects like other object detection architecture.

Two Stage Object Detection Methods

The region-based convolutional neural network (R-CNN) [11] and the spatial pyramid pooling (SPP) [12] networks are examples of two stage detectors. Faster R-CNN [14] uses a region proposal network and is an improvement over Fast R-CNN [13]. In the first phase, the object detection model produces region proposals. Then, the object is classified and detected using a bounding box in the second stage. The first two-stage object detection model is Region proposal with Convolutional neural network (R-CNN). The R-CNN's first stage creates a sparse set of potential item locations, and its second stage uses a convolutional neural network to categorise each candidate position as background or one of the foreground classes. Since R-CNN relies on the selective search method to generate suggestions, real-time RCNN execution is not feasible due to its lengthy processing time. The R-CNN model has some significant drawbacks, despite its improved accuracy over a previous object identification model. For example, CNN requires a fixed-size input image, but R-CNN generates arbitrary-sized region proposals. The scale and aspect ratio of the original image were compromised in order to obtain a fixed size input image at that time. The original image's scale and aspect ratio were compromised during cropping and warping in order to obtain a fixed size input image.

The working speed of Fast-RCNN is higher than the R-CNN due to not requirement to feed every region proposal to the convolutional neural network every time. In the original R-CNN, the image is split into several region proposals using a selective search algorithm. Each region proposal is then treated as an independent image and passed through a pre-trained CNN to extract features. his approach is time-consuming as it involves running the CNN independently for each region proposal, leading to redundancy and inefficiency. ast R-CNN improves upon the inefficiencies of R-CNN by introducing a Region of Interest (RoI) pooling laye. Instead of independently passing each region proposal through the CNN, Fast R-CNN takes the entire image as input to the CNN to extract feature maps. he region proposals are then projected onto the feature maps, and RoI pooling is applied to extract fixed-size feature vectors for each region. These region

features are used for both object classification and bounding box regression. Faster R-CNN uses shared computation to use of a single forward pass through the CNN for the entire image reduces redundant computations. Fast R-CNN allows for end-to-end training, optimizing both the CNN and region-based tasks jointly. he elimination of redundant CNN computations makes Fast R-CNN much faster than R-CNN, making it more practical for real-time applications. The working speed of Fast-RCNN is higher than the R-CNN due to not requirement to feed every region proposal to the convolutional neural network every time. Instead, it employs shared computations and RoI pooling to efficiently extract features from the entire image, resulting in improved speed and performance for object detection tasks.

Faster R-CNN is a single and unified network and end to end trainable faster than R-CNN. The deep features extracted by using a region proposal network (RPN) is utilized to suggest approximate objects and finally bounding box are detected. It is a single and unified network and end-to-end trainable and shares computations across all proposals rather than doing the calculations for each proposal independently. The last convolutional layer is used to predict the class scores and corresponding bounding boxes. The region proposals are predicted by the RPN. Anchor is a bounding box that detect the region proposal by sliding window location each time. A threshold value over the "objectness" score is applied for selecting a relevant anchor box. In the first stage, the CNN model computes the feature maps and bounding boxes, and these together are fed to the RoI pooling layer for reshaping the RoI pooling layers' output. Then, the FC layer uses the output of RoI pooling layers for classification and bounding box regression. The last convolution layer divide the result into two parts one part is used to predict the class score and another part is used to predict the bounding box position to localize the objects.

In order to build a feature pyramid network (FPN), top-down and bottom-up architecture are taken into consideration. The top-down pathway creates a multilayer feature

pyramid in which the feature maps form the appropriate stage of the bottom-up pathway as inputs, and each level of the pyramid receives the upper layers of feature maps. These feature maps are concatenated to create the output for each level. A region proposal is produced by RPN, and features are then extracted from the region proposal using FPN. To fuse multiscale properties, ROIs are identified at various levels. Afterwards, such ROIs are reshaped using RoI pooling layers, such as Fast R-CNN. The ROI pooling layer's output is fed into the FC layer, which is used for bounding box regression and classification.

2.2 Segmentation Methods

Template matching and mathematical morphological were applied to automate classification, detection, and segmentation of red lesions. Common steps for classification are preprocessing, detection of initial candidate, extraction of feature. The preprocessing is important to enhance red lesions from the background [25]. Normalization of green and shade correction method [26] were applied for preprocessing. To distinguish micro-aneurysms from other dot-like structures [25], watershed transform was applied. Variable bandwidth is used to estimate karnel density [26], template matching [27] for extraction of haemorrhages. Different circular structures were detected using Radon Cliff operator [28]. For lesion detection, wavelet transform [29] and support vector machine(SVM) [30] were used. The thresholding operation is used to detect micro-aneurysms by making a score map of micro-aneurysms by Lazar et al *et al.*. Rocha et al. used a visual dictionary to bypass the preprocessing or post processing step and detect the point of interest inside the lesion *et al.* To identify micro-aneurysms, directional cross-section profiles were estimated on the local maximum pixels of the pictures. Next, peaks were detected by calculating their size, shape, and height [31].

To detect microaneurysms, a Hough transformation with a feature extraction tech-

nique is used successfully [32]. Kar et al. separated the red lesions from the background using a curvelet-based enhancement technique [33]. The red lesion detection was improved by using information of matched filter with Laplacian and Gaussian filters together [34]. Micro-aneurysms were detected using an unsupervised classification method by Zhou et al. [34]. Dark object filtering process, along with singular spectrum analysis, is used to localize the candidate in the image by Wang et al. [35].

:

CHAPTER 3

Multi-Scale Feature Pyramid for Detection of Red Lesions

3.1 Introduction

DR is an asymptomatic disease in nature, not showing any difficulty in vision in the early stage, but after a long time, vision of the eye may be lost even before 50 years old [1]. The risk of vision loss can be avoided by early DR detection and initiating medical treatment. The severity of DR depends on the degree of glucose level control in blood and the duration of the progression of DR. In general, red lesion micro-aneurasyms, cotton wool spots, white lesions, soft and hard-exudates, and haemorrhages are the pathology seen in the fundus image of eyes. At the initial stage of DR, micro-aneurysms like tiny red circular spots appear in the fundus imageas shown in Fig. 3.1. The severity of DR is determined by the size and number of haemorrhages presents in the fundus image.

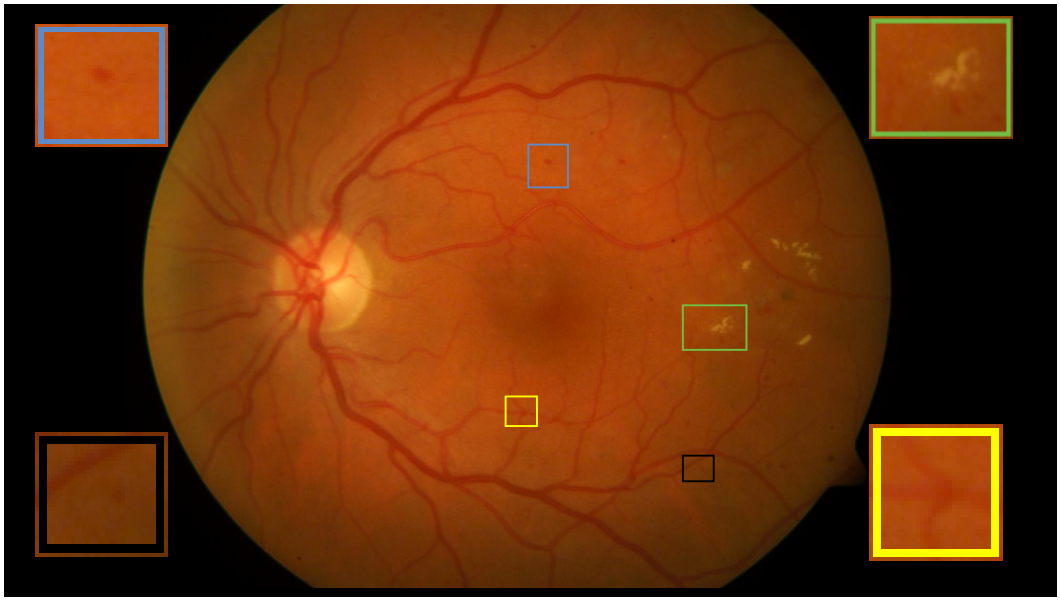

Figure 3.1: Hemorrhages are marked in blue, vessel crossover are marked in yellow and microaneurysms are marked in green.

Ophthalmologists manually detect the micro-aneurysms and haemorrhage spots in image,to understand the severity of DR. In computer-based automation, the detection of DR is challenged due to differences in image contrast, variation in image sizes, color intensity, shape, and size of the pathology present in the image. It is observed that

ambiguity comes to distinguish the small red lesion and vessel junction.Small lesion sizes vary from 25 micrometers to 100 micrometers; therefore, it is challenging to detect them accurately with an automatic diagnosis system. Variation of shape and size of red lesion increases the possibility of false detection of pixel candidates. It enhances the chances of missing out, mainly in the case of small-sized red lesions. Clinical doctors face a challenge in segmenting small lesions and are frequently prone to error. In India, the number of ophthalmologists is much less than the number of DR patients. An automatic diagnosis tool is necessary to handle the massive number of DR patients properly and reduce the burden for clinical persons so ophthalmologists can accurately and effectively diagonalize DR.

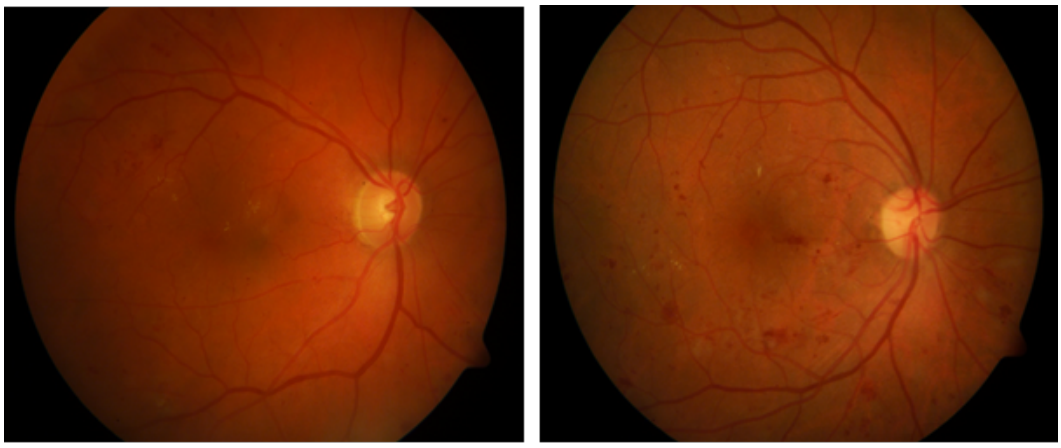

Figure 3.2: (a) and (b) sample relational fundus images from DIARECTDB1 with non-uniform illumination

Several challenges for detection of red lesion are poor image contrast, non-uniform illumination 3.2 and wide variety of red lesions in terms of shapes and lead to missing of red lesions. In India, screening for DR is conducted by means of eye camps, where fundus evaluation is performed for diabetic patients attending the camps. The digital retinal images are analyzed by experts and referable DR cases are segregated for further treatment. The healthy subjects are suggested to attend the next eye camps for follow-up.

3.2 State-of-the-art-methods for red lesion detection

Automated detection of red lesions is affected by low image contrast, non-uniform illumination, and very small structures of microaneurysms. The junction of blood vessels has similarity with red lesions and cause false positives. In classical image analysis, Mathematical morphology [25], template matching [27], cross-section profiles analysis [31], and multiscale filtering [35] were explored for red lesion detection. The microaneurysms were extracted from other dot-like structures using Watershed transform. The hemorrhages of several shapes were taken care by template matching [27]. The circular red lesions were detected by Radon Cliff operator [28]. Wang et al. [35] performed singular spectrum analysis of cross-sections profiles of microaneurysms. Wavelet transform [28] and curvelet-based method [33] were also explored for red lesion detection.

Deep neural network is taught to recognize pathological lesions from retinal fundus images for diagnosis of DR. Grinsven et al. [36] applied CNN to detect hemorrhages based on patch classification, where the negative patches are extracted from images without hemorrhages and positive patches are extracted only from hemorrhages location. The training was made faster by dynamically selecting misclassified negative samples. This method works on the patch level and is not trainable in end-to-end.

We present a single-stage detection framework, which works on image level and end-to-end trainable. The architecture combines a backbone for generating feature map and feature pyramid network (FPN) [37] with skip connections for effectively combining the features of lower resolution with features of the higher resolution images. Finally, classification and regression heads are integrated to recognize the red lesions with their locations. The FPN take care several limitations of the previous architectures and creates feature pyramid with strong semantics at several scales. The proposed network is could be trained in image level in all possible scales and works for a wide range of red lesions.

3.3 Multi-Scale Feature Pyramid for Red Lesion Detection

Object detection at a wide range of scales is a fundamental challenge, and feature pyramid is used as a neck in the architecture with classification and regression head. The feature extracted by convolutional networks is used to create feature pyramid. The pyramids are scale-invariant the model is able to detect objects of different shapes. The proposed network consists of a top-down and a bottom-up architecture with skip connections for generating better semantic features. Non-uniform illumination and poor contrast are observed in fundus images. Therefore, a preprocessing method is developed to improve image contrast so that red lesions are visible in preprocessed images.

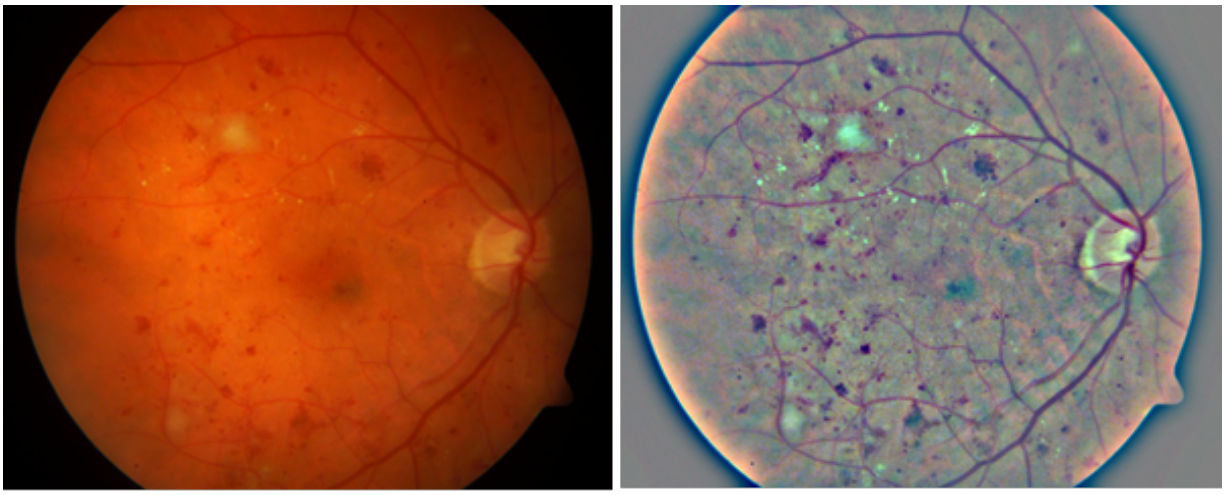

Figure 3.3: Preprocessing step: (a) fundus image and (b) fundus image after preprocessing

3.3.1 Preprocessing

The maximum illumination in fundus image is observed near the optic disc and boundary region has less illumination. Therefore, illumination equalization is very important as a preprocessing step. The fundus portion is extracted from the big-size fundus images using template matching and then the images are resized to 512×512 to make the process faster. A contrast enhancement is performed as shown in Fig. 3.3 has high

contrast with better visibility of red lesions for the low contrast input image in Fig

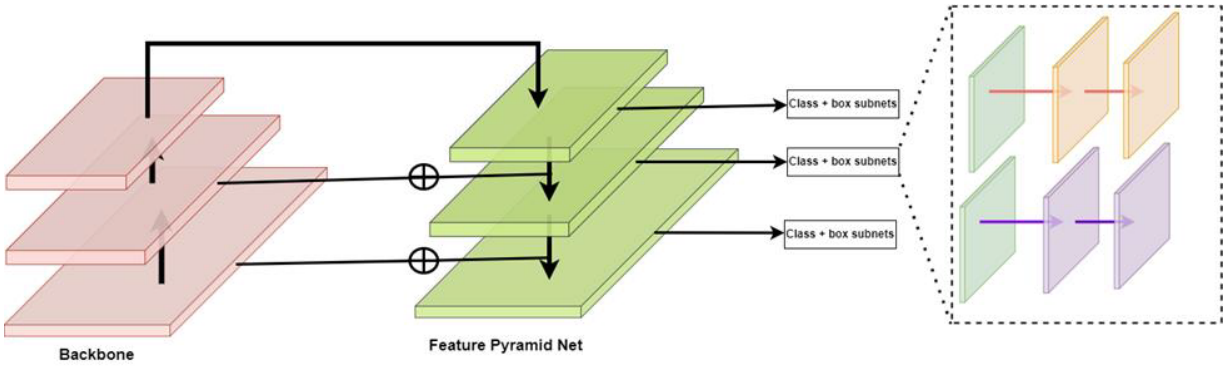

Figure 3.4: Multi-scale feature pyramid with classification and regression head for red lesion detection

3.3.2 Network Architecture

The one-stage detectors are very popular because of their simpler design and acceptable performance. The proposed detection framework (Fig. 3.4) is one stage and has similarities with RetinaNet [24]. The network consists of a backbone for the extraction of features, the FPN, and the object detector head. The CNN is used to build the feature pyramid and throughout the pyramidal structure, high-level semantics is observed. The proposed method takes a preprocessed fundus image as input and creates feature map with several resolutions. The FPN helps to combine lower-resolution features with features of higher resolution. The FPN provides a semantically strong multi-scale feature map with all possible scales.

The bottom-up pathway computes a feature map with several resolutions and feature pyramid is generated. The top-down up-samples spatially coarser features, then combined with the feature pyramid via skip connections. The skip connection merges feature maps of the bottom-up path-way and the top-down pathway, where the same spatial size is considered during merging. The bottom-up feature map has the semantics of lower-level as it is sub-sampled fewer times. The top-down pyramid has high semantic information and it has fine resolutions. The classification and regression heads are

similar to standard object detection method. The focal loss is an improvement of cross entropy by dynamically scaling the cross entropy loss so that little loss is assigned to well-classified examples and incurs high loss value for red lesion candidates, which are difficult to detect.

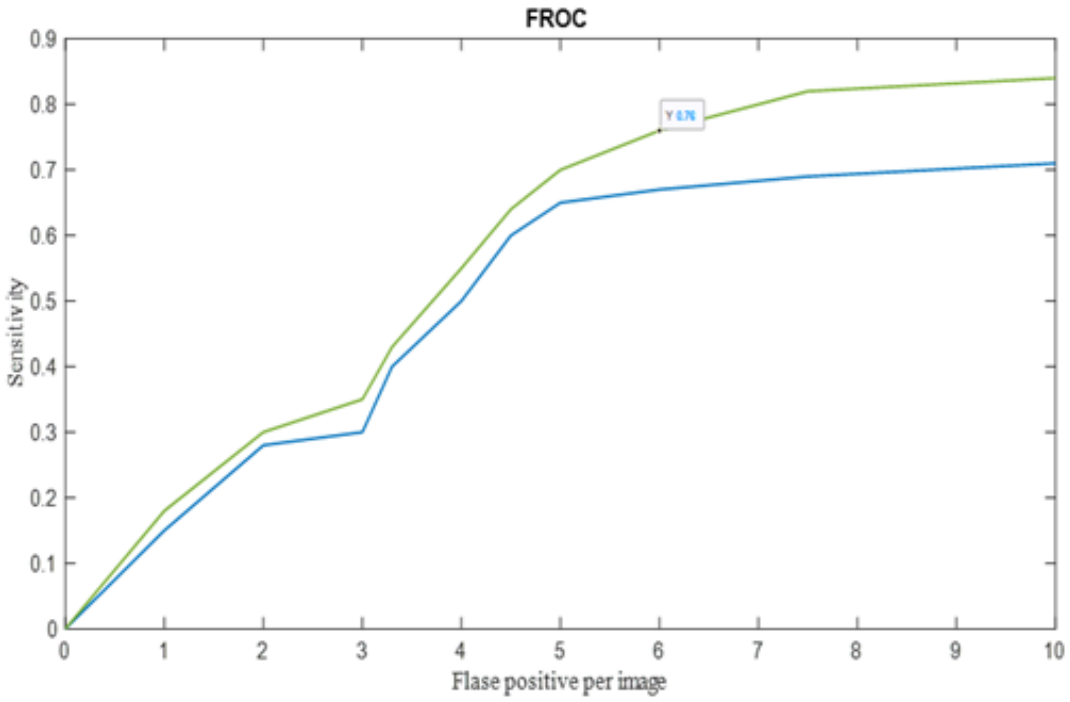

Figure 3.5: FROC of multi-scale feature pyramid network for red lesion detection

Anchor box tuning

Several detectors, such as Faster R-CNN [38], SSD [22], and RetinaNet [24] use pre-defined boxes and could hamper the ability of generalizing detectors because they need tuning onmfresh detection tasks. In this work, suitable anchor boxes were generated adaptively based on the statistical analysis of sizes of red lesions on the test set of DIARECTDB1 [13] public dataset. In this application, anchor boxes are of three scales and four aspect ratios to detect red lesions large with imbalances in object sizes and

shape variations as well as very small red lesions.

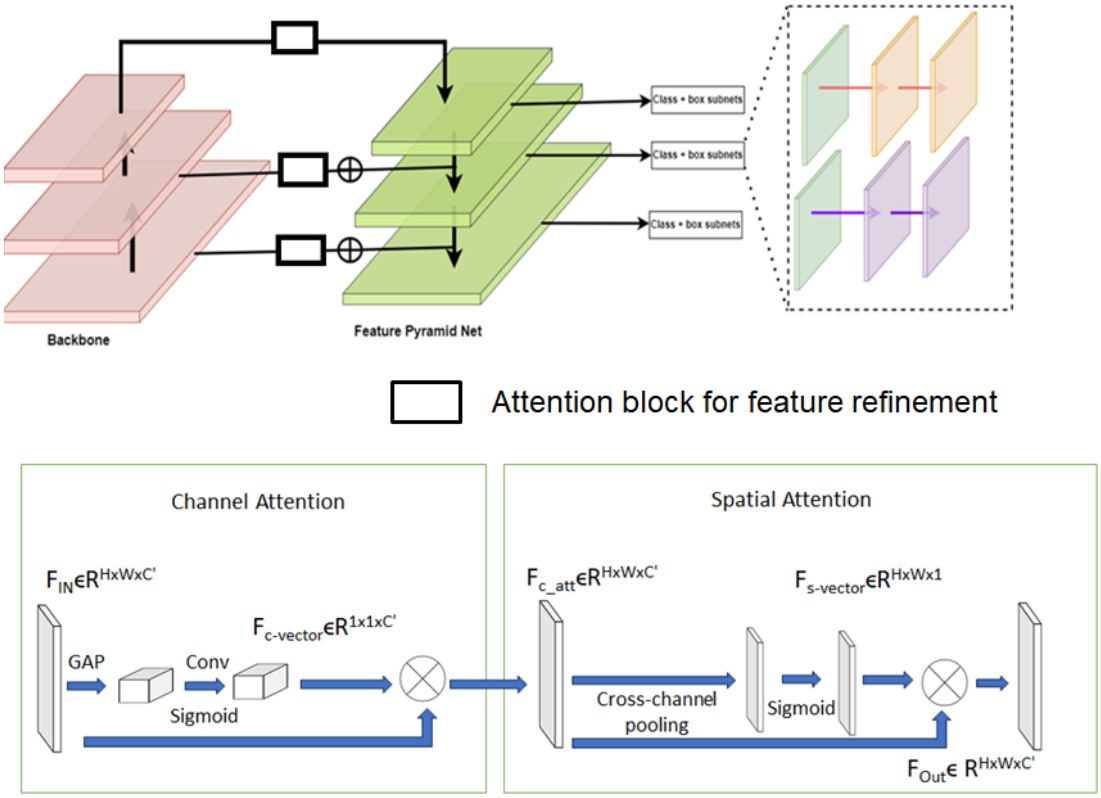

Figure 3.6: Attention augmented feature pyramid network with classification and regression head for small red lesion detection

Network Training

The proposed method is implemented in PyTorch and training is performed using a workstation with 16 GB GPU and 64 GB RAM. The training is performed using the pre-processed fundus images of the training set of DIARECTDB1 [13] public dataset up to 150 epochs. The learning rate is 0.00001. Area of the regions of interest is very small as compared to size of image and could results severe class imbalance. Therefore, training is dominated by irrelevant background region and the easily classifiable background region lead to poor learning of the model. The focal loss could handle the class imbalance and the network is being trained using hard examples, i.e., red lesions which are difficult to detect as, compared to easy examples, i.e., red lesions which are easy to

detect.

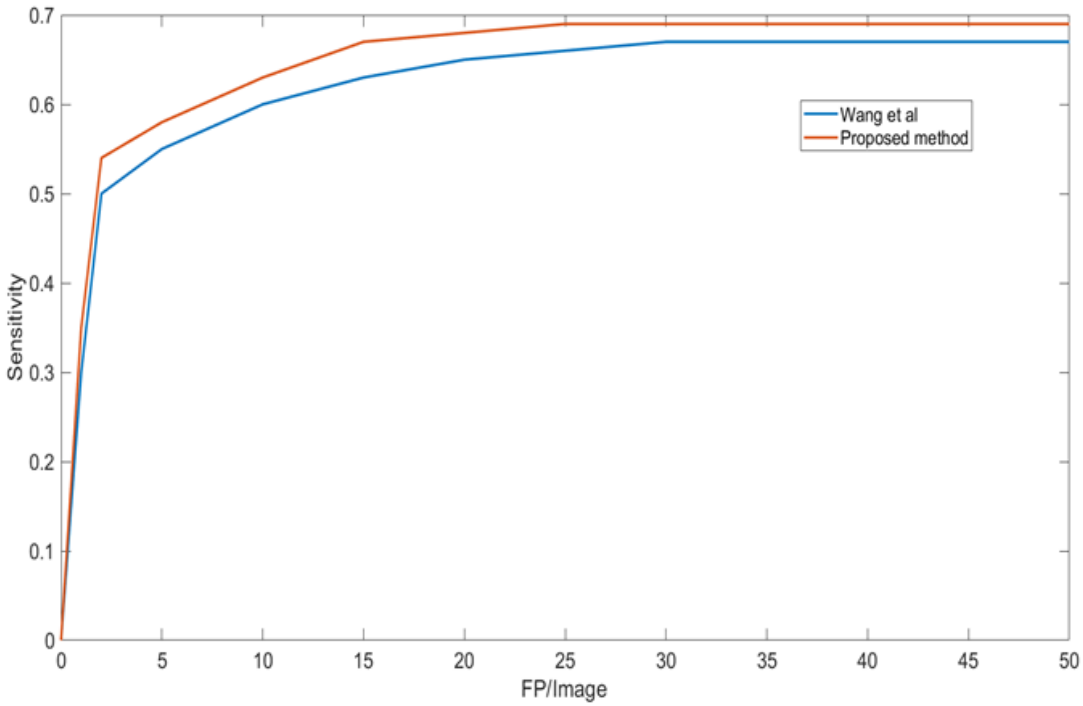

Figure 3.7: FROC of attention augmented feature pyramid network

Attention augmented feature pyramid network (Fig. 3.6) with classification and regression head is implemented for small red lesion detection. The network is constructed with multiple attention blocks to generate attention aware features to highlight MAs. The attention block is constructed by channel attention and spatial attention block and serves as a feature selector in the forward pass. The test set of DIARETDB1 dataset is used for the evaluation of methods on per-lesion basis using FROC as shown in Fig. 3.7. The proposed network is end-to-end trainable with multiple scales and leads to detection of red lesions with wide range of shape and sizes. The focal loss [24] is considered as compared to cross entropy and Dice loss for training of the network on hard examples as compared to easy negatives lesions or background.

3.4 Results and Discussions

The proposed method is compared with method of Seoud et. al. [1] and the methods are evaluated on the publicly available DIARETDB1 dataset. Total 89 fundus images are available in the database and the images are taken using 50-degree field of view. Out of 89 images, 84 images are mild or non-proliferative DR and 5 images are normal. Four expert ophthalmologists annotated several diabetic fundus lesions such as soft exudates, hard exudates, microanneurysm, and hemorrhages and the ground truths are at coarse level. The detection algorithms are evaluated on using those annotations. In this dataset, detection of red lesions is really challenging because of variation in imaging conditions which includes saturation, lighting condition, and different types of blurs. In this experiment, the lesions annotated in fundus image by at least one ophthalmologist is considered as ground truth

The methods have been evaluated on per-lesion basis, where a delineation of all the lesions was provided. The test set of DIARETDB1 dataset is used for the evaluation of methods. The performance is analyzed on every single lesion by means of free-response receiver operating characteristic (FROC). In the per-lesion evaluation strategy, sensitivities are plotted with respect to the average number of false positives per image for several operating points. The FROC plot of the proposed method and the method of Seoud et. al. [1] is shown in Fig. 3.5. The results depict that the proposed method outperforms the competing methods

The proposed network is end-to-end trainable at image level with all possible scales and achieves a sensitivity of 0.76 with six false-positive per image on the test set of publicly available DIARECTDB1 database and outperforms the method of Seoud et. al. [1]. Visuals result of the proposed method is shown in Fig. 3.8 and it is observed that red lesions with a wide range of shapes are detected. The performance of detection is better for hemorrhages as compared to microanneurysm. A few vessel-junction has

Figure 3.8: Visual results for a few test images of DIARECTDB1 database, where true positives, false positives, and false negatives are marked with green, red and blue boxes.

similarity with red lesion and appears as false positives. The method has been evaluated with and without preprocessing, and it is observed that preprocessing has a significant role in performance of red lesion detection

3.5 Conclusion

The proposed network detects red lesions with a fewer number of false-negative and could be a part of DR screening tool. More work is needed to focus on developing red lesion detection methods with a minimum number of false-negative and false-positive and developing robust methods for grading the fundus images for diagnosis of DR. The preprocessing technique could be improved for better visualization of small pathologies in fundus images.

CHAPTER 4

Segmentation of Red Lesions using Nested U-Net

4.1 Introduction

In traditional procedures, the segmentation work for red lesions depend on hand-crafted low-level features of the image. The morphological approaches have issues with the empirically selecting the different parameters. The high-level feature is extracted by the decoder, and the encoder extracts the low-level feature. Using dense convolutional neural network and the feature maps which are taken from the background and foreground helps to upgrade the result in comparison with U-Net [39] and attention U-Netand attention U-Net [40]. Res-Net-18 [41] is used to improve the findings through sub-image classification. The proposed U-Net++ [12] architecture can produce less numbers of false detection in comparison with U-Net [39] and attention U-Net. An ablation of the study of U-Net++ [12] can find its depth. The suggested network is robust and efficient, quickly segmenting the red lesion with higher accuracy.

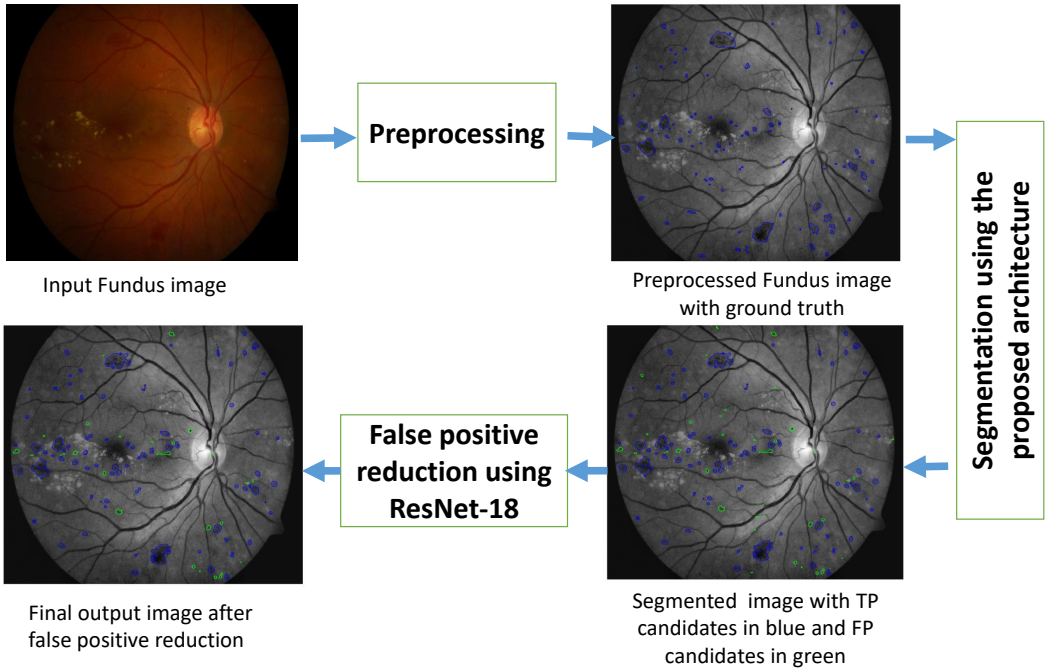

Figure 4.1: Block diagram of the proposed methodology

4.2 Methods for Segmentation of Red Lesions

The suggested method has two stages. In the first stage, red lesion segmentation is done using a Nested U-Net [12]. In the second stage, the number of false-positive candidate reduction is done through sub-image classification. The block diagram of the suggested method is shown in Fig. 4.1, and in the subsequent section, details of the methodology is explained.

4.2.1 Preprocessing

Uneven illumination and poor contrast between the background and objects is an important issue to segment the pathology from retinal fundus images. The illumination is maximum near the optic disc area, gradually decreasing towards the periphery of the fundus image. Out of three color channels, the green channel provides the maximum contrast for red lesions. To preserve the semantic information and maintain high resolution, the large-size fundus images are subdivided into 256×256 pixels, which is possible due to the GPU memory capacity.The visibility of red lesions is enhanced by using contrast-limited adaptive histogram equalization(CLAHE), as shown in Fig. 4.2.

(a) (b)

Figure 4.2: (a) Normal image and (b) result after preprocessing.

4.2.2 Network architecture for red lesions segmentation

The U-Net++[12] network comprises an encoder path and decoder path for medical image segmentation. The decoder extracts the semantic, high-level, and course features, whereas the encoder extracts low-level and fine feature maps. In the suggested network, several dense CNN blocks are linked between the decoder and encoder paths. The combination of the ecoder and decoder path is used to segment red lesions in U-Net++[12] architecture. To get better segmentation results high-level and low-level information is necessary, which is actually done by the U-Net++[12]. The main path of this network, i.e, upsampling and downsampling, is shown in Fig. 4.3. A series of convolutional blocks is used as a skip connection between the encoder and decoder path. A total of twenty-one blocks is used in the network, and each block is built by two convolutional layers. Six convolutional blocks are used with single padding and kernel size 3×3. Except the last convolution block after each encoder block, the max-pooling operation is performed with a filter size 2×2 and stride size two. The feature map size increases from 64 to 2048 along the down-sampling path.

The image size decreases along encoder path from 256×256 to 8×8. Five convolution blocks make the decoder path maintain a sequence of up-sampling and concatenation after each block. Along decoder path feature map size decrease from 2048 to 64, and image size increased from 8×8 to 256×256 . Horizontal padding is used by dense convolution blocks between the decoder and encoder. Four dense convolution blocks are used at the top-most portion between the encoder and decoder, after the number of dense convolution blocks is reduced by one along the downward path of the network. Each block in the horizontal path is connected via skip connections. The block does up-sampling operations that are parallel to the decoder path. The outputs from the previous dense convolution blocks are added to the up-sampling outputs from the lower dense convolutional block by a concatenation operation before each convolution block of the dense layer. The dense blocks properly combine the encoder and decoder feature maps to be semantically identical. The main focus of this network is the segmentation

of red lesions with minimum false-negative candidates.

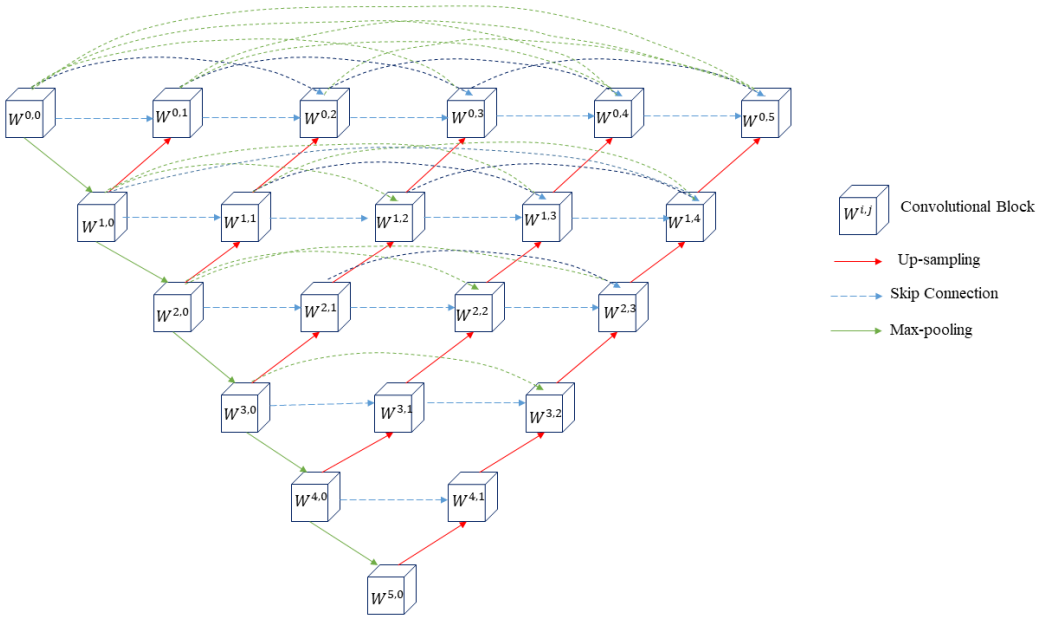

Figure 4.3: Network architecture of proposed segmentation method.

In the above figure, W^{ij} represents the output of the corresponding block and w^{ij} is it's output, where i represents the downward block along the decoder connections, and j denotes the dense convolution block attached with skip connections. Blocks with $j = 0$ receive the output from its preceding block of the encoder following max pooling. Block with $j = 1$ collects the output result from the previous encoder block following the maxpooling operation. For $j > 1$, the convolution blocks collect j number of input from the previous dense convolution block through the skip connection and another from the dense layer below after the up-sampling. As a result each block gets the feature map from its earlier blocks. At the final stage of the network, the feature maps are created by input image after propagating in all possible paths of the network.

$$w^{ij} = \Gamma(\{w^{i-1,j}\}) \quad \text{for} \quad j = 0, i > 0 \tag{4.1}$$

$$w^{ij} = \Gamma([w^{i,k}]_{k=0}^{j-1}, \mu(w^{i+1,j-1})]) \quad \text{for} \quad i, j > 0 \tag{4.2}$$

where $\Gamma(.)$ denotes convolution, $\{.\}$ denotes max pooling, $[.]$ denotes concatenation and $\mu(.)$ denotes up sampling.

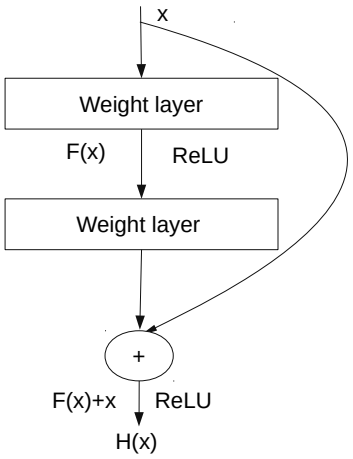

Figure 4.4: The Residual Block

Sub-image classification for false-positive reduction

Res-Net-18 [41] is commonly used to classify the images. Res-Net [41] is specially used to sub-image classification of red lesions or without red lesions in images. It has been seen that 80×80 is the optimal sub-image size is used in this network. The Res-Net architecture consists of 18 numbers of convolutional layers, with each kernel size 3×3. Along the downward path, the feature size increases from 64 to 512. The last layer of Res-Net is a fully connected neural network. The skip connections are connected after each of the two convolution layer intervals except the first convolutional layer.Such type of unit is called residual block [41]. The architecture of the residual block is shown in Fig. 4.4, where a curved line with an arrow mark indicates the skip connection. The mathematical representation of the residual block is as follows.

$$H(x) = F(x) + x \tag{4.3}$$

$$F(x) = H(x) - x \tag{4.4}$$

Here, the input and output of the residual block are represented by x and H(x) respectively. According to Fig. 4.4, each residual block consists of two weighted layers, and a skip connection is used to preserve the previous layer information. The increased convolution layer in the residual block reduces the network distortion. The network learns low-level features like texture, objects' edge, and objects' structure from shallow layers and high-level features like semantic and spatial features from deep layers. It observed that the deep layer, most of the time, loses the low-level feature. But in Res-Net[18], residual block plays a vital role in retaining the low-level feature, which makes better output than the other network architecture.

Training

MASSIDOR data set [42] is used to training the proposed network. Each image was divided into eight pieces and t it resized with 256×256 pixels. ADAM [43] optimizer is used to train the network to maintain the learning rate 0.00001 from one to fifty epochs; then, it is observed that the learning rate decreases to 0.000001 from 51 to 75 epochs. For updating the network 80×80 patch size is used. The network is trained up to 60 number of epochs. A learning rate of 0.01 and momentum of 0.6 is used for stochastic gradient descent (SGD) [44]. It is also observed that the learning decreases by 0.0001 after each 15 epoch.

4.3 Description of Database

In the training process MASSIDOR data [42] set is used, publicly available on the internet. The network is taken from scratch. A total 1200 number of colored fundus images are available in MESSIDOR dataset [42]. The images are captured by using non-mydriatic camera with with a 45 degree field of view. Images were of 1440×960, 2240×1488 or 2304×1536 pixels. Out of 1200 images, 742 are under grade-1,grade-2 and grade-3 categories of DR. Only 572 images are used for network, and 170 numbers are used for result validation.As the ground truth of MESSIDOR data set is publicly

unavailable, manual annotation is done under the supervision of an ophthalmologist.

DIARETDB1 data set [45] has a total 89 numbers of colored fundus images of eyes. Out of 89, 84 images contain at least one pathology of mild or non-proliferative DR, and the remaining five are normal and healthy. The images are captured consider 50 degree field of view. Four expert doctors annotate the publicly available ground truth of DIARETDB1; therefore, it is used for DR detection. In this work, the boundary of the red lesion is again redefined under the supervision of an ophthalmologists.

4.4 Results and Discussions

4.4.1 Metric

The proposed network's performance and competing methods were evaluated by conventional metrics such as sensitivity, accuracy, precision, specificity, and FI score, as given below.

$$Sensitivity = \frac{TP}{TP + FN} \tag{4.5}$$

$$Specificity = \frac{TN}{TN + FP} \tag{4.6}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.7}$$

$$Precision = \frac{TP}{TP + FP} \tag{4.8}$$

$$F1 - Score = \frac{2TP}{2TP + FP + FN} \tag{4.9}$$

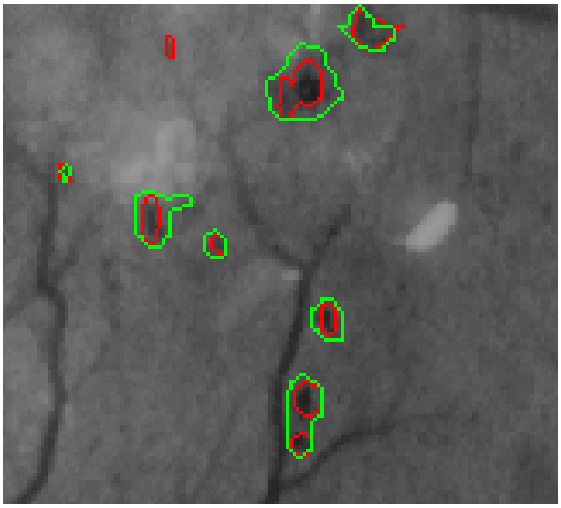

Figure 4.5: predicted lesions and overlap with ground truth.

For the purpose of validation, connected component level validation is considered as follows [46], [47]. In Fig. 4.5, ground truth is shown by red and green contours, indicating the predicted lesions by the proposed methodology. In performance evaluation consider a minimal overlap ratio 0.2 between truly positive pixels from connected components and ground truth instead of counting the classified pixels. If the overlap ratio is less than 0.2, then the non-overlapping portion of predicted lesions could be considered false-positive pixels. The non-overlapping portion of ground truth is considered false-negative pixels; the rest are considered true negatives. In particular, component level validation is used when the boundary of red lesions are not identified to evaluate the performance metrics.

Segmentation results

To evaluate the performance of the proposed network and compare the results with other methods DIARETDB1 [45] data set is used. Out of 89 numbers of images, 28 are used for fine-tuning of pre pre-trained network, and the remaining 61 images are used to report the result. The exact number of images from DIRETDB1 is used for

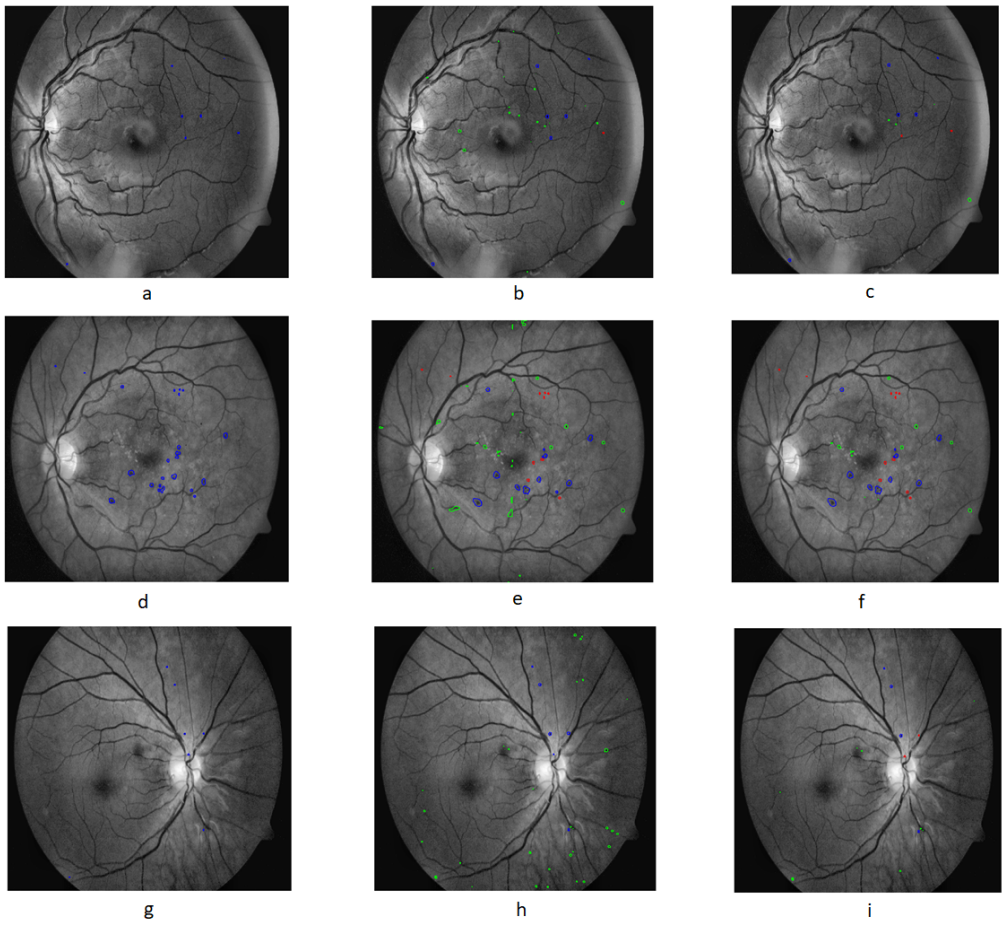

Figure 4.6: ground truths: (a), (d) and (g), segmentation results using proposed network: (b), (e) and (h), result after false-positives removal: (c), (f) and (i). Red corresponds to false negative candidates, green represents the false positive, and true positive candidates are true positive candidates are in blue.

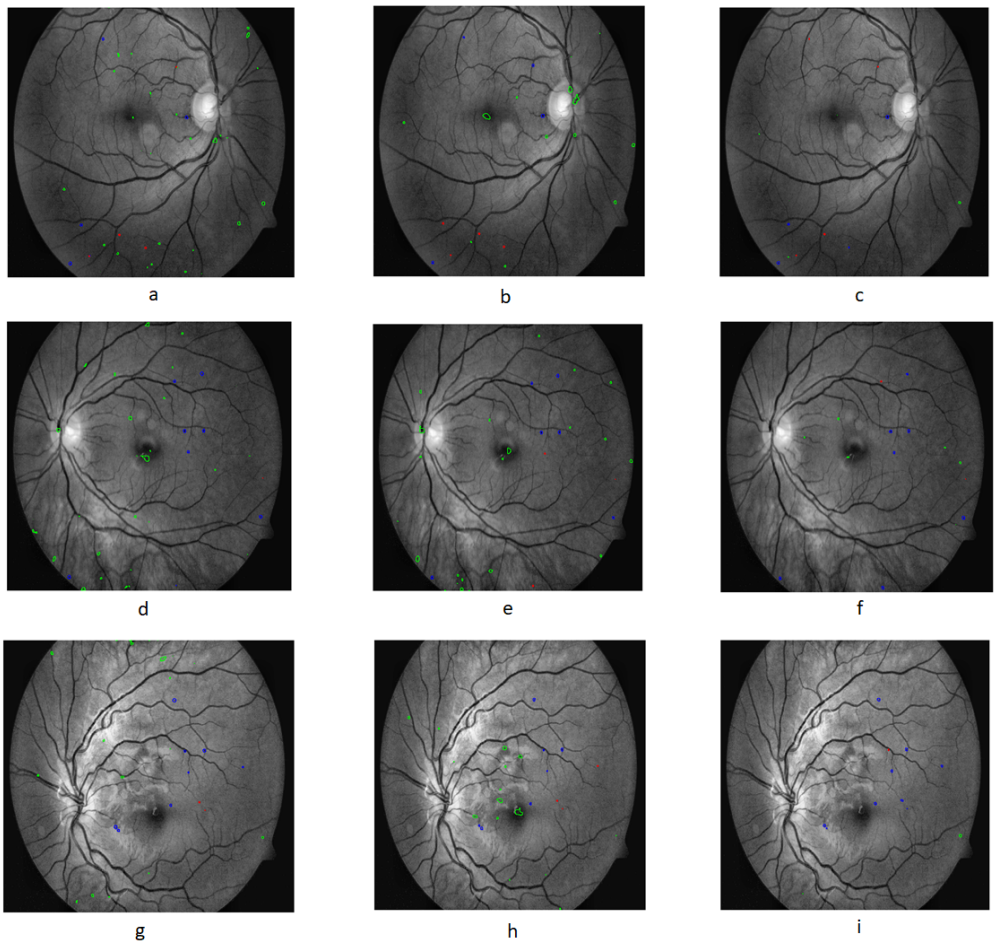

Figure 4.7: Results of U-Net: (a), (d) and (g), results of attention U-Net (b),(e) and (h), and results of proposed method presented in (c), (f) and (i). Red corresponds to false negative candidates, green represents the false positive, and true positive candidates are true positive candidates are in blue.

pre-train and reported the result for U-Net [40] and attention U-Net [40]. The proposed network architecture is showing a significant improvement result in terms of sensitivity and precision(Table 4.1). The sensitivity is improved to 8.87% over the U-Net [40], and precision is improved to 8.83% over the attention U-Net [40].

The ground truth of red lesions is represented in the first column in Fig. 4.6. The initial segmentation result of the suggested network is given in the second column, and the third column shows the segmentation results after false-positive reduction. The small size with poor contrast red lesions are successfully segmented by this U-Net++ architecture. This proposed network can also segment large size as well as very small size red lesions.The segmentation results of red lesions using U-Net [39], attention U-Net [40], and the proposed method are given in the 1st,2nd, and 3rd columns in 4.7. The result shows that the number of false-positive and false-negative candidates are significantly less in the proposed method in comparison with other methods.

Table 4.1: Performance Evaluation for Red lesion Segmentation

Segmentation Architecture Used	**Sensitivity**	**Specificity**	**Accuracy**	**Precision**	**F1-Score**
U-Net [39]	79.92	99.65	99.48	67.88	73.41
Attention U-Net [40]	85.10	99.60	99.49	62.67	72.18
Multi task Architecture [47]	85.18	99.89	99.83	78.96	81.95
Proposed method	88.79	99.64	99.53	71.50	79.21

4.5 Conclusion

The proposed technique has sufficient potential to assist ophthalmologists in DR diagnosis. This technique uses a minimum number of images to train the network. Therefore, with this minimum information, the method can produce accurate outcomes.

The suggested technique can successfully segment red lesions from the given fundus image with a few numbers of false-negative candidates.In this technique the number of false-positive candidates also decrease to a large extent. However, further rigorous study on this network is required to minimize the number of false-positive and false-negative candidates. There is a chance to improve the preprocessing technique for visualization and segmentation of small size red lesions. The detection and segmentation of new vascularisation is not addressed in this work, which is an limitations of the work. Micro-aneurysms can successfully segmented with the minimum number of false positive by this method. This cost-effective proposed method can be used to detect DR at the beginning stage of DR, and as a result, permanent blindness due DR can be preventable. This computer-based diagnosis method may play a vital role to making clinical decisions by opthalmologists, especially NPDR patients.

CHAPTER 5

Conclusions and Future Directions

5.1 Summary of Studies

The rapid growth of DR needs immediate attention to reduce blindness. Early detection and medication are very important to prevent the development of DR. The accuracy and reliability of computer-based diagnosis of DR greatly depend on the accurate detection of pathology of DR in retinal fundus images. This thesis presents novel methods for detection of red lesions in retinal fundus images as well as quantification of red lesions. A brief description of the prevalence of DR around the world is described in Chapter 1. State-of-the-art methods on object detection are described in Chapter 2. From the review it is evident that there is a need and scope for further improvement of detection of red lesions in fundus images. Specially, early detection of microaneurysm is very important as it is early indication of DR. Chapter 3, comprises methodology of an accurate and fast red lesions detection. Chapter 4 comprises about segmentation technique of red lesions from fundus image of retina.

A novel multi-scale feature pyramid network is developed for automatic detection of microaneurysms and hemorrhages from color fundus images. A pyramid of features is generated with strong semantics. Feature pyramid architecture extract the multiscale feature of image. Attention augmented feature pyramid network is developed to detect small red lesions, mainly microaneurysms. The advantage of the suggested network is that it is end-to-end trainable using the direct image as input, can handle a wide range of red lesions with several scales, and can provide acceptable outcomes.

The suggested network can able to reduce the false-positive cadidate to a large extent,and also it can segment red lesions with few numbers of false-negative. The procedure used can efficiently segment microaneurysms with minimum numbers of false-positive candidates. The proposed method is cost-effective and can help in the early detection of DR to prevent the permanent blindness of DR patients. The proposed method can be used for a computer-based diagnosis system and help ophthalmologists

make clinical decisions for DR patients.

5.2 Contributions of the Thesis

The contributions of the thesis towards the early diagnosis of DR are summarized as follows.

- The main contribution of the proposed method is to develop a computer-based deep learning framework that can handle severe class imbalance and size imbalance of red lesions. Deep learning based DR detection system has an advantage in case of DR screening program of referable DR .
- Novel attention augmented feature pyramid network is developed to detect small red lesions, mainly microaneurysms. Cross entropy, dice, and focal loss were used and focal loss provides best results as compared to other loss function.
- Novel deep neural network is developed for segmentation of red lesions. With this minimum information, the method can produce accurate outcomes. There is a chance to improve the preprocessing technique for visualization and segmentation of small size red lesions. Micro- aneurysms can successfully segmented with the minimum number of false positive by this method. The proposed technique has sufficient potential to assist ophthalmologists in DR diagnosis.

5.3 Future Scope

The investigation carried out in the present work leaves scope for extension of the concepts presented in this thesis for diagnosis of referable DR from fundus images. Potential future lines of investigation related to the thesis are listed below:

- To increase the accuracy and robustness of the proposed method, further research is required to improve the preprocessing technique, as it could help identify small-sized pathology in fundus images.

- More work is needed to focus on developing attention based detector to detect red lesions with a minimum number of false-negative and false-positives.
- The detection and segmentation of new vascularisation is not addressed in this work, which is an limitations of the work.
- Exploring on domain attention to handle domain shift and robust detector.
- To reduce false-negative and false-positive candidate in red lesion segmentation, furtherresearch work is necessary.
- In this work, detection of neovascularization was not addressed; therefore, it could be considered a limitation of this work.

◇

Appendices

APPENDIX A

Machine Learning Fundamentals

A.1 Machine Learning Fundamentals

A.1.1 Introduction

Artificial intelligence (AI) is an emerging technology, and all are interested in learning this subject. Numerous articles, even in non-tech-focused publications, frequently discuss machine learning, deep learning, and AI. Fig. A.1 is showing Van diagram about the relation between AI, Machine learning and deep learning. The three terms are artificial intelligence (AI), machine learning, and deep learning. Deep learning is a subset of machine learning,again machine learning is a subset of AI, therefore AI is the super set of deep learning and machine learning. Even though deep learning is an effective tool in machine learning, not all machine learning applications require deep learning. They are arranged in a hierarchy.

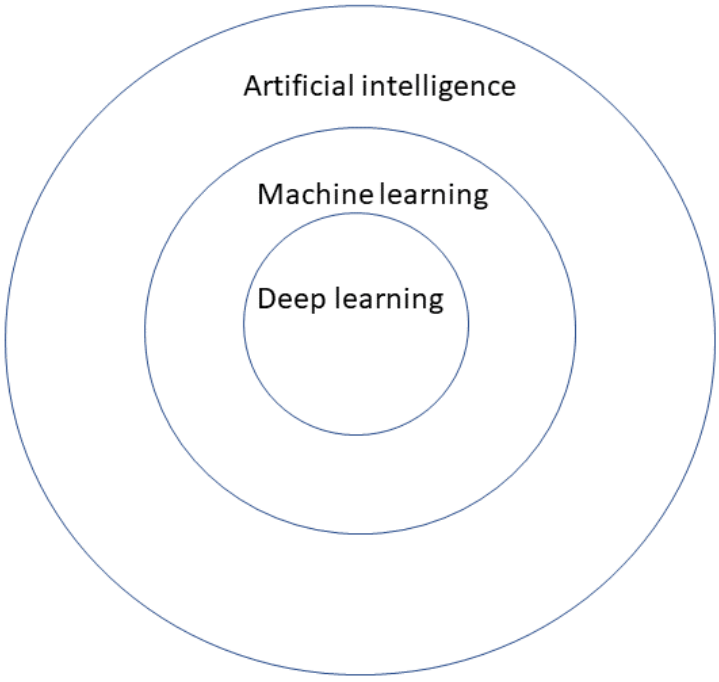

Figure A.1: Pictorial presentation of Artificial Intelligence, machine learning and deep learning

The AI concept emerged in the 1950's when a small group of pioneers in the field of computer science began to inquire whether it was possible to make a machine that can think and decide like human being. This question, with its ongoing implications, continues to be a subject of exploration even today. Machine learning, deep learning, and other non-learning approaches are all covered by the extensive area of AI. For example, early chess algorithms did not fall under the category of machine learning because they mainly used hard-coded rules programmed by programmers. This type of

programming approach that involves the manual crafting of an extensive set of explicit rules for manipulating knowledge is known as symbolic AI. This was the prevailing approach in AI from the 1950s to the late 1980s and many experts popularised this by making expert systems during 1980's. While symbolic AI demonstrated effectiveness in addressing clearly defined, logical challenges like chess-playing, it faced challenges in formulating explicit rules for tackling more complicated, ambiguous problems such as image classification, speech recognition, and language translation. In response to this, a new approach emerged to replace symbolic AI: machine learning.

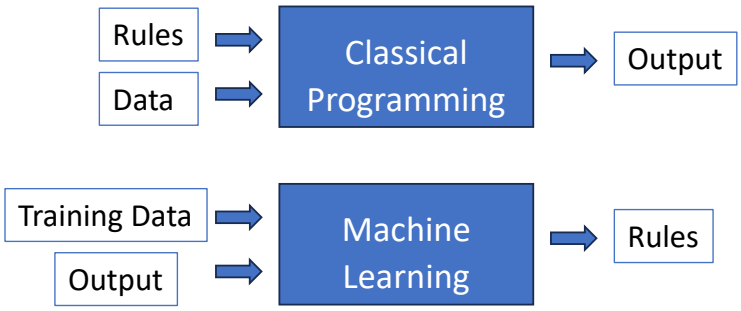

Figure A.2: Pictorial representation of Machine learning methodology.

Machine learning is a technology where machines learn from examples and use this experience to evaluate a fresh task. Machine learning is a subset of AI. The origin of machine learning can be traced back to the fundamental question: Can a computer extend its capabilities beyond predefined instructions and autonomously acquire the ability to execute a given task? Is it possible for a computer to autonomously acquire data-processing rules by analysing data, instead of relying on programmers to manually formulate these rules? In classical programming, we write programs containing rules and give data as input on which it runs and provides output. On the other hand, A machine learning algorithm based on mathematical modeling where the machine learns from examples during the training period and then evaluates a new task without explicit programming as shown in Fig. A.2. A machine-learning system does not have explicit programming; instead, it is trained. It is exposed to numerous examples related to a specific task from where it derives statistical patterns within these examples, enabling the system to formulate rules for automating the task. For example, if the goal is to automate the tagging of vacation pictures, the machine-learning system would be fed numerous instances of pictures already tagged by humans. Subsequently, the system learns statistical patterns to associate particular pictures with specific tags. Machine learning techniques are helpful in various fields, including computer vision and email filtering, where traditional algorithms are impractical or burdensome to develop for the task.

A special area of machine learning known as deep learning emphasizes learning suc-

cessive layers of progressively more significant representations, offering a novel approach to learning representations from data. It is basically a neural network with three or more layers aiming to replicate the functioning of the human brain. It represents the concept of progressively more complex representations. The depth of a model is the number of layers that make up the model of the data. The kinds of data that deep learning uses and the ways that it learns set it apart from traditional machine learning. Deep learning algorithms automate feature extraction, reducing the need for human specialists having the capability to intake and analyze unstructured data, such as text and images. Everyday applications of deep learning in the real world have seamlessly become a routine aspect of our lives. Deep learning based applications are used in many industries, including Automated driving, Speech recognition, Fraud detection, Visual recognition, Chatbots, Recommender systems, Facial recognition and many more.

A.1.2 Learning algorithms

Machine learning algorithms primarily cater to four distinct learning categories.

Supervised Learning

In supervised learning training of of the machine done by providing some inputs example called labaled data such as price of house,some image with ground truth,spam or not spam mail. It involves learning to map using a given set of annotated samples to a known target. The majority of current applications in the spotlight, including optical character recognition, speech recognition, image classification, and language translation etc. fall into this particular category of deep learning. In this approach, labeled datasets are created to guide or supervise the algorithms to accurately classify data or make predictions. By utilizing inputs and outputs that are labeled, the model can assess its accuracy and progressively improve through learning over time. Until the model attains the required degree of accuracy on the training set, the training procedure is repeated. Supervised learning is mainly of two types - classification and regression.

Classification problems deal with classifying the input sample or object into a predefined list of categories e.g. cat, dog, bird, house, car, bus, etc. We have seen that our spam emails go into a separate folder in the mailbox after marking specific types of emails marked as spam by the user over time. In supervised learning diffrent algorithm are used for classification purpose like random forest algorithm,decision making algorithm,linear classifier algorithm and support vector machine.

Linear regression method is also a supervised learning technique. This technique predicts a continuous outcome variable based on one or more predictor variables. The purpose of regression is to create a mapping between the target variable and the input feature. The key components of regression in supervised learning are mainly - Features, target variables, training data, regression model, and prediction. Regression models can be helpful when predicting numerical values based on multiple data points, such as

predicting earnings from sales for a particular business.

Apart from classification and regression, there are some more variants in supervised learning, e.g. i) Sequence generation, e.g. using an image, anticipating and generating a descriptive caption for it ii) Syntax tree prediction, e.g. predicting how a sentence will break down into a syntax tree given the sentence. iii) Object detection iv) image segmentation etc.

Unsupervised Learning

In this context, the input data lacks labels and definitive outcomes. A model is constructed by identifying inherent structures within the input data, aiming to derive general rules or cluster the unlabeled data sets. This process might involve a mathematical approach to systematically minimize redundancy, or it could follow grouping the data based on similarities. These algorithms show concealed patterns in data autonomously, eliminating the necessity for human involvement. Example problems are clustering, dimensionality reduction, and association rule learning.

Clustering is a type of algorithm that groups similar data points into clusters, where items within the same cluster share common characteristics, while those in different clusters are dissimilar. The algorithm does not have prior knowledge of the specific classes or categories; instead, it organizes the data based on inherent similarities. K-means clustering techniques, for instance, group together comparable data points; the K value denotes the granularity and size of the grouping. This method works well for image compression and market segmentation, among other things.

Dimensionality reduction Certain cases, when in a dataset the number of variable or feature is decrease due to preserve the information or patterns a learning technique called dimensionally reduction technique is used. It does not use labeled output data but rather explores the internal structure and relationships within the data. This technique is frequently applied to feature extraction, feature selection, and other preprocessing stages of data.

Association is another type of unsupervised technique is Association,that employs several rule to identify the relation between parameters in a given dataset. This technique have been used in various fields, including retail, healthcare, and recommendation systems, where discovering hidden patterns and connections within data can lead to insights or improvements in decision-making processes. The pictorial presentation is shown in Fig. A.3

Self-supervised learning

In this learning paradigm the model is trained to learn representations from the data itself without human-annotated labels. Although labels are still necessary because someone has to supervise the learning, they are produced from the input data, usually by a heuristic algorithm. Some Common approaches in self-supervised learning are -

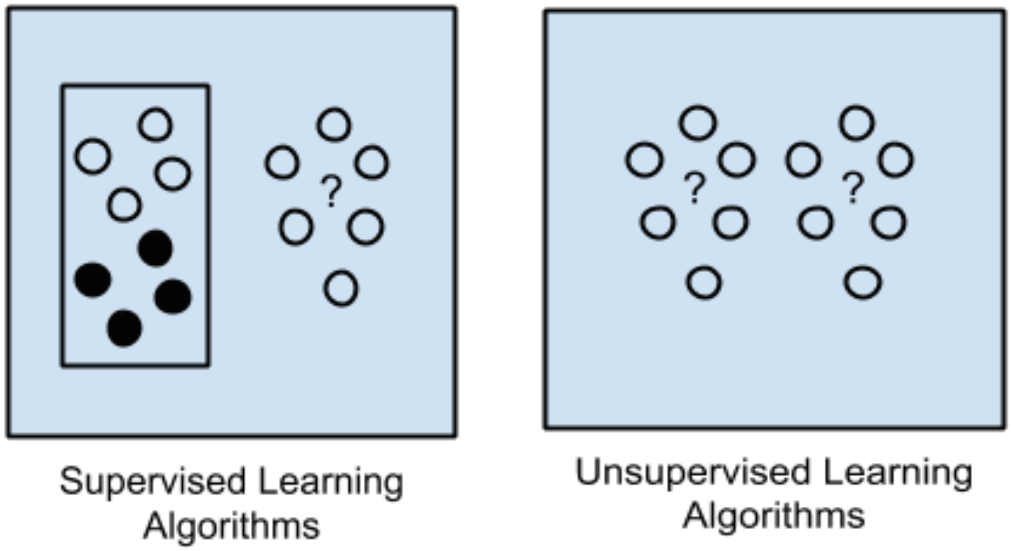

Figure A.3: Pictorial presentation of Supervised and Unsupervised algorithm

Contrastive Learning, Autoencoders, Generative Models etc. Trying to predict the subsequent frame in a video based on previously captured frames or the subsequent word in a text based on previously captured words are examples of self-supervised learning.

Reinforcement learning

Nowadays, the majority of research is focused on reinforcement learning. In this technique(e.g. algorithm that interacts with the environment and makes decisions) through specific behaviors and feedback in the form of rewards or punishments, an agent learns appropriate behavior in a given environment. Over time, the agent aims to maximize its commutative rewards the agent seeks to maximize its cumulative reward over time. Reinforcement learning is learning from interactions with an environment without explicit supervision. This technique has been applied in the areas of game playing, robotics, autonomous vehicles, and natural language processing.

A.1.3 Understanding how deep learning works

Deep learning is a specific type of machine learning that emphasizes learning successive layers of more significant representations and from that it tries to extract representations from the input data. Here the word 'deep' refers to the depth of the model that is how many layers contribute to a model. The modern-day deep learning network can even consist of hundreds of layers. This design is implemented by using a neural

network model that is stacked one upon another. Fig. A.4 shows how a deep neural network is transforming an image of a digit to recognize what digit it is using multiple layers. The network converts the image of the digits into representations that deviate from the original image more and more as it moves through the layers. There are three main steps to understand how the deep learning mechanism works step by step (refer Fig. A.5).

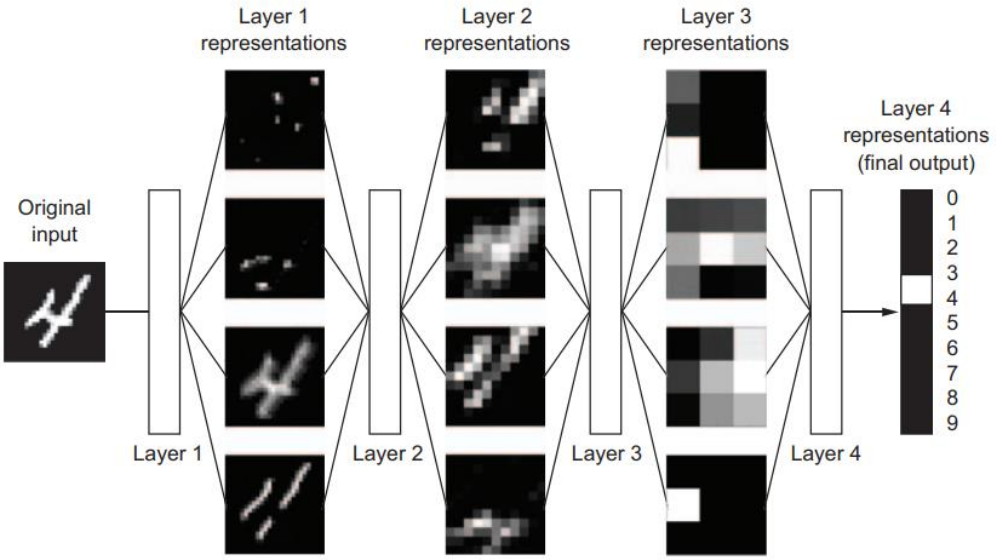

Figure A.4: A deep neural network for digit classification (Source: Deep Learning with Python by François Chollet)

Initially, some random weights are assigned to each layer. It is also known as the parameters of the layer. In forward propagation, input data is given to the network, and computation continues layer by layer up to the output layer. Based on the weights each layer transforms the input values. Finding a set of values for each layer's weight in a network so that it can accurately map sample inputs to their corresponding target is referred to as learning in this scenario. However, the challenge lies in the fact that a deep neural network may handle millions of parameters. It may seem impossible to find the right value for every one of them, especially when changing the value of one parameter can change the nature of all others.

In the next step, it is very important to control the predicted output from the neural network. The output of the network is compared to the actual output, and the difference is measured using a loss function. This difference in value indicates how well the network performed in this particular example based on the values of different parameters set in different layers.

Finally, using this difference or loss score as a feedback signal deep learning network slightly modifies the weight values in a way that lowers the loss score for the current

example. This is done by an optimization algorithm. As it involves propagating the error backward through the network, it is also known as backpropagation algorithm.

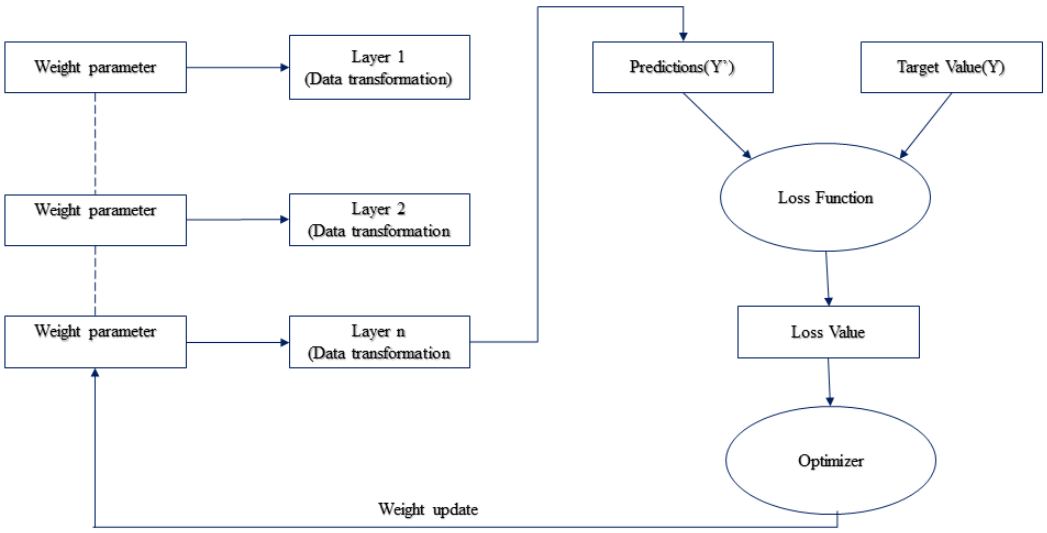

Figure A.5: A deep neural working mechanism

References

[1] L. Seoud, T. Hurtut, J. Chelbi, F. Cheriet, and J. P. Langlois, "Red lesion detection using dynamic shape features for diabetic retinopathy screening," *IEEE transactions on medical imaging*, vol. 35, no. 4, pp. 1116–1126, 2015.

[2] J. E. Shaw, R. A. Sicree, and P. Z. Zimmet, "Global estimates of the prevalence of diabetes for 2010 and 2030," *Diabetes research and clinical practice*, vol. 87, no. 1, pp. 4–14, 2010.

[3] R. Anjana, R. Pradeepa, M. Deepa, M. Datta, V. Sudha, R. Unnikrishnan, A. Bhansali, S. Joshi, P. Joshi, C. Yajnik *et al.*, "Prevalence of diabetes and prediabetes (impaired fasting glucose and/or impaired glucose tolerance) in urban and rural india: Phase i results of the indian council of medical research–india diabetes (icmr–indiab) study," *Diabetologia*, vol. 54, pp. 3022–3027, 2011.

[4] R. Raman, P. K. Rani, S. R. Rachepalle, P. Gnanamoorthy, S. Uthra, G. Kumaramanickavel, and T. Sharma, "Prevalence of diabetic retinopathy in india: Sankara nethralaya diabetic retinopathy epidemiology and molecular genetics study report 2," *Ophthalmology*, vol. 116, no. 2, pp. 311–318, 2009.

[5] R. Raman, S. Srinivasan, S. Virmani, S. Sivaprasad, C. Rao, and R. Rajalakshmi, "Fundus photograph-based deep learning algorithms in detecting diabetic retinopathy," *Eye*, vol. 33, no. 1, pp. 97–109, 2019.

[6] E. Rahimy, "Deep learning applications in ophthalmology," *Current opinion in ophthalmology*, vol. 29, no. 3, pp. 254–260, 2018.

[7] K. Solanki, C. Ramachandra, S. Bhat, M. Bhaskaranand, M. G. Nittala, and S. R. Sadda, "Eyeart: automated, high-throughput, image analysis for diabetic retinopathy screening," *Investigative Ophthalmology & Visual Science*, vol. 56, no. 7, pp. 1429–1429, 2015.

[8] A. Tufail, C. Rudisill, C. Egan, V. V. Kapetanakis, S. Salas-Vega, C. G. Owen, A. Lee, V. Louw, J. Anderson, G. Liew *et al.*, "Automated diabetic retinopathy image assessment software: diagnostic accuracy and cost-effectiveness compared with human graders," *Ophthalmology*, vol. 124, no. 3, pp. 343–351, 2017.

[9] D. S. W. Ting, C. Y.-L. Cheung, G. Lim, G. S. W. Tan, N. D. Quang, A. Gan, H. Hamzah, R. Garcia-Franco, I. Y. San Yeo, S. Y. Lee *et al.*, "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes," *Jama*, vol. 318, no. 22, pp. 2211–2223, 2017.

[10] R. Paisan, J. Krause, C. Peranut, S. Rory, R. Rajiv, W. Kasumi, P. Sonia, H. Kornwipa, T. Mongkol, S.-A. Sukhum *et al.*, "Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program," *NPJ Digital Medicine*, vol. 2, no. 1, 2019.

[11] V. Gulshan, R. P. Rajan, K. Widner, D. Wu, P. Wubbels, T. Rhodes, K. Whitehouse, M. Coram, G. Corrado, K. Ramasamy *et al.*, "Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in india," *JAMA ophthalmology*, vol. 137, no. 9, pp. 987–993, 2019.

[12] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2018, pp. 3–11.

[13] T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kälviäinen, and J. Pietilä, "The diarectdb1 diabetic retinopathy database and evaluation protocol." in *BMVC*, vol. 1, no. 1. Citeseer, 2007, p. 10.

[14] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, "Dream: diabetic retinopathy analysis using machine learning," *IEEE journal of biomedical and health informatics*, vol. 18, no. 5, pp. 1717–1728, 2013.

[15] C. Angermueller, T. Pärnamaa, L. Parts, and O. Stegle, "Deep learning for computational biology," *Molecular systems biology*, vol. 12, no. 7, p. 878, 2016.

[16] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *jama*, vol. 316, no. 22, pp. 2402–2410, 2016.

[17] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, 2001, pp. I–I.

[18] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1. Ieee, 2005, pp. 886–893.

[19] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 9, pp. 1627–1645, 2009.

[20] C. Ying, M. Qi-Guang, L. Jia-Chen, and G. Lin, "Advance and prospects of adaboost algorithm," *Acta Automatica Sinica*, vol. 39, no. 6, pp. 745–758, 2013.

[21] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *2008 IEEE conference on computer vision and pattern recognition*. Ieee, 2008, pp. 1–8.

[22] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21–37.

[23] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of yolo algorithm developments," *Procedia Computer Science*, vol. 199, pp. 1066–1073, 2022.

[24] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[25] A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson, and P. F. Sharp, "Automated microaneurysm detection using local contrast normalization and local vessel detection," *IEEE transactions on medical imaging*, vol. 25, no. 9, pp. 1223–1232, 2006.

[26] T. Walter, P. Massin, A. Erginay, R. Ordonez, C. Jeulin, and J.-C. Klein, "Automatic detection of microaneurysms in color fundus images," *Medical image analysis*, vol. 11, no. 6, pp. 555–566, 2007.

[27] J. P. Bae, K. G. Kim, H. C. Kang, C. B. Jeong, K. H. Park, and J.-M. Hwang, "A study on hemorrhage detection using hybrid method in fundus images," *Journal of digital imaging*, vol. 24, no. 3, pp. 394–404, 2011.

[28] L. Giancardo, F. Mériaudeau, T. P. Karnowski, K. W. Tobin, Y. Li, and E. Chaum, "Microaneurysms detection with the radon cliff operator in retinal fundus images," in *Medical Imaging 2010: Image Processing*, vol. 7623. International Society for Optics and Photonics, 2010, p. 76230U.

[29] G. Quellec, M. Lamard, P. M. Josselin, G. Cazuguel, B. Cochener, and C. Roux, "Optimal wavelet transform for the detection of microaneurysms in retina photographs," *IEEE transactions on medical imaging*, vol. 27, no. 9, pp. 1230–1241, 2008.

[30] K. M. Adal, P. G. Van Etten, J. P. Martinez, K. W. Rouwen, K. A. Vermeer, and L. J. van Vliet, "An automated system for the detection and classification of retinal changes due to red lesions in longitudinal fundus images," *IEEE transactions on biomedical engineering*, vol. 65, no. 6, pp. 1382–1390, 2017.

[31] I. Lazar and A. Hajdu, "Retinal microaneurysm detection through local rotating cross-section profile analysis," *IEEE transactions on medical imaging*, vol. 32, no. 2, pp. 400–407, 2012.

[32] S. S. Rahim, C. Jayne, V. Palade, and J. Shuttleworth, "Automatic detection of microaneurysms in colour fundus images for diabetic retinopathy screening," *Neural computing and applications*, vol. 27, no. 5, pp. 1149–1164, 2016.

[33] S. S. Kar and S. P. Maity, "Automatic detection of retinal lesions for screening of diabetic retinopathy," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 3, pp. 608–618, 2017.

[34] W. Zhou, C. Wu, D. Chen, Y. Yi, and W. Du, "Automatic microaneurysm detection using the sparse principal component analysis-based unsupervised classification method," *IEEE access*, vol. 5, pp. 2563–2572, 2017.

[35] S. Wang, H. L. Tang, Y. Hu, S. Sanei, G. M. Saleh, T. Peto *et al.*, "Localizing microaneurysms in fundus images through singular spectrum analysis," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 5, pp. 990–1002, 2016.

[36] M. J. Van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sánchez, "Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1273–1284, 2016.

[37] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.

[38] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[39] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[40] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[41] X. Ou, P. Yan, Y. Zhang, B. Tu, G. Zhang, J. Wu, and W. Li, "Moving object detection method via resnet-18 with encoder–decoder structure in complex scenes," *IEEE Access*, vol. 7, pp. 108 152–108 160, 2019.

[42] "Feedback on a publicly distributed database: the messidor database," vol. 33. [Online]. Available: http://www.ias-iss.org/ojs/IAS/article/view/1155

[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[44] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[45] R. Kälviäinen and H. Uusitalo, "Diaretdb1 diabetic retinopathy database and evaluation protocol," in *Medical image understanding and analysis*, vol. 2007. Citeseer, 2007, p. 61.

[46] X. Zhang, G. Thibault, E. Decencière, B. Marcotegui, B. Laÿ, R. Danno, G. Cazuguel, G. Quellec, M. Lamard, P. Massin *et al.*, "Exudate detection in color retinal images for mass screening of diabetic retinopathy," *Medical image analysis*, vol. 18, no. 7, pp. 1026–1043, 2014.

[47] C. Playout, R. Duval, and F. Cheriet, "A novel weakly supervised multitask architecture for retinal lesions segmentation on fundus images," *IEEE transactions on medical imaging*, vol. 38, no. 10, pp. 2434–2444, 2019.

Publications from this Thesis

Journals

- **Goutam Kumar Ghorai**, Swagata Kundu, Gautam Sarkar, Ashis kumar Dhara, "Multi-Scale Feature Pyramid for Red Lesions in Fundus Images", International Journal of Recent Technology and Engineering (IJRTE),Volume-12 Issue-4, November 2023

- Swagata Kundu, Vikrant Karale, **Goutam Kumar Ghorai** Gautam Sarkar, Sambuddha Ghosh, Ashis Kumar Dhara, "Nested U Net for Segmentation of Red Lesions in Retinal Fundus Images and Sub Image Classification for removal of False Positives", Journal of Digital Imaging 35, no. 5 (2022): 1111-1119.

- Sandip Sadhukhan, Arpita Sarkar, Debprasad Sinha, **Goutam Kumar Ghorai**, Gautam Sarkar and Ashis Kumar Dhara "Attention Based Fully Convolutional Neural Network for Simultaneous Detection and Segmentation of Optic Disc in Retinal Fundus Images", International Journal of Medical and Health Science Vol. 14 No: 8, pp. 200-204, 2020.

- Sandip Sadhukhan, **Goutam Kumar Ghorai**, Debprasad Sinha, Souvik Maiti, Gautam Sarkar and Ashis Kumar Dhara "Segmentation of optic disc in retinal images using fully convolutional network" Journal of Current Indian Eye Research Vol. 6, Issue 2, December 2019, pp. 40-47.

Conferences

- Sandip Sadhukhan, **Goutam Kumar Ghorai**, Souvik Maiti, Gautam Sarkar and Ashis Kumar Dhara, "Optic Disc Localization in Retinal Fundus Images using Faster R-CNN" Proceedings of Fifth International Conference on Emerging

Applications of Information Technology (EAIT 2018), January 12-13, Kolkata, India.

- Sandip Sadhukhan, **Goutam Kumar Ghorai**, Souvik Maiti, Vikrant Anilrao Karale, Gautam Sarkar and Ashis Kumar Dhara, "Optic Disc Segmentation in Retinal Fundus Images Using Fully Convolutional Network and Removal of False- Positives Based on Shape Features" Proceedings of 4th International Workshop, DLMIA 2018 and 8th Internation Workshop, ML-CDS 2018, September 20, 2018, Granada, Spain.

- Sandip Sadhukhan, **Goutam Kumar Ghorai**, Souvik Maiti, Debiprasad Sinha, Gautam Sarkar and Ashis Kumar Dhara "Fully Convolutional Network for Segmentation of Optic Disc In Retinal Fundus Images" Proceedings of IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) April 8-11, 2019, Venice, Italy.

- Debasis Maji, Souvik Maiti, Moyenak Biswas, **Goutam Kumar Ghorai**, Sandip Sadhukhan, Debprasad Sinha, Ashis Kumar Dhara, Gautam Sarkar, "Automatic Patch Based Tortuosity Retinal Vessel Classification using VGG16 Network". IEI Impact in Changing Energy Mix in the Power Sector, ISBN:978-81-942561-2-0, 2019, Kolkata, India.

Author's Biography

Goutam Kumar Ghorai is a Ph.D candidate from the Department of Electrical Engineering,at Jadavpur University, WB, India. He completed B.Tech and M.tech degrees from the University of Calcutta. He also completed B.Scs degree in Physics honours degree from Vidyasagar University. He is currently working as an Assistant Professor of the Electrical Engineering Department at Ghanikhan Choudhury Institute of Engineering and Technology,Malda, West Bengal. He developed a Modular pattern course curriculum at B.tech Degree level. He served the Institute as Deputy register in charge, Departmental HoD, and was member and chairman of various committees of the Institute. He also serves the institute as Fire Safety and security officer. He was the training coordinator for Nirman Sahayak training conducted by Govt. of West Bengal. His teaching experience is about 17 years as a faculty of the Electrical Engineering Department. He teaches the student various subjects like, Electrical machines, control systems, Electrical Circuits theory, Electrical Design, etc.

Contact Information

Email :goutamghorai@rediffmail.com

Research Interests:

Machine Learning and Medical Image Analysis, Electrical vehicles, Control system.

Goutam Kumar Ghorai
14/02/2024

14/02/2024

Prof. Gautam Sarkar
Electrical Engineering Department
Jadavpur University
Kolkata, INDIA

14/02/24

Dr. Ashis Kumar Dhara
Assistant Professor
Electrical Engineering Department
National Institute of Technology
DURGAPUR - 713209

1.0