SEGMENTATION OF OPTIC DISC, EXUDATES AND MACULA IN RETINAL FUNDUS IMAGES

Thesis submitted by

Souvik Maiti

Doctor of Philosophy (Engineering)

Department of Electrical Engineering
Faculty Council of Engineering and Technology
Jadavpur University
Kolkata, India
2023

Index No.: 50/16/E

1. **Title of the Thesis:**

   SEGMENTATION OF OPTIC DISC, EXUDATES AND MACULA 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, India

3. **List of Publications:**

   Journal Published

- **Souvik Maiti**, Debasis Maji, Ashis Kumar Dhara, and Gautam Sarkar, "Automatic detection and segmentation of optic disc using a modified convolution network," Biomedical Signal Processing and Control, vol. 76, p. 103633, 2022.
- **Souvik Maiti**, Debasis Maji, Ashis Kumar Dhara, and Gautam Sarkar, "An Attention Enriched Encoder-Decoder Architecture with CLSTM and RES Unit for Segmenting Exudate in Retinal Images," Signal, Image and Video Processing, pp. 1–11, 2024.

   National/ International Conferences:

- **Souvik Maiti**, Debasis Maji, Ashis Kumar Dhara, and Gautam Sarkar, "Spatial Attention Enhanced Network for Segmentation of Exudate," in 2022 IEEE Calcutta Conference (CALCON), Kolkata, West Bengal, India, 2022, pp. 93–97.
- **Souvik Maiti**, Debasis Maji, Ashis Kumar Dhara, and Gautam Sarkar, "Channel Attention Enhanced Deep Network for Segmenting Exudate," in 2022 IEEE 6th International Conference on Condition Assessment Techniques in Electrical Systems (CATCON), Durgapur, West Bengal, India, 2022, pp. 94–98.

- **Souvik Maiti**, Debasis Maji, Ashis Kumar Dhara, and Gautam Sarkar, "Automated Segmentation of Macula in Retinal Images Using Deep Learning Methodology", in Springer Lecture Notes in Electrical Engineering, International Conference on Emerging Electronics and Automation, Silchar, Assam, India, 2022, pp. 201–213.

4. **List of Patents:** Nil

Statement of Originality

I, Souvik Maiti, registered on 24th November, 2016, do hereby declare that this thesis entitled "Segmentation of Optic Disc, Exudates and Macula 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 3%.

Souvik Maiti

Signature of Candidate

Date: 20/11/2023

Certified by Supervisors: (Signature with date and seal)

20/11/2023

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

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

Ashis 20/11/23

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

CERTIFICATE

This is to certify that the thesis entitled "Segmentation of Optic Disc, Exudates and Macula in Retinal Fundus Images" submitted by Sri. Souvik Maiti, who got his name registered on 24th November, 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.

20/11/2023

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

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

20/11/23

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

Dedicated to

My Parents, Sister and Grandmother

ACKNOWLEDGEMENT

During the journey in obtaining my Ph.D., several domains were explored which might not be possible without the support and encouragement of numerous persons. This thesis is the end of my journey in obtaining my Ph.D. which has been kept on track and been seen through to completion with the kind support of my family, friends, teachers, colleagues and well wishers. It would not have been possible to write this doctoral thesis without the help and support of these kind people around me. It is a pleasant task to express my thanks to all those who contributed in many ways to the success of this study and made it a memorable experience for me. At the end of this thesis, I would like to thank all those people who made this thesis possible.

First, I would like to thank my supervisors Dr. Gautam Sarkar and Dr. Ashis Kumar Dhara for their advice, guidance and support. Under their guidance, I successfully overcame many difficulties and learned a lot. I am extremely indebted to them for providing the necessary infrastructure and resources to accomplish my research work. Their advice, reviews and suggestions around my work were extremely valuable during the period of research.

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.

For any errors or inadequacies that may remain in this work, of course, the responsibility is entirely my own.

Souvik Maiti

Souvik Maiti
Haldia
November, 2023

ABSTRACT

Diabetic retinopathy, a microvascular complexity prevalent in long-term diabetic patients, is a serious vision menacing condition that leads to blindness in working citizens worldwide. Prolonged high sugar level in the blood weakens and damages the vessels in the eye, blurring the vision. Moreover, diabetic retinopathy is painless and individuals do not experience any symptoms unless sight loss prevails, hence the fundus must be checked regularly to assure timely treatment. It is time-consuming and tedious to manually examine the retinal pictures. An automated screening method can become a popular tool for detecting the pathologies of diabetic retinopathy. The advancement in research works has developed several strategies to automatically identify the retinopathy lesions.

Numerous retinal anomalies manifest in the optic nerve head in the beginning of diabetic retinopathy, which, if identified in the initial phase, will aid ophthalmologists to treat patients effectively. Glaucoma monitoring approach considers the optic cup to optic disc ratio which necessitates the identification of optic disc. It is also vital to recognize the macula for assessing the seriousness of diabetic retinopathy. The consequences of the degeneration of macula perform a crucial role in the evolvement of visual deterioration. The development of microaneurysms, exudates and haemorrhages are the key symptoms of diabetic retinopathy. Therefore, exudate segmentation is considered to be an important task for early diagnosis and treatment. Since the pathologies of diabetic retinopathy are typically difficult to discern, finding the appropriate characteristics to automatically recognize the exudate in retinal pictures are critical. This study presents a computerized technique to aid experts in retinopathy screening services by indicating the lesions of diabetic retinopathy in fundus photograph.

The high efficacy of deep neural networks makes them a popular choice for evaluating medical images. This work presents encoder-decoder architectures which employ convolutional network to detect the exudate, optic disc and macula region automatically in fundus pictures. The convolutional neural networks are designed to learn distinctive features

and patterns associated with the various retinopathy pathologies, enabling precise segmentation. To overcome the limitations of the unavailability of significant amount of labelled training data, data augmentation techniques such as rotation, scaling, and flipping are applied during the training phase.

Extensive experiments are conducted on publicly available datasets to evaluate the performance of the proposed system. The results demonstrate its effectiveness in accurately detecting and segmenting the exudates in retinal fundus images. Furthermore, the robustness of the designed algorithms are confirmed by implementing them on a broad range of fundus images with varying image qualities acquired from several datasets like DRIVE, IDRiD, MESSIDOR, STARE, CHASE-DB1, DIARETDB0 and DIARETDB1. Performance indicators namely F1-score, sensitivity, accuracy and specificity are computed and compared to those of the existing techniques. This work outperforms the existing methods, showcasing its potential as a reliable tool for computer-aided diagnosis of diabetic retinopathy.

The proposed models provide promising approach for efficient and accurate segmentation of exudates in fundus photographs. The performance of the models surpasses the existing methods making them a valuable tool to assist an ophthalmologist in making decisions quickly. The experimental results provide insights into the system's ability to generalize well across different imaging conditions, enhancing its applicability in real-world clinical settings.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
Ac	Accuracy
CLSTM	Convolutional Long Short Term Memory
CNN	Convolutional Neural Network
DR	Diabetic Retinopathy
DC	Dice Coefficient
DRIVE	Digital Retinal Images for Vessel Extraction
FCN	Fully Convolutional Network
IDRiD	Indian Diabetic Retinopathy Image Dataset
LSTM	Long Short Term Memory
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
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
Se	Sensitivity
Sp	Specificity
DIARETDB0	Standard Diabetic Retinopathy Database Calibration Level 0
DIARETDB1	Standard Diabetic Retinopathy Database Calibration Level 1
SGD	Stochastic Gradient Descent
STARE	Structured Analysis of the Retina
VEGF-A	Vascular Endothelial Growth Factor-A

Chapter 1

Introduction

Diabetes is a condition which happens either due to inadequate pancreatic insulin secretion or due to the human body's ineffective utilization of insulin. The insulin hormone controls the sugar level in the blood. An increased sugar level over time causes substantial harm to the different organs in the body even the blood vessels. As stated by the International Diabetes Foundation Diabetes Atlas tenth edition 2021, 537 million individuals of age in the range of 20-79 years are affected with diabetes and by 2045, the count is predicted to grow to 783 million worldwide. According to the report, India had 74.2 million diabetic individuals in 2021, and by 2045, the number is predicted to reach to 124.9 million. It is unfortunate that approximately one out of two persons who have diabetes go undiagnosed. Diabetic individuals are more vulnerable to experience stroke and heart attack. The uncontrolled and undiagnosed diabetes leads to diabetic retinopathy which is a threat to vision impairment.

1.1 Diabetic Retinopathy

Diabetes has evolved to be a significant issue for public health in this modern era. Prolonged rise in glucose levels in the blood leads to diabetes. Eventually the diabetes causes complications in human eyes, referred as Diabetic Retinopathy (DR). The vessels within the retina are affected by the excessive blood glucose concentration. These compromised blood vessels begin to spill fluid and blood onto the retina, threatening the eyesight.

DR often endangers the vision by affecting the retinal vascular structures. In the initial stage, DR is asymptotic or manifests minor symptoms but eventually results in blindness if not diagnosed on time. Patients with type 1 and type 2 diabetics are susceptible to develop this disease. The microvasculature structure of the eye is impacted in this condition. With the passage of time, the excessive amount of blood sugar blocks the tiny vessels responsible for supplying nutrients to the retina. This results the eye to develop new blood vessels which often leaks due to their improper development. A regular screening of the eye fundus helps in identifying the morphological alterations, inflammations, vascular defects, internal bleeding and anomalies in the macula region. A computer-assisted technique can complete this task quickly and effectively.

1.2 Stages in Diabetic Retinopathy

DR in its initial phase does not show symptoms. However, gradually the patients claim trouble in reading or focusing on distant objects. The alterations in eye sight worsen with time and eventually result in irreversable blindness. The stages of DR are referred as

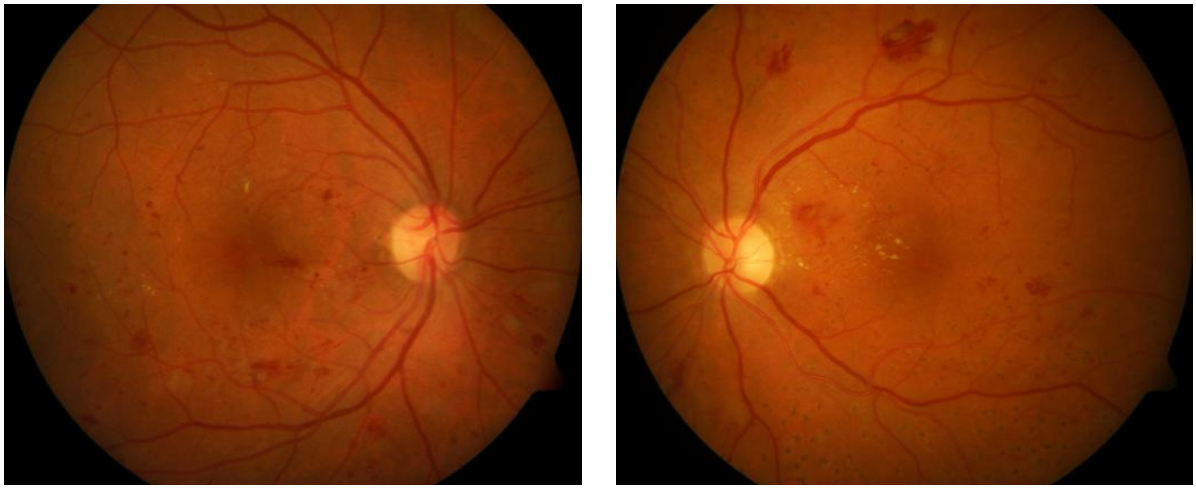
i. Non-Proliferative Diabetic Retinopathy (NPDR)
    (a) Mild
    (b) Moderate
    (c) Severe
ii. Proliferative Diabetic Retinopathy (PDR)

1.2.1 Non-Proliferative Diabetic Retinopathy

In this phase the walls of the blood vessel swells and even sometimes bulges. Fluid can leak from vessels into the retina, causing swelling. The diameter of the vessels in the retina starts to change and becomes thinner. There is no new emergence of additional blood vessel. Sometimes the retinal macula area swells giving rise to a condition termed as macular edema. The NPDR has been typically categorized into three variants, such as Mild NPDR, Moderate NPDR and Severe NPDR.

1.2.1.1 Mild NPDR

DR is now at its most basic level. During this stage, red coloured round shaped minute blood-filled lumps of size ranging from 25 µm to 100 µm in diameter [1] are found in the retinal artery walls. These structures are termed as microaneurysms. They might leak and cause fluid to seep into the retina leaving eyesight to become blurry. The chance of DR severity increases as the quantity of microaneurysms in the retina rises. A mild NPDR retinal photograph is exhibited in fig. 1.1 (a).



(a) (b)

Fig. 1.1. (a) A mild NPDR and (b) a moderate NPDR retinal photograph

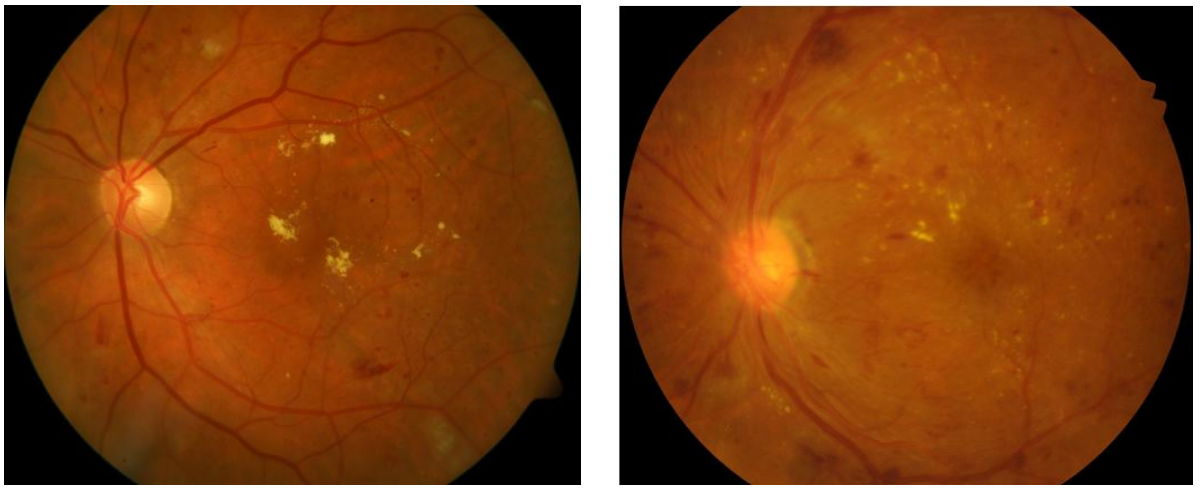
1.2.1.2 Moderate NPDR

With time, the retina's vessels swell and their capacity to transport blood decrease. The retina is therefore deprived of its essential nutrition. When the microaneurysms burst or the vessel wall leaks, blood accumulates on the retina giving rise to haemorrhages. As a result the retina's appearance may alter. There are two forms of haemorrhages: dot-blot haemorrhages and flame haemorrhages. Dot-blot haemorrhages, which are circular in appearance, develop at different locations in the retina, especially near the capillary's venous end. Flame haemorrhages that emerges from the pre-capillary arterioles found in the retina's inner lining,

appear at the nerve fibers. Moderate NPDR if left untreated, may result in macular edema, the most well-known cause of visual loss in diabetics. This condition develops as the fluid from the ruptured vessels seeps into the macula region. As a result, the macula swells and blurs the vision. Macular edema results in the lifelong loss of vision, if not properly diagnosed. A moderate NPDR retinal photograph is depicted in fig. 1.1 (b).

1.2.1.3 Severe NPDR

During this stage, the retina's blood circulation is compromised because of the clogging in the vessels. In the extreme condition of NPDR acute intra-retinal micro vascular alterations, venous beading, leakage of protein and lipid occur. The protein and lipid that seeps from the ruptured vessels result in the development of bright yellowish irregular shaped areas termed as exudates. The exudates usually emerge as clusters, small spots, singular strips, or ringed structures surrounding the leaky capillaries, macular edema, or microaneurysms. The exudates also manifest in the retina as little patches of yellowish white coloured cotton wool structures and are termed as soft exudates. These structures develop as the retina's surface layer nerve fibers swell. The presence of retinal haemorrhages, venous beading in minimum two quadrants, microaneurysms in all the four quadrants, or intra-retinal microvascular abnormalities in minimum one quadrant [2] indicate the progression of a patient's condition towards PDR. A severe NPDR retinal picture is depicted in fig. 1.2 (a).



(a) (b)

Fig. 1.2. (a) A severe NPDR and (b) a PDR retinal photograph

1.2.2 Proliferative Diabetic Retinopathy

This phase is a more severe condition of DR. Neovascularization, a phenomenon of the formation of new additional retinal vessels, occurs in this stage. The fibrovascular growth enlarges with time. The new, abnormal blood vessels, after leaving the retina, spread in the vitreous humour which is a gel-like material found in between the retina and eye lens. Age-related vitreous expansion strains these abnormal vessels and even causes them to rupture, oozing fluid into the vitreous and an unanticipated visual impairment. If it leaks a bit, dark floaters appear whereas large amount of discharge impairs eyesight. In extreme situation, the retina might get separated leading to lifelong blindness. The eyeball may experience pressure if the newly formed vessels hinder the normal fluid flow in the eye. Cataracts may develop if the optic nerves that convey information from the eyes into the brain get damaged. The macula might seem elevated, which can be a sign of retinal traction detachment occurring because of the splitting of the retinal pigment epithelium from the neurosensory retina. A PDR retinal photograph is depicted in fig. 1.2 (b).

1.3 Need for Digital Image Processing

With the growing patient population and the limited staffs in hospitals, an automated detection approach is very essential. The human eye wears out while assessing lot of patients. Hence there remains a possibility of missing important information in the report. Moreover, several individuals from the rural area who participate in initial phase of eye screening programme do not return to receive the report. Eventually the patients with DR could not even be informed the disease. Thus, the development of an automated technique is crucial for producing accurate reports rapidly so that the patients can be updated about their condition in a short time. The implementation of machine learning and digital imaging techniques are expanding rapidly in various domains of research. In biomedical sector significant advancements have been attained using computerized techniques.

1.4 Fundus Camera

The fundus camera comprises of a specially built low-powered microscope along with a mounted camera. It is used to capture images of the internal eye, which involves the macula, optic disc (OD), retina, retinal vessels, posterior pole etc. There are two kinds of fundus cameras, the Mydriatic Fundus Camera (involve dilating the pupils) and the Non-Mydriatic Fundus Camera (does not involve dilating the pupils). A camera is characterized based on the lens's angle of view. A wide angle camera has an angle of view ranging from 45° and 140°, but their magnification power is comparatively lower. A narrow angle camera has a limited angle of view of 20° or less. A 30° angle of view of is regarded as the standard acceptance angle, which produces an image two and a half times bigger than actual size. A fundus camera used for capturing the image of the retina is exhibited in fig. 1.3.

Fig. 1.3. A fundus camera

1.5 Anatomy of a Human Eye

The wall of an eye, which is a spherical shaped hollow structure, comprises of three layers namely the outer sclera, the middle choroid, and the inner retina. The internal area is packed with liquid that helps to preserve the shape of the eye. A human eye's anatomy is depicted in

fig. 1.4. Four out of every five information which the brain perceives derives from the eyes. This demonstrates how crucial they are. The elliptical shaped clear outer portion encasing the pupil, iris, anterior chamber and that aids in focusing incoming rays of light is known as the cornea. In between the cornea and the lens, there exists a fluid called aqueous humour which helps in providing oxygen and nutrients to the organs.

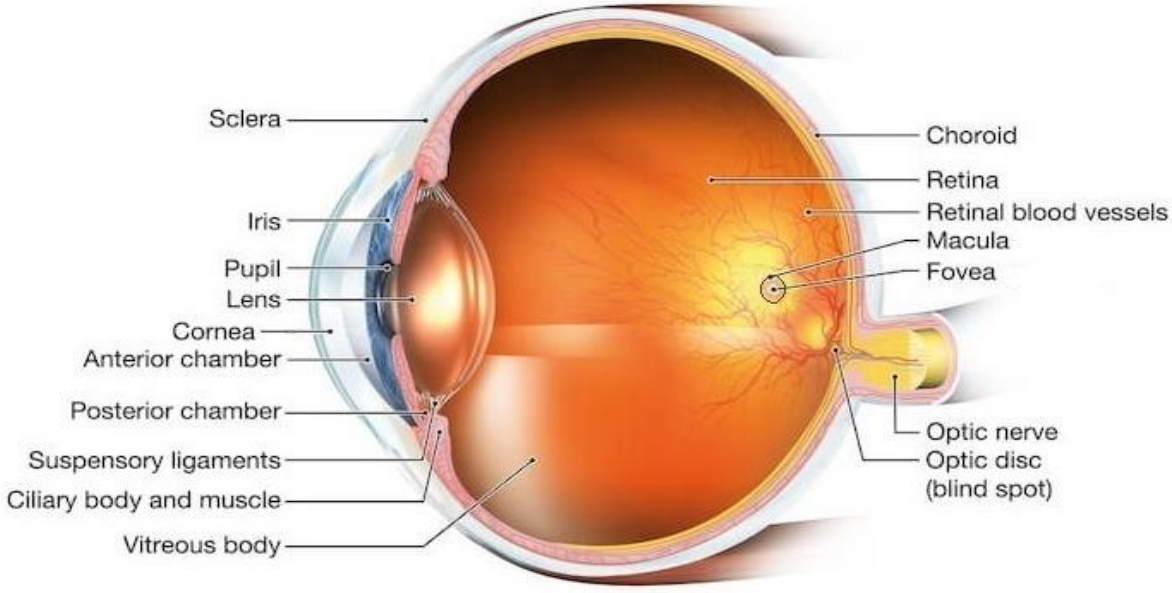


Fig. 1.4. The anatomical structure of a human eye

The iris, a round, pigmented tissue containing the pupil in the center, controls how much light will enter the eye. The role of iris is to regulate the pupil's size based on the illumination intensity. The iris's sphincter and dilator muscles regulate the size of the pupil. As the tissues absorb the majority of the incident light, the pupil looks dark in colour. A biconvex, clear lens together with the cornea, aids in refracting the light, so as to focus them on the retina. The eyeball is filled with a colorless fluid between the retina and the lens called vitreous humour. 95% of vitreous humour fluid is water.

The opaque, white structure, called the sclera serves as the outer, protective layer of the eye. It is attached to six small muscles surrounding the eye, which regulate eye movement. The optic nerve that is coupled with sclera in the backside of the eye, gets into the retina at the optic disc, which is also addressed as optic nerve head or blind spot, since no light-sensitive cone cells or rod cells are present here to sense the light stimulation.

The inner rear portion of the eye is lined by a thin sheet of nerve cells termed as the retina. It is highly sensitive to light and is responsible for transmitting the received image signals to the brain. The cone cells and the rod cells are the two categories of light receptor cells present in the retina. A complicated network of around 1.2 million neurons participates in linking the light receptors with the brain. The rod cells which produce a black and white response enable us to view at night. The cone cells only react to bright light and are capable of sensing colour. The primary retinal artery and the subsidiary of the choriocapillaris are the blood supplying channels in the retina. The interior two-thirds of the retina are supplied by the branches of the primary retinal artery, whereas the exterior one-third of the retina is fed by the choriocapillaris.

In the backside of the human eye, nearby the middle of the retina, there is a round shaped small zone containing a substantial number of light sensing cells termed as the macula. The macula is essential to perceive the details of items viewed. The macula's core is referred as the fovea. The fovea that includes the most tightly packed photoreceptor cells, accounts in making the vision the sharpest. The fovea lacks blood vessels, and thus facilitates light to reach the foveal cone mosaic unhindered.

1.6 Symptoms of Diabetic Retinopathy

DR is asymptotic in the initial phase, but with time the following symptoms appear.

i. Blurring of vision
ii. Impairment in night vision
iii. Deterioration in colour vision
iv. Development of blind spots
v. A sudden loss in eyesight.
vi. Emergence of floaters which appear as small specks, circles, dots in the field of vision

1.7 Risk Factors of Diabetic Retinopathy

Individuals who have diabetes are susceptible to retinopathy. With other chronic conditions, certain risk factors remain beyond the control although few might be handled with

alterations in lifestyle and medical assistance. The DR risk factors may therefore be categorized as controllable and non-modifiable type.

1.7.1 Controllable Risk Factors

i. Poor blood sugar management: Researches have found a link between poor blood sugar level management and the development and advancement of DR.
ii. Increased blood pressure level: If an individual's blood pressure is not correctly regulated, DR is more likely to develop. DR advances more rapidly in individuals suffering from diabetes over ten years and having increased blood pressure level.
iii. Obesity: Person suffering from diabetes and having high body mass index are two times more likely to develop DR.
iv. Pregnancy: If the blood sugar level is not well managed throughout pregnancy, the DR can get worse. Specifically, the deregulation of glycaemia during conception, pregnancy and even in the postpartum phase is responsible for the fast advancement of DR.
v. Illness: Proteinuria, a disorder where excessive quantity of protein is found urine, is an important clue of the occurrence of DR. The increased levels of urea and creatinine are another indication. DR is more likely to appear in persons with underlying health issues like diabetic kidney disorder, heart disease, and excessive cholesterol.

1.7.2 Non-Modifiable Risk Factors

i. Genetic variation: A protein named Vascular Endothelial Growth Factor-A (VEGF-A) is responsible for the the improper development of new vessels in DR. A specific gene, termed VEGF-A gene conveys instructions to this protein. Medical experts have identified a discrepancy in this gene's sequence is connected to the onset of severe DR.
ii. Growing age: DR is rare in children aged 10 and less. Teenagers between the ages of 15 and 19 who have DR make up around 10% of the population. In the age range of 20 and 29, the percentage increases from 10% to 40%. When they reach the age of 30, approximately 60% of diabetic patients develop DR, and as they turn 45 years, the percentage increases to 70%.

iii. Ethnicity: Certain ethnic communities are more prone to get diabetic retinopathy. Experts have not been able to provide a convincing explanation for this disproportionately high prevalence of illness across different ethnicities.

iv. Duration of diabetic: The chance of developing DR in individuals with diabetes increases with time. The severity of the deasease progresses in proportion to the duration an individual is suffering from diabetes.

v. Stage of diabetic: Cardiovascular health risks and hypertension can have an impact on the beginning of DR. The probability of getting affected with DR becomes higher with the stage of diabetic and might endanger vision.

vi. Gender: A male individual is more likely to suffer from severe retinal impairment and DR. Researchers are still looking for an explanation for this gender disparity.

1.8 Medical Treatments for Diabetic Retinopathy

DR, when detected in its initial phase, has various efficient therapies that can slow down the advancement of DR and assist in preventing serious visual impairment. Moreover, researchers are trying to formulate new medicines to diagnose DR. The frequently utilized method of treating DR is the laser photocoagulation therapy. In this technique a laser beam in conjunction with optical devices and lenses, is focused on the retina to cure the damaged areas. Short laser bursts might be utilized to repair the injuries in the retina, seal ruptured vessels, demolish abnormal cells, and disrupt weak blood vessels.

1.9 Diabetic Retinopathy Datasets

There are numerous publicly accessible databases that offer retina related photographs, together with annotations of the various pathologies and information regarding the degree of DR severity. The databases utilized in study are described below:

i. DRIVE Dataset [3]

A screening programme for DR organized in the Netherlands provided the Digital Retinal Images for Vessel Extraction (DRIVE) dataset. A Canon CR5 non-mydriatic

3CCD camera with an angle of view of 45 degree was used to capture each image. Every picture is 8 bits per colour plane with dimension 768×584 pixels. There are 40 total pictures in the database along with a mask image corresponding to each fundus picture.

ii. MESSIDOR Dataset [4]

The non-mydriatic camera, Topcon TRC NW6 was used for collecting the fundus images in Methods to Evaluate Segmentation and Indexing Techniques in the Field of Retinal Ophthalmology (MESSIDOR) dataset. The pictures were captured with an angle of view of 45 degree. Every picture is 8 bits per colour plane with dimensions 1440×960, 2240×1488, and 2304×1536 pixels. The database consist 1200 pictures in 3 sets, each of which is subdivided into four smaller sets, each of which has 100 photos. An excel file containing the information regarding each photograph has been supplied with each subgroup.

iii. STARE Dataset [5]

The STructured Analysis of the Retina (STARE) database consist of 400 fundus photograhs of dimention 700×605 pixels. The pictures were captured with an angle of view of 35 degree. The database also includes 80 labeled photographs for OD segmentation and 40 labeled photographs for blood vessel segmentation.

iv. DIARETDB0 Dataset [6]

The DIARETDB0 database comprises of 130 fundus photos. Among them 20 photos are of normal eye and the remaining 110 photos include pathologies like soft exudates, hard exudates, hemorrhages, micronaneuyrysms and neovascularization. The pictures were captured with an angle of view of 50 degree. The dimension of every photograph present in the database is 1500×1152 pixels.

v. DIARETDB1 Dataset [7]

The DIARETDB1 database comprises of 89 fundus photos. Among them 5 photos are of normal eye and the remaining 84 photos include mild non-proliferative symptoms (Microaneurysms). The pictures were captured with an angle of view of 50 degree. The dimension of every photograph present in the database is 1500×1152 pixels.

vi. CHASE-DB1 Dataset [8]

This dataset, which involved 14 kids, comprises of 28 photographs of dimension of 999 by 960 pixels. The photographs that were obtained exhibit poor contrast and illumination challenges. The experts independently annotated every image of the database.

vii. IDRiD Dataset [9]

The Indian Diabetic Retinopathy Image Dataset (IDRiD) database comprises of 81 fundus photos of dimension 4288×2848 pixels. Both, photographs of normal fundus as well as that containing lesions are included in this database, which was created for the DR grading and segmentation competition. In each image the lesions have been labelled by specialists.

1.10 Convolutional Neural Network

A Convolutional Neural Network (CNN) is a prominent machine learning technique which is capable of accepting an image as input, provides significance (biases and weights) to numerous attributes and objects in the picture and has the ability to distinguish between them. A CNN necessitates considerably fewer pre-processing against other recognition techniques. It is capable of identifying the temporal and spatial relationships in an image effectively by the using appropriate filters. The specific neurons react to stimuli within a constrained area, referred as the Receptive Field. Several instance of these fields combine together to comprise an overall work area. The layers of CNN architecture primarily include the convolutional layer, the pooling layer and the fully connected layer along with suitable activating procedures.

i. Convolutional layer

The determination of the input picture's characteristics is the main aim of the convolution operation. A CNN can have more than one convolutional layer. Conventionally, the low-level characteristics especially colour, edges, etc. are tracked by the early convolutional layers. The network gradually captures the high-level

characteristics with more layers. A convolutional layer's channel count equates to the kernel's count

ii. Pooling layer

The convolved feature's size is lowered by the pooling layer. Thus the amount of computing power necessary for processing the information gets decreased. Moreover, it assists in appropriately training the network by facilitating the acquisition of preeminent attributes which are positional and rotational invariant. The pooling operation reduces the image's width and height, while the channel count remains unaltered.

The two frequently employed pooling strategies are the average and max pooling. The average of all data over the kernel's coverage region is conveyed by the average pooling method. The highest score of all data over the kernel's coverage region is conveyed by max pooling method.

The max pooling method completely mitigates the noisy activations along with the dimensionality reduction operation. The average pooling, on the other hand, only carries out the dimensionality reduction operation. Hence, the max pooling strategy outperforms the average pooling strategy significantly.

iii. Fully Connected layer

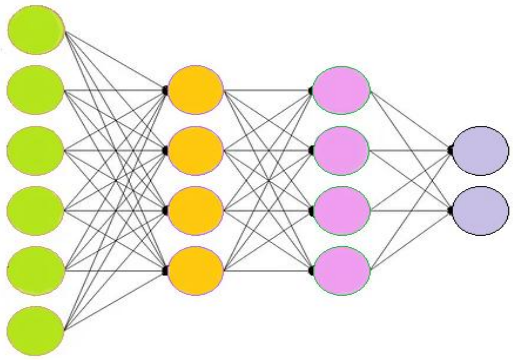


Fig. 1.5. Fully connected layers

Every input node and every output node are coupled in a fully connected layer. The fully-connected layers are exhibited in fig. 1.5. Typically, an architecture’s final few layers include the fully connected layers. One of the approaches for acquiring non-linear combinations of the high-level characteristics, as computed by the convolutional

layer's output, is carried out by introducing a fully-connected layer which is a feed forward network.

The outcome of the last convolutional or pooling layer is obtained in a matrix form. This outcome has to be flattened, which means all the values has to be unrolled into a column vector as shown in fig. 1.6. The flattened outcome is then transmitted to a feed-forward network and subsequently backpropagation is employed in every training iteration. After executing a specific number of epochs, the framework gains the capability to identify the dominant characteristics and the low-level characteristics in the pictures. Finally the model categorizes the objects utilizing the softmax activation in the last layer. The probability signifying an object's belongingness to a specific class is being provided by the softmax function.

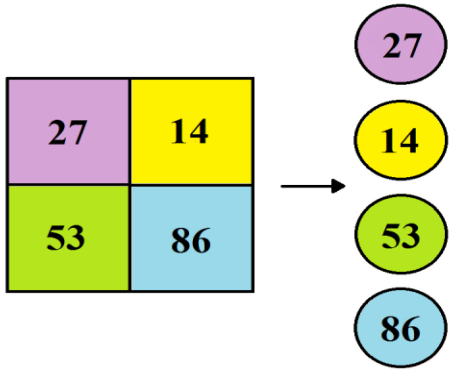


Fig. 1.6. Flattening operation

A group of dependent non-linear functions constitute the neural networks. Every function is made up of a neuron (or a perceptron). The perceptron is regarded as the basic building block of neural networks. It accept inputs, namely x_1, x_2, x_3, ... , x_n, which is linked to a coefficient known as a weight, namely W_1, W_2, W_3, … , W_n. The perceptron then calculates the weighted sum Z, a linear combination of weights and inputs. The weighted sum is then subjected to a non-linear transformation using an activation function. f . The mathematical expression is given as,

$$Y = f\ [\sum_{i=1}^{n}(x_i W_i)+ b] \tag{1.1}$$

where, bias is denoted by b. The weights control the impact of every input on the network architecture.

A network layer is a collection of connected neurons. There are various kinds of layers, like the input layer, hidden layer, and output layers. The network's input layers accept the input data, the hidden layers carry out the intermediary calculations, and the output layers provide the final predictions as the outcome.

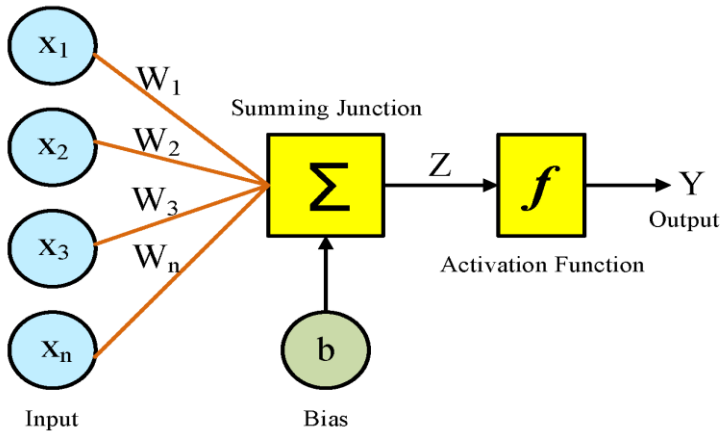


Fig. 1.7. The working of a perceptron

The term "forward pass" describes the method of calculating the numerical scores of the output layers from the input details. In the forward pass the data traverses through every neuron from the topmost to the bottommost layer. The term "backward pass" describes the method of calculating the modifications in the assigned weights utilizing the gradient descent approach. In case of the backward pass the bottommost layer is computed initially before moving to the top layer. The following equation 1.2 depicts the mathematical calculation to determine the output tensor's dimension from the input tensor.

$$W_2 = \frac{W_1 - F + 2P}{S} + 1 \qquad (1.2)$$

where W1, W2, F are the input tensor's, output tensor's and the kernel's width/height respectively, S and P are the stride and padding respectively.

1.11 Backpropagation

Backpropagation is a technique used to modify the weights during the network training for obtaining more accurate result by transmitting error rates back to the neural network layers. It entails the modification of the network's weights in accordance with the losses achieved in

the earlier iteration. Low error rates are obtained by the appropriate tuning of weights, and this improves the model's applicability and makes it more reliable.

The random initialization of the biases and weights in a network can produce high error value while predicting the object. A popular approach to minimize the error value is the applicability of the backpropagation methodology for training the neural network. The backpropagation methodology calculates the gradient of the error rate in accordance with every weight implementing the chain rule, calculating the gradient successively for every layer, and traversing backward from the final layer in order to reduce redundancy in calculation of the intermediary terms. A raise in the weight reduces error in case where the gradient is negative. A reduction in the weight minimizes the error when the gradient is positive.

1.12 Gradient Descent Algorithm

The losses acquired in a neural network model is minimised using the potent optimisation technique known as gradient descent. Finding the best collection of weights that minimises the loss function is the main objective of gradient descent algorithm. This technique iteratively modifies the weights along the course of loss function's steepest descent, to arrive at a local or the global minimum. Local minimum refers to the minimum values of the parameter found within a specific span of loss function. The lowest value of the parameter throughout the whole range of loss function is referred as the global minimum.

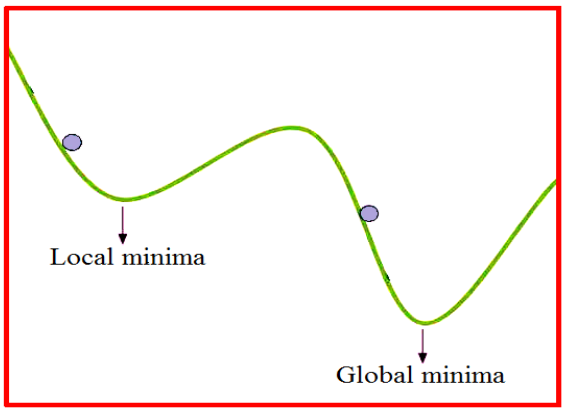


Fig. 1.8. Visualization of gradient descent

The gradient, which is a vector of partial derivatives, signifies the rate of change of the loss function in relation to the weights. The algorithm advances in the direction of the minimum loss function as the weights are upgraded towards the negative gradient. The gradient descent method determines the gradient of a differentiable function and traverses in the reverse direction of the gradient for updating the set of weights so as to minimise the loss function at a local or global minimum.

The gradient descent algorithms are available in various types, each of which has a unique method for calculating the update in the parameter.

i. Batch Gradient Descent: In this version, the gradient and parameter updates are computed using the whole training database. For large databases, this procedure can be sluggish, but it assures convergence to the global minimum.

ii. Stochastic Gradient Descent (SGD): In this version, the gradient and parameter updates are computed using a single random training sample. Although this procedure is quicker than Batch Gradient Descent but the updates in parameter can be noisy and may not converge to the global minimum.

iii. Mini-Batch Gradient Descent: In this version, the gradient and parameter updates are computed using a minute part of the training database. This is less noisy than SGD and quicker than Batch Gradient Descent.

iv. Momentum-based Gradient Descent: In this version, the parameters are updated depending on both the previous updates and the current gradient. This facilitates the algorithm's ability to avoid local minima and speed up the convergence process.

v. Adagrad: In this version, the learning rate is effectively scaled for every parameter according to the previous gradient data. As a result, the commonly used parameters face smaller updates and the seldom used parameters face larger updates.

vi. RMSprop: In this variation, the sliding average of the squared gradient is being used to effectively adjust the learning rate for every parameter. This method converges more quickly when noisy gradients are present.

vii. Adam: In this variation, both the sliding average of the gradient and the squared gradient are used to effectively adjust the learning rate for every parameter. This mehod combines the benefits of Adagrad, Momentum-based Gradient Descent and RMSprop and is the most widely used deep learning optimization technique.

1.13 Learning Rate

It is a hyperparameter which regulates the size of the step to be considered while updating the weight. Slow convergence is a consequence of a low learning rate, whereas a high learning rate can overshoot the minimum position and result in oscillation around the minimum position. It is crucial to select a learning rate that maintains a balance between the optimization stability and the convergence speed.

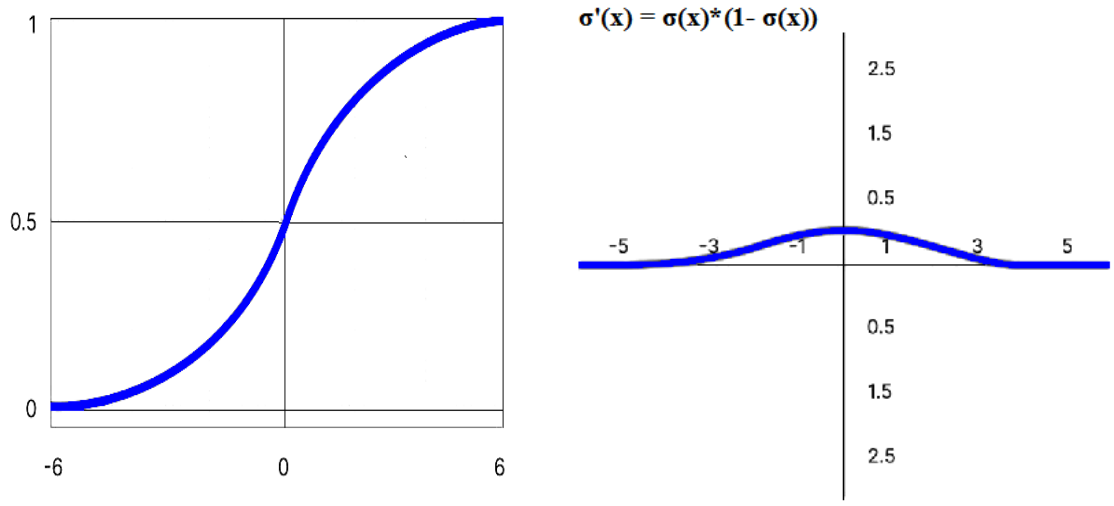
1.14 Non-Linear Activation Functions

An activation function decides the activity status of a neuron. Thus an activation function determines whether a neuron's input into the architecture is crucial for prediction. The activation function's fundamental objective is the conversion of the cumulative weighed input from a node to an output info that may either be communicated into the following hidden layers or be used as an output.

1.14.1 Sigmoid Activation Function

As an input parameter, this function adopts any real value and yields output ranging from 0 to 1. The more positive is the input, the closer the outcome approaches 1, and the more negative is the input, the closer the outcome approaches 0. The function can be formulated as:

$$\text{Sigmoid, } \sigma(x) = \frac{1}{1 + e^{-x}} \tag{1.3}$$



(a) (b)

Fig. 1.9. (a) The sigmoid activation function and (b) its derivative

The sigmoid activation function's derivative is given as

$$\sigma'(x) = \sigma(x) * (1 - \sigma(x)) \tag{1.4}$$

As portrayed in fig.1.9 (b), the gradient values are very prominent within the interval -3 to 3, and the curve becomes considerably flatter in other areas. This signifies that the function will have extremely low gradients for any number larger than 3 or smaller than -3. As the gradient proceeds toward zero, the network confronts the Vanishing Gradient issue and stops learning.

The benefit of using the sigmoid as an activation function:

- The activation function is differentiable and produces a smooth gradient, therefore prevents the output value from jumping.

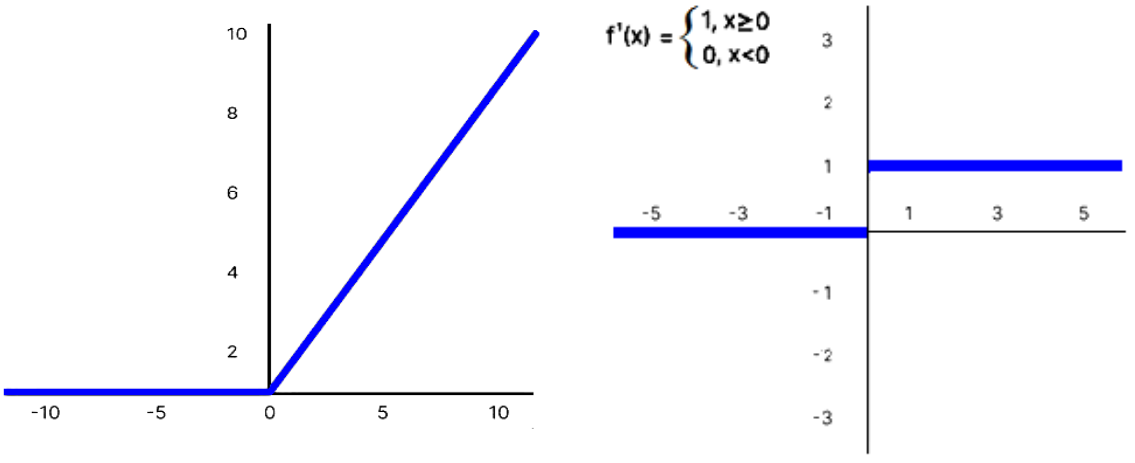
The drawbacks encountered while using sigmoid activation function:

- The function's output is not symmetric near zero. As a result, every neuron will produce output of same sign. This causes the neural network training to be highly challenging and unreliable.
- The network experiences the Vanishing Gradient issue.

1.14.2 ReLU Activation Function

Rectified Linear Unit (ReLU), unlike its name, is not linear. The function offers same advantages as the sigmoid activation, but provides superior performance. It supports backpropagation, and is also computationally effective. Here, the fundamental aspect is that the ReLU does not stimulate all the neurons concurrently. In case, the inputs are negative, the neurons become inactive. The ReLU activation function and its derivative are depicted in fig. 1.10. The function is mathematically formulated as:

$$\text{ReLU, } R(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{1.5}$$



(a) (b)

Fig. 1.10. (a) The ReLU activation function and (b) its derivative

The benefits of using the ReLU activation:

- The ReLU activation is significantly more computationally effective in comparison to the tanh and sigmoid function because all the neurons are not activated at the same time.
- Due to non-saturating characteristic of the ReLU function, it speeds up the convergence process of gradient descent towards the loss function's global minimum.

The drawback encountered while using ReLU activation function:

- The network experiences the Dying ReLU issue.

The dying ReLU issue addresses a condition where numerous ReLU neurons produce a value of 0 as output. This scenario occurs when the majority of the inputs lie in the negative range. Fig. 1.10 (b) depicts that the gradient value is 0 on the graph's negative side. As the majority of neurons return 0 as an output, the gradients cannot propagate during backpropagation and hence the biases and weights fail to get update. Eventually, a significant portion of the network dies and remains inactive, thus losing its capability to learn further. Since ReLU's slope in the negative input range is 0, the network cannot be revived to learn features, after it has been dead (i.e., locked in the negative range and producing 0 as output). As a consequence the training efficiency of the model significantly declines.

The dying ReLU issue arises generally due to the following factors:

- High learning rate
- Large negative bias

The popular strategies addressed to handle the dying ReLU issue are

- Use of a lower learning rate
- Use of different variation of ReLU namely Parametric ReLU (PReLU), Leaky ReLU.

1.14.3 Softmax Activation Function

The softmax activation function is being employed to transform a vector containing K number of real values into another vector with K real values which sums up to 1. This activation function converts the input that might be zero, negative, positive, or even a value greater than one, to numbers ranging from 0 to 1, facilitating their interpretation as probability. The function can be mathematically expressed as:

$$\sigma\left(\vec{z}\right)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{1.6}$$

$$\sigma\left(\vec{z}\right)_1 = \frac{e^{z_1}}{e^{z_1} + e^{z_2}} \tag{1.7}$$

where, z_i represent the components of the input vector which may include any real value. The denominator indicates that the normalization has been performed and assures that all

values of the function's output will total to 1, establishing an appropriate probability distribution.

❖ The softmax activation function vs. the sigmoid activation function

The sigmoid operates on a scalar, but the softmax utilizes a vector. Actually, the sigmoid is a particular variant of the softmax function for a classifier with two input classes. This can be demonstrated by setting the input vector to [x, 0] and computing the first element of the output using the conventional softmax mathematical expression:

$$\sigma\left(\vec{z}\right)_1 = \frac{e^{z_1}}{e^{z_1} + e^{z_2}} = \frac{e^{x}}{e^{x} + e^{0}} = \frac{e^{x}}{e^{x} + 1} \tag{1.8}$$

The numerator and denominator when divided by e^{x}, the equation 1.8 becomes

$$\sigma\left(\vec{z}\right)_1 = \frac{1}{1 + e^{-x}} \tag{1.9}$$

$$\sigma\left(x\right) = \frac{1}{1 + e^{-x}} \tag{1.10}$$

The equation 1.10 highlights that the softmax function becomes identical to the sigmoid function, where there are two classes. There is no need to compute the second vector element, since the two probabilities must add to 1. Thus, while designing a two-class classifier, instead of working with vectors, the sigmoid function can be utilized. However, the softmax function must be used when there are more than two classes that are mutually exclusive.

When dealing with more than two mutually non-exclusive classes (i.e. multi-label classifier), the classifier is splitted into several binary classifiers, each of which uses its own sigmoid activation function.

1.14.4 Tanh Activation Function

The tangent hyperbolic activation function is abbreviated as tanh function. This function inputs any real value and outputs values in the range -1 to 1. An increase in the positivity of the input, results in an output value close to 1, and an increase in the negativity of the input, results in an output value close to -1.

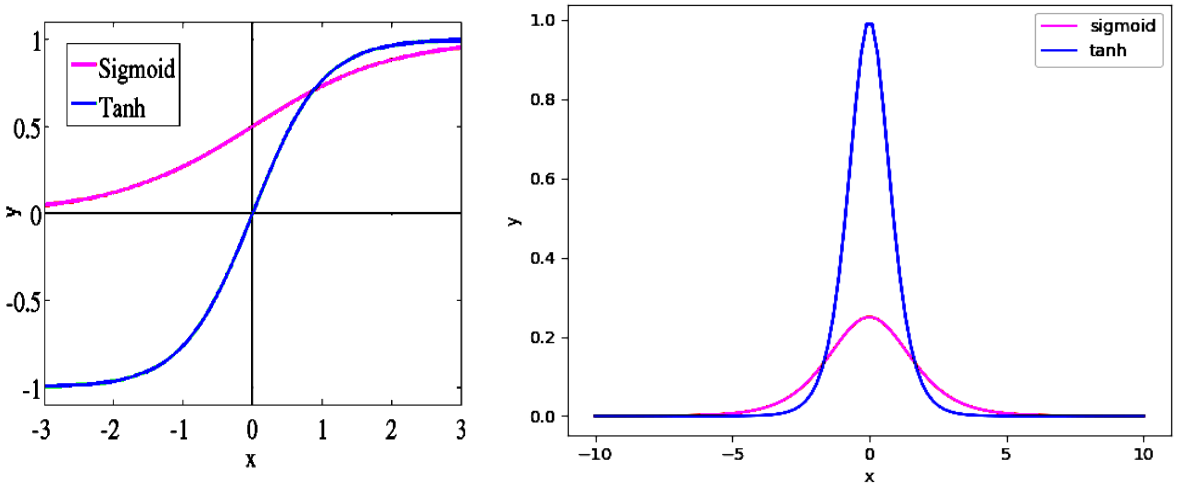
In case of multi-layer neural networks, the tanh activation is favored over sigmoid activation function because of its superior performance but it fails to resolve the vanishing gradient issue which the sigmoid activation encounters. The ReLU function has the capability of handling the vanishing gradient problem more efficiently. The tanh and sigmoid function are remarkably similar. A stretched and shifted variation of the sigmoid is the tanh function. The function can be mathematically expressed as:

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{1.11}$$

The tanh activation function's derivative is given as

$$\tanh'(x) = 1 - \tanh^{2}(x) \tag{1.12}$$

The behaviour of the gradient is a significant distinction between the tanh and sigmoid activation function. Fig. 1.11 (b) portraits the gradient of the sigmoid and tanh function.



(a) (b)

Fig. 1.11. (a) The tanh and sigmoid activation function and (b) their derivatives

The data typically remains centered around 0 when these functions are utilized in a network. Therefore, special attention should be provided to the behavior of every gradient in the vicinity of 0. It has been observed that the gradient of tanh function is larger than the sigmoid. This indicates that the tanh function when employed in architecture, introduces larger gradients in the training process resulting in huge upgradation to the network's weights. The tanh activation should thus be employed when high gradients and large learning steps are

required. The outcome of tanh function is symmetric around 0, which promotes quicker convergence.

The benefit of using the tanh as an activation function:

- The gradient in case of tanh function is larger in comparison to that of the sigmoid activation function.

The drawback encountered while using tanh activation function:

- The tanh function suffers from the vanishing gradient issue similar to the sigmoid function.

1.15 Convolutional Long Short Term Memory Network

The Long Short Term Memory (LSTM) architecture is employed to tackle the flaws of the exploding and vanishing gradients in a Recurrent Network. The LSTM recurrent unit is designed to remember the prior information that the network has encountered and to forget the insignificant information. To accomplish this, different gates (such as input, forget, output gate) are introduced for executing various operations. The gates and the memory cell comprise the most important elements of the LSTM. Every LSTM unit simultaneously supports a vector known as the Internal Cell State that specifies the data retained by the preceding LSTM unit.

The Convolutional Long Short Term Memory (CLSTM) layer has the similar architecture as that of the LSTM layer, only the convolution operations are used instead of internal matrix multiplications. The CLSTM layers are usually employed when dealt with sequential image processing.

1.16 Overfitting

The overfitting is an unacceptable behaviour of learning, where a neural network model makes correct predictions for the training dataset, but not for the test dataset. The overfitting happens when a model attempts to include all the data points, or more than the necessary data

points from the available dataset. As a result, the model begins to accept incorrect and noisy values from the dataset, which decreases the model's effectiveness and accuracy to a great extent.

The various factors responsible for the overfitting are as follows:

i. Use of a small training dataset which does not have sufficient data samples to precisely define all the possible input data values.
ii. Use of a dataset, which is noisy (i.e. it includes a lot of irrelevant information) to train a network.
iii. Consumption of an excessive amount of time for training a network with a single sample set of data.
iv. Use of very complicated networks, which allow the learning of the noises available in the training dataset.

❖ K-fold cross-validation

Cross-validation is a tactic used to detect the overfitting. In the K-fold cross-validation process, the training set is being split into K equal sized sets of samples termed as the folds. The training phase involves several iterations. The following steps are carried out in every iteration:

i. One set is retained as the validation dataset while the rest of the K-1 sets are used to train the network.
ii. The network’s outcome is rigorously studied on the validation set.
iii. The potential of the network is then scored depending on the quality of prediction.

The iterations continue till the network has been validated on all sample sets. The average of the scores obtained from all the iterations is evaluated for assessing network's overall efficacy.

❖ The overfitting can be prevented by employing the following techniques:

i. Early stopping

The early stopping mechanism the pauses model's training process before the noisy data in the input are learnt. Moreover, selecting the right parameter value for the early stopping is crucial, otherwise the model will not provide correct predictions.

ii. Pruning

The precise outcome of a model is depends on a number of parameters or features. The pruning methodology helps to discard the unimportant features and selects the most crucial ones from the training dataset.

iii. Training done with more data

An expansion of the training dataset for including more information in the model, can elevate the prediction efficiency as the model gets opportunity to explore the inter-relationships between the output and input variables.

iv. Data augmentation

Data augmentation is an approach in which the sample dataset is slightly modified each time it is processed. This can be accomplished by making minor transformations like flipping, translation and rotation to the training data. After data moderation, the training data seems to be unique to the model.

v. Regularization

Regularization consists of a set of optimization/training approaches which aids in reducing overfitting. These approaches exclude those aspects which does not influence the model's prediction. The regularization process imposes penalty on the input parameters having higher coefficient, thus reducing the model's variance to a certain extent. There are several regularisation techniques, like dropout, Lasso regularisation, and L1 regularisation, each of which helps to minimize the noises present in the data.

vi. Ensembling

In ensembling technique, the predictions from numerous different algorithms are combined. Models which usually provide inaccurate results are referred to as weak learners. To obtain precise outcomes, the ensemble technique integrates all the weak learners. The sample data is analyzed utilizing several models and the most precise results are selected. Boosting and bagging are the two significant ensemble techniques. While the boosting method trains various models sequentially one after another to obtain the final outcome, the bagging method trains the models in

parallel. The ensembling technique is frequently employed to minimize the variance in a noisy dataset.

1.17 Underfitting

Underfitting is an undesirable behavior of a model which is encountered when the model fails to establish a significant inter-relationship between the output and input data. If a model is not trained with sufficient sample data points for enough time, or if the input characteristics are not important enough to establish a correlation between the input and output data, then the underfitting happens.

❖ The following are the techniques used to prevent underfitting.

i. An increase in the model's training duration
ii. An increase in the number of features.

❖ Underfitting vs.Overfitting

Small variance and high bias in predictions are experienced by underfitted models. These models produce inaccurate outcomes both when training dataset and testing dataset are considered. On the contrary, large variance and small bias are experienced by overfitted models, which provide precise responses for the training dataset but fails in case of the test dataset. It is very essential to identify a model's optimum point that lie between overfitting and underfitting stages and achieve an equilibrium between the variance and bias. A well-fitted model has the capability to effectively determine the dominating trend for both the train and the test dataset.

- Signal: It signifies a pattern of the information which enables a model to learn the features.

- Noise: It is the unwanted and insignificant data which degrades the model's performance.

- Bias: The bias is a prediction error which arises due to incorrect assumptions in a model. It represents the disparity between the predicted and actual outcome. A model that experience large bias does not pay much attention to the training data and oversimplifies the machine learning algorithm.

- Variance: The variance is an error which occurs when a model includes the fluctuations and noises in the data. A model that experience large variance learns too much from training data, but fails to generalize on consideration of new data. As a consequence, these models exhibit high accuracy on training data, whereas produces significant error on test data.

1.18 Goodness of Fit

In terms of statistics, the goodness of fit, signifies how closely the predicted and actual values match each other. The machine learning approaches aim to attain the goodness of fit in order to produce the best results. Ideally, a model which has achieved a good fit, makes predictions with zero errors and is said to be well situated between the overfitted and the underfitted stages.

The prediction errors for the training and testing dataset gradually decrease as the training time of a model is gradually increased. However, if training time is too long, the model also learns the dataset's noise and hence the accuracy of the model can suffer from overfitting. As a consequence of overfitting the test dataset's prediction errors start to rise. Thus the point right before the increase in the error is the good fit point, which is very crucial to identify in order to design an efficient model.

1.19 Dropout

The random elimination of nodes in the training phase facilitates a single model to process a framework with several distinct network architectures. This process, which is known as dropout, provides a highly efficient regularization approach to eliminate overfitting at a very

low computational cost. On every update in the learning phase, the dropout randomly assigns a value of zero to the outgoing edges of the hidden units.

A massive network can overfit when trained with smaller datasets. The ensemble approach necessitates the fitting and storing several models, which becomes challenging, when the models involved are massive and require large amount of time to train. The use of dropout on a layer during training leads to the random sub-sampling of the output which results in the thinning of the network. Dropout can be performed on the input layer, convolutional layer, fully connected layer, recurrent layer, but not on the output layer.

Advantages of dropout

- The use of dropout, in each iteration, helps to deal with a smaller network as compared to the previous version and thus the process of regularization is achieved to reduce overfitting.

Disadvantage of dropout

- Dropout makes the model's learning procedure noisy by imposing responsibility on the nodes of a layer.

1.20 Challenges and Objectives of the Work

The recognition of DR is affected by numerous factors. The most difficult task is distinguishing the stages of DR. Increasing the accuracy of segmentation of the DR lesions in retinal photographs is another tough challenge. Thus, it is crucial to create a model for automatically assessing the presence of lesions in a retinal photograph and assisting the ophthalmologists in the process of rapid screening of patients.

The objectives of this study include:

i. Creation of automatic models to localize optic disc, exudates and macula region in fundus photographs.

ii. Rapid and accurate segmentation of optic disc, exudates and macula in retinal pictures during the mass screening programmes of diabetic patients.
iii. Improving the efficacy of the DR diagnosing methodologies for assisting ophthalmologists in early treatment by identifying the retinopathy lesions.

1.21 Outline of the Thesis

The thesis is being outlined in the following manner. The **chapter 1** describes the details of the anatomy of human eye. This chapter also discusses the details of the probable causes, risk factors, symptoms, stages of severity and the effective treatment of DR. The **chapter 2** addresses the various techniques adapted for the localizing and segmenting the OD, exudate and macula region in fundus photographs of DR affected individuals. The **chapter 3** presents the convolutional network based OD localizing and segmenting methodology. The introduction of the Convolutional Long Short Term Memory structure and the procedure of training the network with the pre-trained weights has improved the ability to achieve fast convergence and high accuracy. The **chapter 4** elaborates the details of the exudate detection approaches employed for rapid screening of diabetic patients. The methodologies make use of the channel and spatial attention module, Convolutional Long Short Term Memory architecture and Residual Extended Skip unit to boost the potency of the encoder-decoder network. The **chapter 5** explores the method adopted to recognize the macula region which is a dark zone near the retina's centre accountable for regulating the central vision, colour perception and visual acuity. The incorporation of Residual Extended Skip unit substantially enhances the segmentation performance by increase in the valid receptive field. The **chapter 6** summaries the contributions of the study and reveals the future scope of the work.

Chapter 2

Literature Survey

2.1 Introduction

The noteworthy advancement in digital image processing techniques has promoted the interest in developing models for the rapid screening of DR patients. Eventually, several works have been accomplished to identify the DR lesions automatically using digital fundus pictures. This chapter includes a review of the pertinent works carried out for the recognition and segmentation of OD, exudate and macula in fundus pictures has been addressed. There are three subsections in which the review is conducted. The first and second part consider the methodologies utilized in localizing and segmenting the optic disc and exudates respectively and the third part deals with the studies pertaining to the recognition of the macula region.

2.2 Survey on the Segmentation of the Optic Disc

Neural Networks have boosted the interest of the researchers to develop algorithms for helping specialists to recognize disorders. Convolutional Neural Network (CNN), which is frequently utilized in computer vision, allows the system to automatically perceive and comprehend a picture. Some application which utilizes CNN algorithms for detecting and segmenting objects are presented in [10]–[16]. Fully Convolutional Network (FCN)

presented by Long et al. [17], forms the basis of the most successful deep learning models. This method uses CNN as its main component to extract features by using convolution layers thereby producing feature maps as output rather than the classification scores. Up-sampling of these feature maps are then carried out to obtain dense pixel-wise output. A distinctive feature of this method is that, here the CNN is trained with input images of similar dimensions in an end-to-end way for properly segmenting the objects. FCN finds its application in segmenting biomedical images [18], and in semantic segmentation [19]. Drozdzal et al. in [20] modified the conventional FCN by using short and long skip connections to obtain better performance. This approach after undergoing additional refinement has been incorporated in U-Net [21] architecture. Skip connections are used in the U-Net framework, whereby higher-level feature maps are integrated with lower ones, that helps in accurate pixel-level localization. The encoder-decoder framework has proved its performance in various image segmentation contests including medical image segmentation [22], [23] and satellite image processing [24]. It is a known fact that deep networks require large datasets with thousands of annotated samples for training the model successfully, which is not always possible to acquire in case of biomedical images. The effectiveness of U-Net with limited dataset of images makes it a popular medical image segmentation technique. U-Net is comprised of a contracting pathway that accumulates contextual information and propagates them to higher resolution layers through a large number of feature channels. It is followed by a symmetrically growing pathway which helps in accurate localisation. Several modifications of the network have been done based on the application of U-Net [25]–[27].

A very renowned network that has widespread use in segmenting biomedical images is the Recurrent Neural Network (RNN) [28]. A distinctive feature of the U-Nets using RNNs is the presence of feedback coupling [29], which makes the framework more appropriate to deal with the dynamic property of the data [30]. Bai et al. [31] fused RNN and CNN framework for utilizing in sequence segmentation of biomedical image. The Fast R-CNN methodology presented by Girshick et al. [32] and Faster R-CNN [33] models have been well accepted for segmentation tasks. Alom et al. [34] enhanced segmentation performance by developing Recurrent Residual Convolutional Neural Network based U-Net model. Among the RNN models which are extensively used, the Long Short Term Memory (LSTM) [28] proved very effective in predicting sequential problems. This approach is widely utilized in image captioning [35], natural language processing [36], audio signal processing [37]. RNNs are unable to retain long-term information while learning by using gradient descent algorithm,

because of the vanishing gradient complication. The problem got resolved on the introduction of LSTM frameworks. These networks handle the vanishing gradient problem by incorporating three gates namely input, output, forget gate and a memory cell. The information which is to be updated for the upcoming step is determined by the input gate. The output gate decides which information is to be displayed as output. The forget gate determines which information are to be thrown away. LSTM networks are very popular in prediction when input vectors are sparse matrices. A special type of LSTM network is the Fully Connected LSTM [38] network. This framework is considered to be a powerful tool while dealing with temporal characteristics, but in case of spatial data the methodology suffers from a lot of redundancy. To mitigate the issue convolutional structures have been added in each input-to-state and state-to-state passes to produce a modified structure named Convolutional LSTM (CLSTM) [39]. The CLSTM has found its application in analyzing the next frame prediction [40] and in volumetric data sets [41].

In the proposed work, the CLSTM which is introduced to handle the spatiotemporal characteristics, not only increases the speed of convergence of the framework, but also enhances the overall potential of the method by extracting more features.

2.3 Survey on the Segmentation of the Exudates

Exudates, which typically develop from lipid deposits in the retina, appear in two varieties, namely soft and hard exudates. While the soft exudates, also termed as the cotton wool spots, emerge as white dots, the hard exudates look bright yellowish in a fundus photograph.

Several efforts have been adopted in previous studies for segmenting the exudates. Numerous image processing techniques are being introduced in recent times [42]–[45] for detecting and segmenting the exudates in fundus photographs. The development of different methodologies for automatically detecting and segmenting exudates over the previous years may be broadly categorized into four classes, morphology based, region growing, thresholding, and machine learning techniques. Convolutional network-based strategies that rely on the quantity of training samples and the quality of the labels are found to have better efficiency. In [46]–[48] morphology based operations have been considered to perform the

segmentation operation. These techniques often start by identifying and removing predominant components, like the vascular structures and OD in the fundus photographs, to lessen the interference that can hamper the process of segmenting the exudates. Sopharak et al. [46] used morphological reconstruction and histogram equalization techniques for segmenting exudate. For segmenting exudate, Fraz et al. in [49] presented a method that incorporates Gabor filter, morphological reconstruction methodologies and bootstrap decision function.

Region growing algorithm [50] has been designed to automatically segment the exudates. Thresholding based techniques depending on the image gray scale level [51] and Otsu [52], [53] which is a clustering-based thresholding methodology are utilized for segmenting the exudates. Sanchez et al. [54] combined the dynamic thresholding algorithm with a mixture model for exudate segmentation. This method used a combination of image processing and pattern recognition methodology to provide a robust system with high adaptability to various images. In order to segment exudates automatically and unsupervisedly, Giancardo et al. [55] formulated a feature set utilizing adaptive colour vector representation and wavelet decomposition technique. García et al. [56] used radial basis function, multilayer perceptron, and support vector machines for detecting the exudate.

Deep learning approaches, which have emerged as superior image segmentation technique as in [57], [58], are able to resolve the feature extraction issues, that are the primary challenges in segmentation task, compared to traditional pattern recognition methods. Deep neural networks have excelled in variety of applications particularly in medical image analysis [59], [60]. The deep neural architectures possess the ability to extract the features effectively but, the selection of the correct architecture and the parameter values are extremely important when creating a model. In medical image analysis, Fully Convolutional Network [61] and U-Net [21] are the two most used machine learning-based techniques. These techniques generally produce a feature vector depending on the brightness, contextual information, shape, colour, edge strength of every pixel which has to be identified using a machine learning algorithm. The U-Net has been widely accepted because of its remarkable capabilities in segmenting lesions from images. The feature fusion technique and the skip-connections constitute the key elements in improving the segmentation outcomes.

Considering a deep convolutional network along with principal component analysis dimensional reduction methodology, which is an unsupervised linear transformation

technique, Chudzik et al. [62] developed an automatic procedure for segmenting exudate. The automatic feature extraction characteristics of approaches using CNN have proved to attain promising outcomes in a various image processing applications [63], [64]. In [64], Tan et al. designed a single convolutional network for identifying haemorrhages, exudates, and microaneurysms. This method addressed the limitations of the model developed by Prentašić et al. [65] and Yu et al. [66] which required segmentation of the vascular structure and removal of optic disk before the localization of exudate. A fully convolutional residual framework was formulated by Mo et al. [67] for segmenting the exudates.

U-Net having encoder-decoder convolutional network architecture and skip-connection, helps in the accurate segmentation of small objects and hence finds wide usage in clinical image analysis. Various U-Net modifications with more sophisticated unit including residual, recurrent block [68], [69] and attention module [70], [71] have been extensively investigated to improve segmentation accuracy. Bahdanau et al. in [72] suggested an attention methodology which was modified by Wang et al. [73] to determine the spatio-temporal inter-dependencies of video sequences. By assessing the interrelationship between the channel characteristics, Cheng et al. [74] developed an attention model for action recognition. Rao et al. in [75] presented the self-attention approaches for analyzing medical images. Zhao et al. [76] designed SCAU-Net by fusing attention module to an encoder-decoder framework. SENet, as proposed in [77], adaptively recalibrates channel features by modeling channel interdependencies. The squeeze-and-excitation module in SE-Net was further extended, considering the positive attributes associated with each of the channel and spatial attention methodology, by Woo et al. [78]. A self-attention technique was fused with generative adversarial network by Zhang et al. [79] for developing key objects from long-distance complementary features in images. A CNN and LSTM coupled architecture developed in [80], recognized the temporal and spatial information and extracted the hidden patterns effectively. Reza Azad et al. [81] considered using the CLSTM algorithm to combine the features extracted from both the pathways.

The unavailability of sufficient labeled datasets in case of medical image processing is also a major concern for training the deep architectures. To overcome the constraints, it is necessary to study various deep learning algorithms and design a network that performs well in the adverse conditions. The proposed work aims to develop an effective attention modulated deep neural architecture incorporating CLSTM mechanism for successfully segmenting the exudate in the fundus image.

2.4 Survey on the segmentation of the Macula

The macula, a dark area situated near the retina's core is accountable for controlling the central vision, colour perception and visual acuity. The small depression densely packed with cone cells near the macula's center is termed as the fovea. Macula abrasions can have a substantial impact on the vision. It is crucial to understand the size and position of the macula for accurate assessment. Welfer [82] utilized morphological operations for segmenting the vessels and OD. Then, the macula is recognized by utilizing the knowledge that the macula is located 2.5 disc diameter away from the OD centre on the temporal side. Asim in [83] recognized the macula region considering the position of OD as a reference. The fovea is determined as the median of the macula area. The fovea point's vicinity is then carefully examined. If the point is significantly distant away from the blood vessels, it is maintained as the fovea; otherwise, the next minimal position in the cluster of darkest patches is considered to see whether it meets the aforementioned requirement. Gegundez-Arias [84] demarcated the fovea area taking into account the vessels and the OD location. The feature extraction and thresholding technique are thereafter applied in the chosen area to obtain fovea centre. In another work, Giachetti [85] outlined the macula by utilizing the circular nature criterion for both the OD and the macula, considering the fact that unlike the bright, vessel-interrupted OD; the macula region is darker and devoid of blood vessels. In this technique, the round shaped dark and bright elements have been identified utilizing Fast Radial Symmetry strategy. In another method, Tewari [86] predicted the position of the macula and the OD by employing parameter optimization technique which helped to define a line that splits a retinal photograph horizontally in two portions with nearly similar vascular density. Aquino [87] integrated the anatomical information with a macula segmentation process in order to maximize the benefits of both strategies. At first a horizontal line is developed by considering a parabolic structure with vertex at the middle of OD. The centre of the fovea is determined on the line, 2.5 disk diameter apart from the OD's middle point. The macula region of size equal to 2 disk diameter is then built all the way around the fovea centre. The actual fovea is then identified after thresholding the outcome of the green and red channel H-minima.

The macula and the fovea can be directly segmented without considering the position of the OD and the anatomy of the vessels using deep learning algorithms. Deep algorithms have

proved to be one of the renowned approaches for assessing a variety of medical images as in [88]. Macular degeneration with ageing has been effectively identified using deep learning algorithms formulated by [89]–[91]. To recognize the macula location, Wong [92] employed a seeded technique and Aravindan [93] adopted CNN algorithm. Yang [94] adopted a two-step method considering gradient based data in dual scales that concurrently made use of local and complementary global gradient details for segmenting the macula. SVM is utilized by Fuller [95] to perform semi-automatic retinal layer segmentation. Ishikawa [96] adapted optical coherence tomography to carry out macula segmentation.

In the proposed work, a fast, accurate and fully automated deep neural methodology to segment the macula region in fundus images has been presented. It is observed that even when dealing with a wide range of input images of various appearances, the proposed framework maintained its stable performance.

2.5 Summary

Various approaches have been discussed in this chapter for localizing and segmenting the optic disc, exudate, and macula region in the fundus pictures. The key purpose of this work is to investigate superior feature learning methodologies. The several morphological alterations that happen in the retina due to long term diabetes in individuals are explored. The literature survey reveals that there are still possibilities to make improvements in segmentation accuracy, specificity and sensitivity.

Chapter 3

Segmentation of the Optic Disc

3.1 Methodology Adopted for the Segmentation of Optic Disc

This work uses a CLSTM [39] incorporated encoder-decoder architecture. The most powerful encoder structure is identified by experimenting on seven encoder networks namely VGG11 [97], VGG13 [97], VGG16 [98], VGG19 [97], Resnet34 [99], Densenet121 [100] and InceptionV3 [101]. The VGG16 framework has been adopted as an encoder in the proposed model because it attended the best segmentation performance by providing minimum loss on all the test datasets. The VGG16 encoder network is then integrated with the decoder which is designed as a symmetric structure of that of the encoder to successfully frame a model, which is capable of providing better segmentation result with wide variation of images available in different public datasets.

The architecture of the method used in this study for detecting and segmenting the optic disc in fundus image is depicted in fig. 3.1. The VGG16 encoder consists of five stacks, each of which contains convolutional layers of kernel size of 3x3. A kernel is a matrix of weights which slides across the image and are multiplied with the input to extract relevant features. Bigger size kernels cover larger area and hence pay less attention to the patches in an image. A 3x3 kernel is used to allow the network to focus more on the patches in an image. The CLSTM is followed by the activation function, ReLU which adds non-linearity to the network and enables it to learn more effectively. This function replicates the input directly if it is positive, otherwise, produces an output equal to zero. The convolution stride and padding are set to 1 pixel. The purpose of padding the outer frame of the image is to allow for more

space for the convolutional filter to cover in the image and to preserve the original size of an image. The transition between the stacks is associated with max pooling from each 2x2 window and thus lowering the feature map size by 2. The initial two convolutional layers contains 64 filters each and number of the filters gradually doubles after every pooling operation till it attains a value of 512, thereafter remaining unaltered. Moreover the sequential operation of convolution CLSTM and max pooling ensure spatial contraction with progressive increase in content information and a reduction in location information.

The expansion block helps the framework in creating a segmentation map of high resolution. The decoder comprises of transposed convolutional layers that helps the feature map to get doubled in size and the number of filters to get cut in half. The output of a transposed convolutional layer is combined with the corresponding feature map from the contraction block. To maintain the number of filters identical to that as in the encoder section, the obtained feature map is processed with convolution operation. The upsampling process is carried out five times in order to match up with the five max pooling operations in the encoder section. Finally, 1×1 convolution operation followed by sigmoid is performed for obtaining the segmented mask as an output.

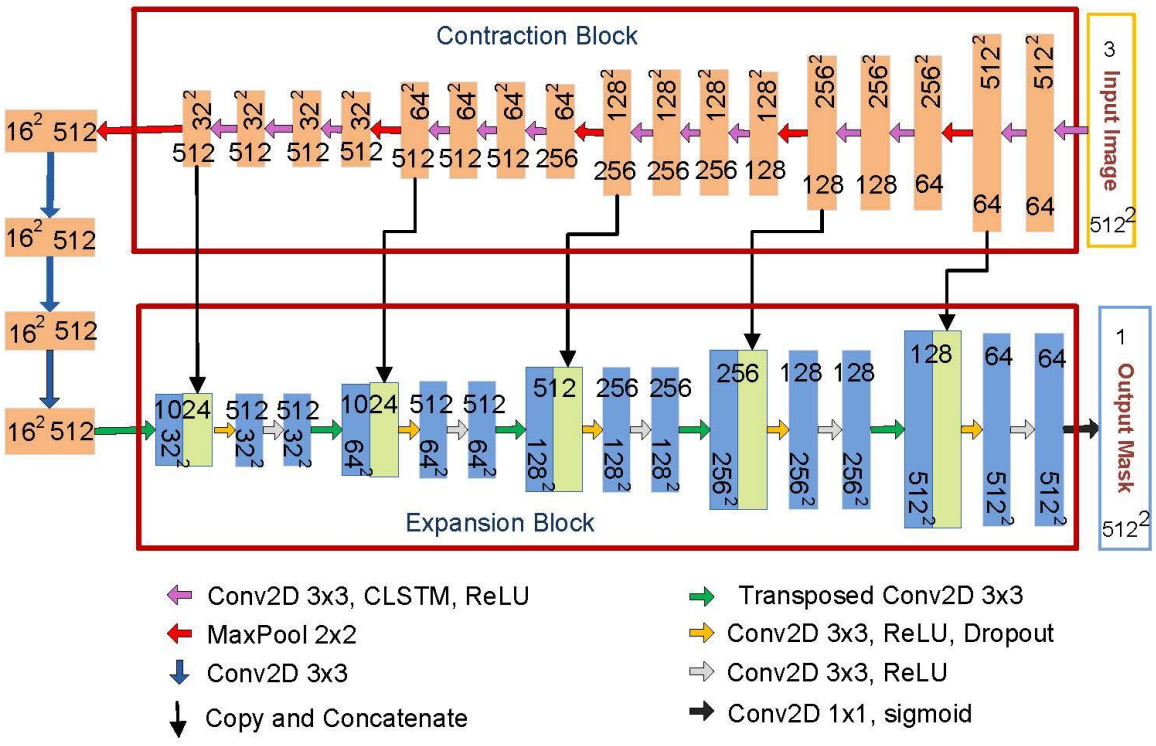


Fig. 3.1. Framework of the proposed network for the segmentation of optic disc

The proposed encoder is enriched with CLSTM, in which all the inputs, hidden states, outputs, and the three gates are represented by tensors where the final two dimensions are spatial dimensions used for denoting the rows and columns. To achieve a clear view, the inputs and the hidden states can be considered as vectors lined-up along a spatial grid. CLSTM predicts the future state of a cell in the grid by analyzing the previous states of its immediate neighbours and the inputs. This has been accomplished by utilising convolution operation in input-to-state and state-to-state passes. An important part of the CLSTM network is its memory cell, which helps in accumulating the state information, and the control gates. The memory can be loaded, cleared and accessed by using the control gates. Whenever an input (X_t) arrives, the information is retained in the memory, if the input gate (i_t) is on. If the forget gate (f_t) is active, the past memory information (C_{t-1}) will be forgotten. The output gate (o_t) when enabled propagates the latest memory information (C_t) to the final state. The following equations 3.1-3.5 reported in [102] control the operation of the CLSTM.

$$i_t = \sigma\left(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \times C_{t-1} + b_i\right) \quad (3.1)$$

$$f_t = \sigma\left(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \times C_{t-1} + b_f\right) \quad (3.2)$$

$$C_t = f_t \times C_{t-1} + i_t \times \tanh\left(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c\right) \quad (3.3)$$

$$o_t = \sigma\left(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \times C_t + b_o\right) \quad (3.4)$$

$$H_t = o_t \times \tanh\left(C_t\right) \quad (3.5)$$

where i_t, f_t, o_t, X_t, H_t, H_{t-1}, C_t , C_{t-1}, W, b, $\times$, $*$ and σ denotes the input gate, forget gate, output gate, input state, hidden state, previous hidden state, cell state, previous cell state, learnable weight, bias term, Hadamard product, convolution operator and sigmoid activation respectively.

The internal structure of the CLSTM framework is shown in fig. 3.2. The working of CLSTM as in [103] consists of three steps. In the first step the forget gate (f_t) decides whether to keep the previous information or forget it by considering both the previous hidden state and new input data. The network produces a vector after the application of a sigmoid function where the elements of f_t are converted into a number between 0 and 1. The f_t value is then multiplied (Hadamard product) with the previous cell state as given below:

$f_t \times C_{t-1} = 0$, if $f_t = 0$ which signifies that the network will forget everything.

$f_t \times C_{t-1} = C_{t-1}$, if $f_t = 1$ which signifies that the network will forget nothing.

where, C_{t-1} denotes the previous cell state.

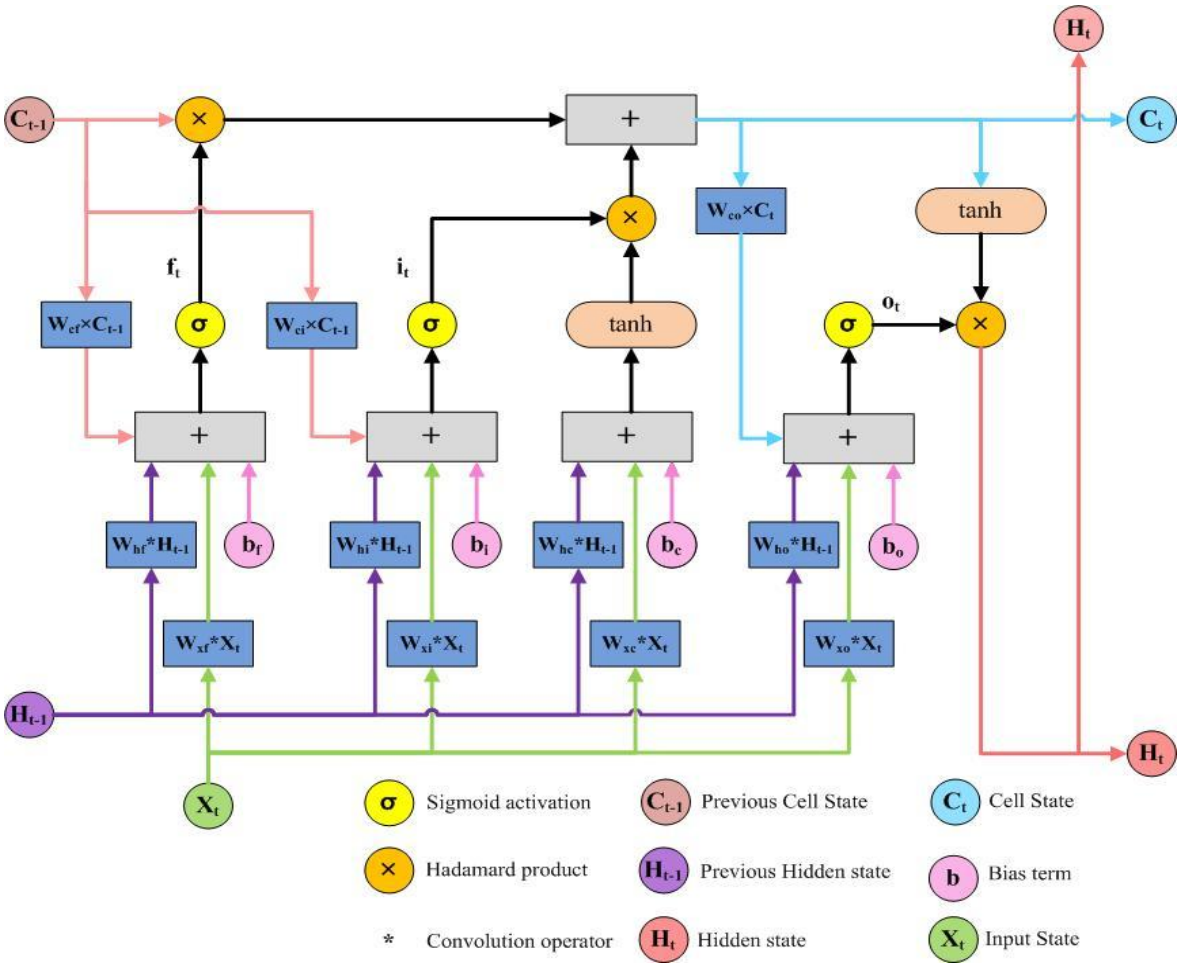


Fig. 3.2. The structure of CLSTM

The second step involves the input gate and the formation of new memory. The new memory network which is a tanh activated neural network, combines the previous hidden state information with the new input data to create a new updated memory vector. As the tanh function is used, the new information value will lie between -1 and 1. If the value is negative, the information is deducted from the cell state and if it is positive, the information gets added to the cell state. The new information, as reported in [104] is given by the equation 3.6.

$$\text{New information} = \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \quad (3.6)$$

where H_{t-1}, W, b_c, X_t and $*$ denotes the previous hidden state, learnable weight, bias term, input state, and convolution operator respectively.

The input gate (i_t), which is a sigmoid activated network, acts as a filter to determine which components of the new memory vector are to be retained. Due to the use of sigmoid function, this network outputs a vector of values in [0,1]. The i_t value is multiplied (Hadamard product) with the new information value, and then combined with the cell state, to update the memory cell. The i_t value near zero implies that the cell state element does not require updating. The updated memory cell (C_t) is given by the equation 3 as $C_t = f_t \times C_{t-1} + i_t \times \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c)$ where i_t, f_t, X_t, H_{t-1}, W, b_c, ×, C_{t-1} and $*$ denotes the input gate, forget gate, input state, previous hidden state, learnable weight, bias term, Hadamard product, previous cell state and convolution operator respectively.

In the third step, the new hidden state is determined by the output gate (o_t) considering the new input data, the previous hidden state and the newly updated cell state. The o_t act as a filter performing similarly as the forget gate network considering the new input data, the previous hidden state and sigmoid activation. This filter is applied on the newly updated cell state to ensure that only necessary information is passed on to the new hidden state. However, before the application of the filter, the newly updated cell state is conveyed through a tanh function to force the values to lie in the interval [-1, 1]. The new hidden state (H_t) is given by the equation 5 as $H_t = o_t \times \tanh(C_t)$ where C_t and × denotes the newly updated cell state and Hadamard product respectively.

CNNs suffer from losing the spatial data in the downsampling section. The maxpooling operation also introduces some heterogeneity in the feature maps. CLSTM resolves these issues by providing hidden states and memory cells in the downsampling section of the network. The CLSTM strategy also considers the low-level features and helps in passing them significantly during transition between the network layers. The initial few layers and the last few layers are the position where the majority of learning operation takes place. The network's initial layers deal in learning the general features specially the unique shapes and various components of the objects with which the network is being trained. For this reason, the first few layers are frozen so as to keep the weights unaltered during backpropagation, and only the last few layers are trained. The benefit of such strategy is to make the learning process go faster. Here, the pre-trained weights of the ImageNet dataset are considered as the initial weights. Adam as an optimiser, Model Checkpoint and Early Stopping as callbacks, are used while training the network. The rate of learning is maintained at 0.001.

CNNs are comprised of several convolutional and pooling layers which are discussed as follows:

- Convolutional Layer: This layer which consists of multiple layer maps of equal dimensions, highlights on spatially-local correlations by allowing linkages only from the direct neighbors of the pixel considered. The nodes, arranged in the form of maps in a convolutional layer are attached to the previous layer nodes with the help of kernels. Every particular set of layer maps has a kernel of its own, and therefore requires a collection of weights. A typical convolution operation is illustrated in the fig. 3.3. The convolutional layer is usually succeeded by activation function ReLU which helps to speed up the convergence process and is found to be similar in operation with that of the neurons in the brain. The ReLU function is stated as in equation 3.7.

$$\text{ReLU, } R(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \qquad (3.7)$$

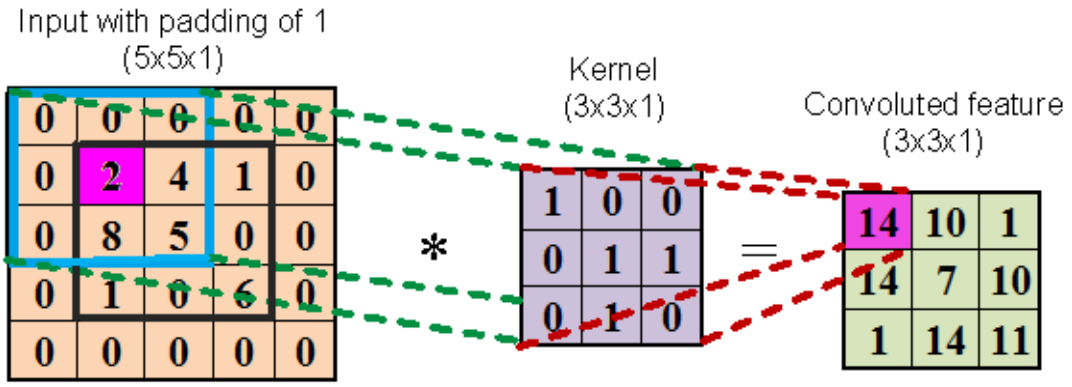


Fig. 3.3. A typical convolution operation.

- Max-pooling Layer: This layer subsamples a convolution layer map, firstly by splitting it to several uniform-sized non-overlapping windows and then by awarding the maximum available value in a relevant window to the corresponding node in the max-pooling layer map as demonstrated in the fig. 3.4. As a consequence of this operation, the count of the nodes decreases and consequently the calculations involved in the succeeding layers also reduces.

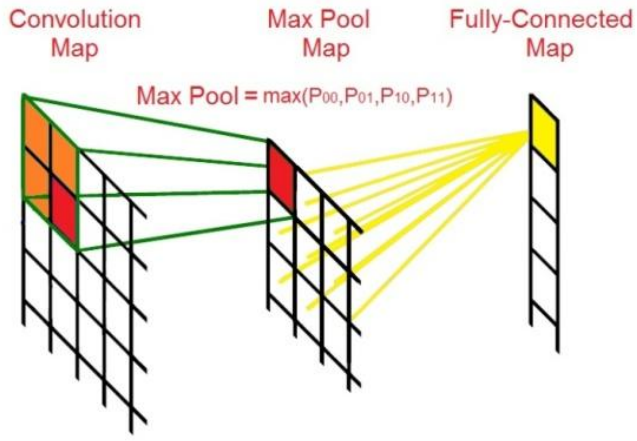


Fig. 3.4. The working of max-pooling layer in a network.

In each training iteration, a randomly oriented image patch considering the current source pixel as a centre, with dimension same as that of the input layer map is treated as the network input. The values obtained as input are passed to the subsequent layers in the network, in such a manner that the output of one layer becomes the input to its immediate succeeding layer and finally generating an output vector at the end of the last layer. The error between the real and the last layer predicted value of output vector performs a vital function in rearranging the weights of every connection by applying gradient descent back propagation [105] algorithm. The network intends to minimize the mean error obtained in the output by considering several iterations. The dropout layer helps to minimize over fitting during the training process by assigning a specific probability to the outputs of the hidden neurons possessing zero value. As a result the neurons which have been dropped out, do not participate in the ongoing training mechanism. Thus a model is forced to learn complex characteristics. The Dice Coefficient which determines the alikeness between two images, has been adapted to formulate a loss function named Dice Loss [106]. Dice coefficient has a value ranging from zero to one, and one minus the Dice Coefficient value equals the Dice Loss as given by equation 3.8.

$$\text{Dice Loss} = 1 - \frac{2 \times Q \times S + 1}{Q + S + 1} \tag{3.8}$$

where, S denotes the network's predicted value and Q denotes the sample's actual label. The addition of numerical value 1 to the numerator and to the denominator is to assure that the function does not get undefined in case when Q = S = 0.

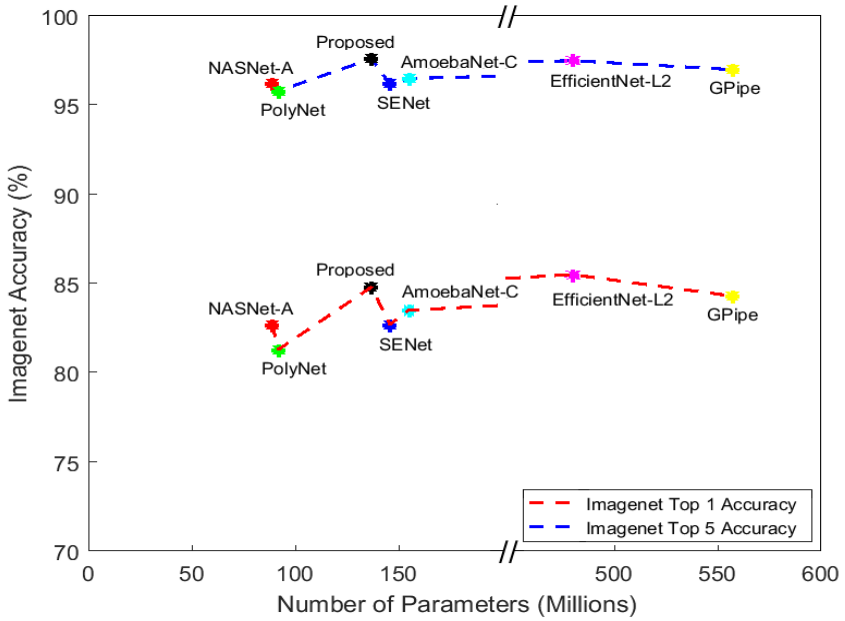


Fig. 3.5. Model parameters vs. Imagenet top1 and top 5 accuracy for NASNet-A, PolyNet, SENet, AmoebaNet-C, EfficientNet-L2 and GPipe as reported in [107], [108], [77], [109], [110] and [111] respectively.

The implementation of the proposed model has been carried out using python with OpenCV 3.6. The work has been executed on a computer having i5 processor with 8 GB RAM and a NVIDIA GeForce GTX 1060 graphics card with 6 GB VRAM to accelerate the training process. Fig. 3.5 depicts that the proposed model provide higher Imagenet Top1 and Top 5 accuracy with comparatively lower number of parameters.

3.2 Result

3.2.1 Dataset Used

To analyze the reliability of the proposed technique, the task of OD segmentation is carried out on seven public databases namely MESSIDOR, DRIVE, CHASE-DB1, DIARETDB0, DIARETDB1, IDRiD and STARE. The colour fundus photographs from all the databases are scaled to 512×512 pixels.

3.2.2 Performance Evaluation

The evaluation of the suggested framework has been appraised utilizing metrics namely Dice Coefficient (DC), Accuracy (Ac) and Sensitivity (Se) which are mathematically given by equations 3.9-3.11.

$$DC = \frac{2\times TP}{2\times TP+FP+FN} \tag{3.9}$$

$$Ac = \frac{TP+TN}{TP+FP+TN+FN} \tag{3.10}$$

$$Se = \frac{TP}{TP+FN} \tag{3.11}$$

where,

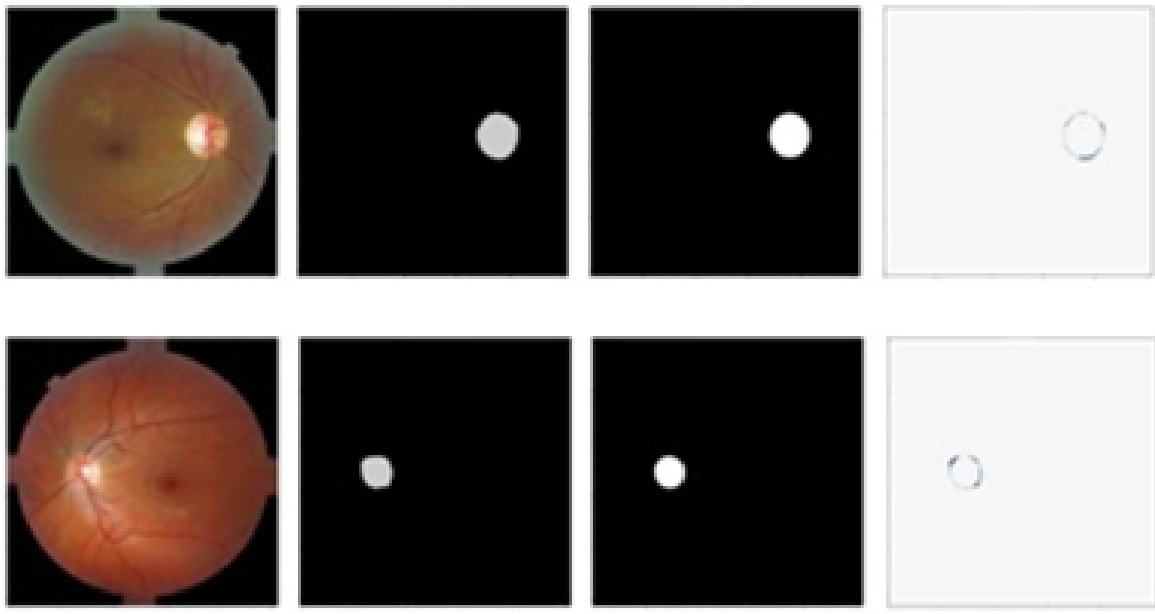
TP signifies the pixels within the OD boundary that are correctly recognized.

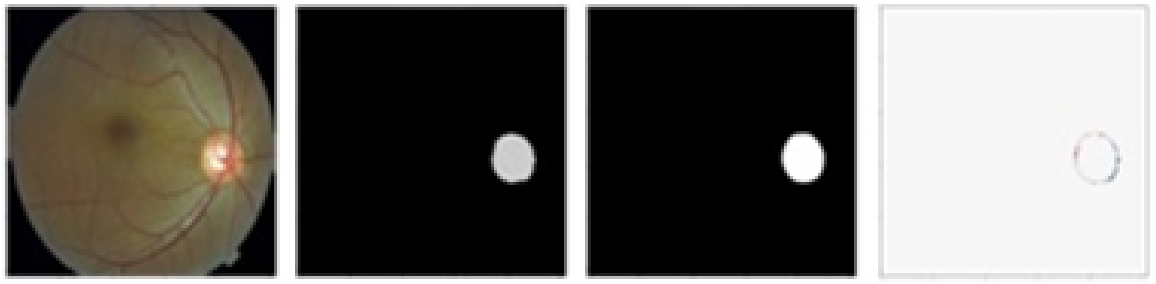
TN signifies the pixels outside the OD boundary that are correctly recognized.

FP signifies the non OD pixels that have been recognized as OD pixels.

FN signifies the OD pixels that have been recognized as non OD pixels.

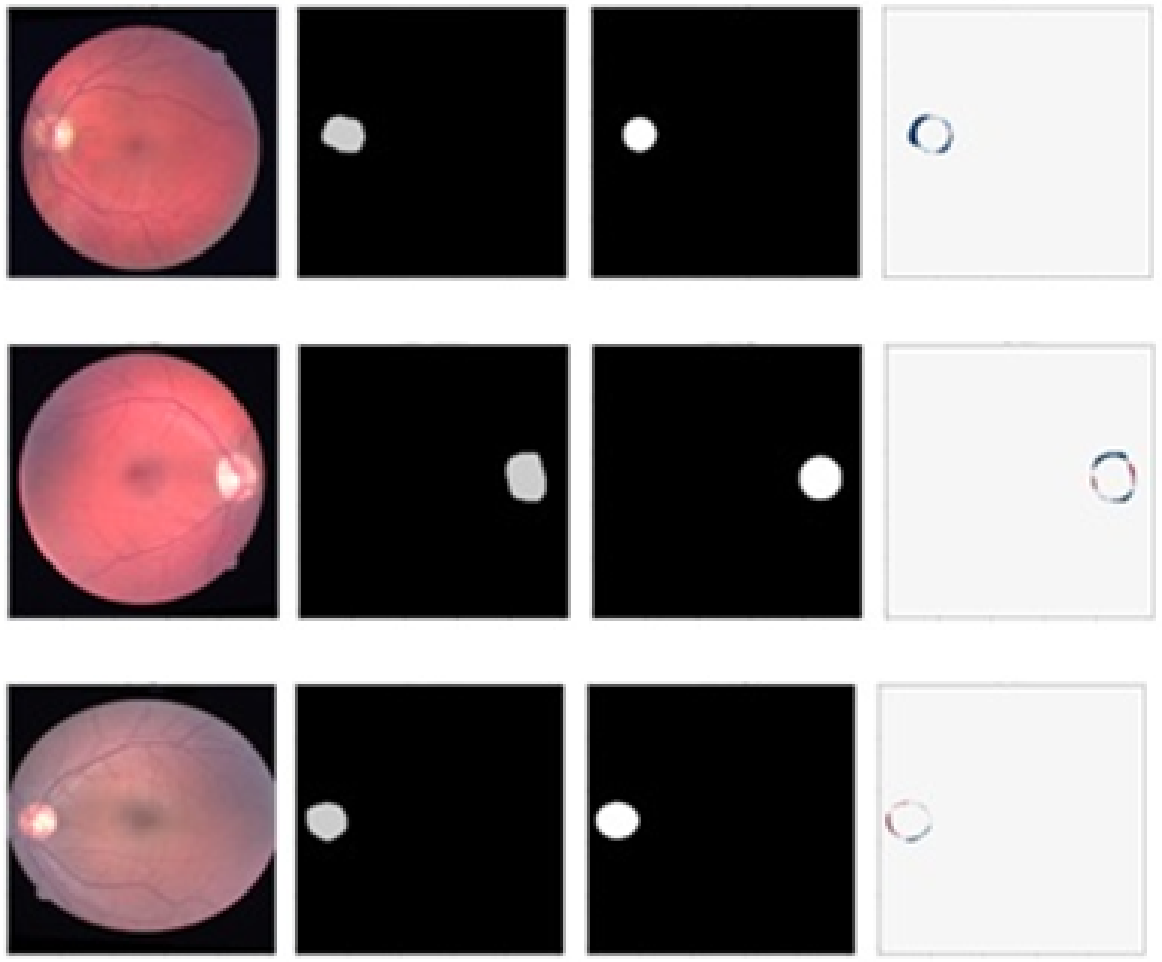
3.2.3 Experimental Outcomes





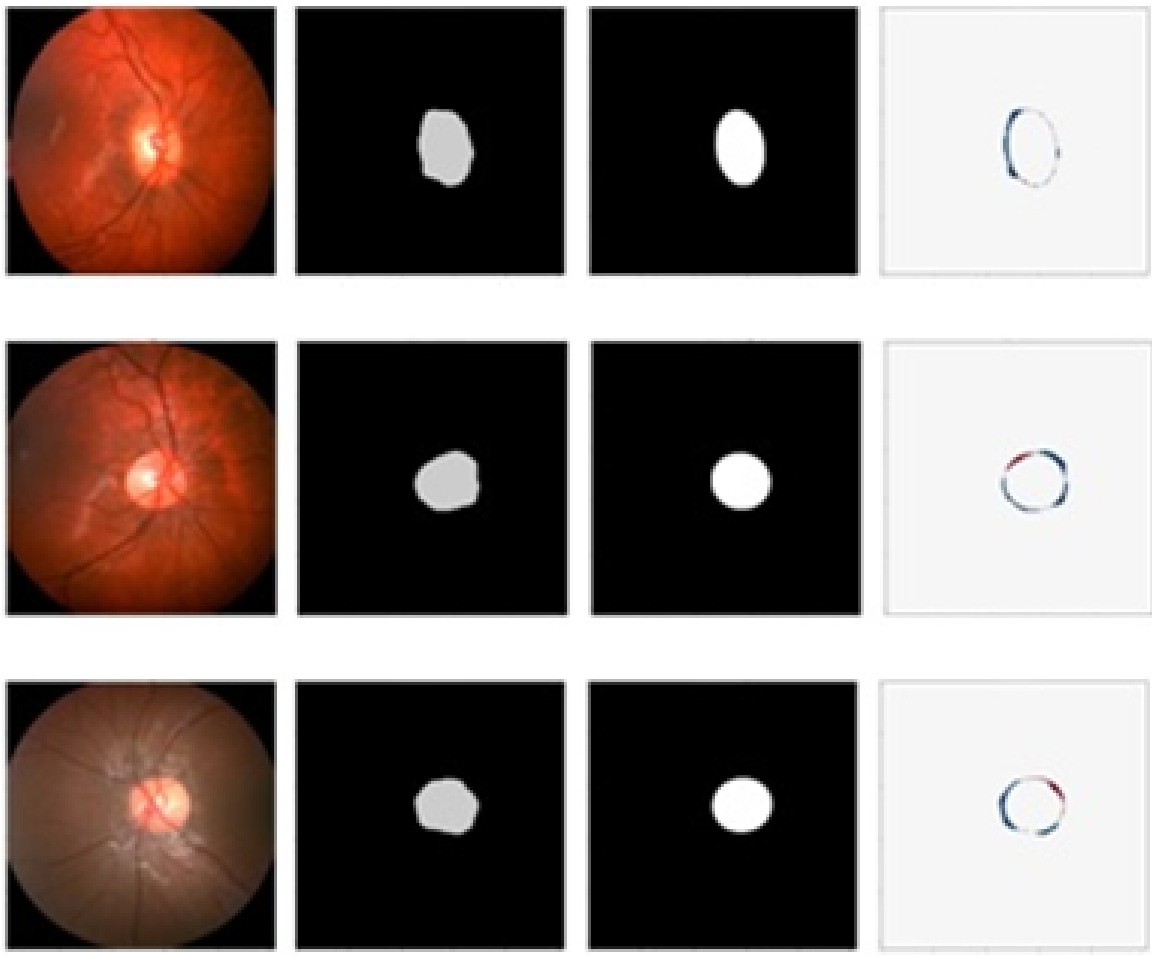
(a) (b) (c) (d)

Fig. 3.6. (a) Fundus photograph, (b) predicted optic disc mask by the proposed framework, (c) ground truth mask of the optic disc as identified by the experts and (d) error in the prediction while considering MESSIDOR dataset.



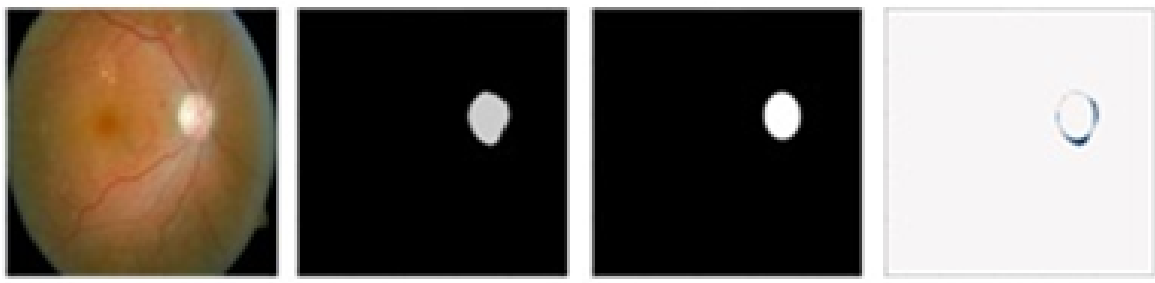
(a) (b) (c) (d)

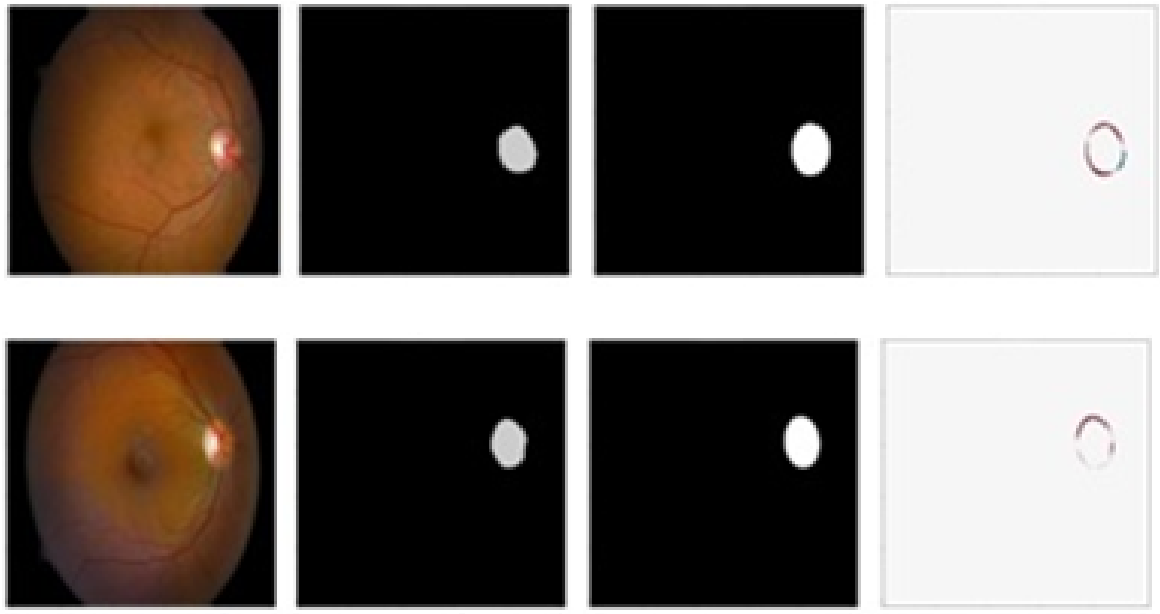
Fig. 3.7. (a) Fundus photograph, (b) predicted optic disc mask by the proposed framework, (c) ground truth mask of the optic disc as identified by the experts and (d) error in the prediction while considering DRIVE dataset.



(a) (b) (c) (d)

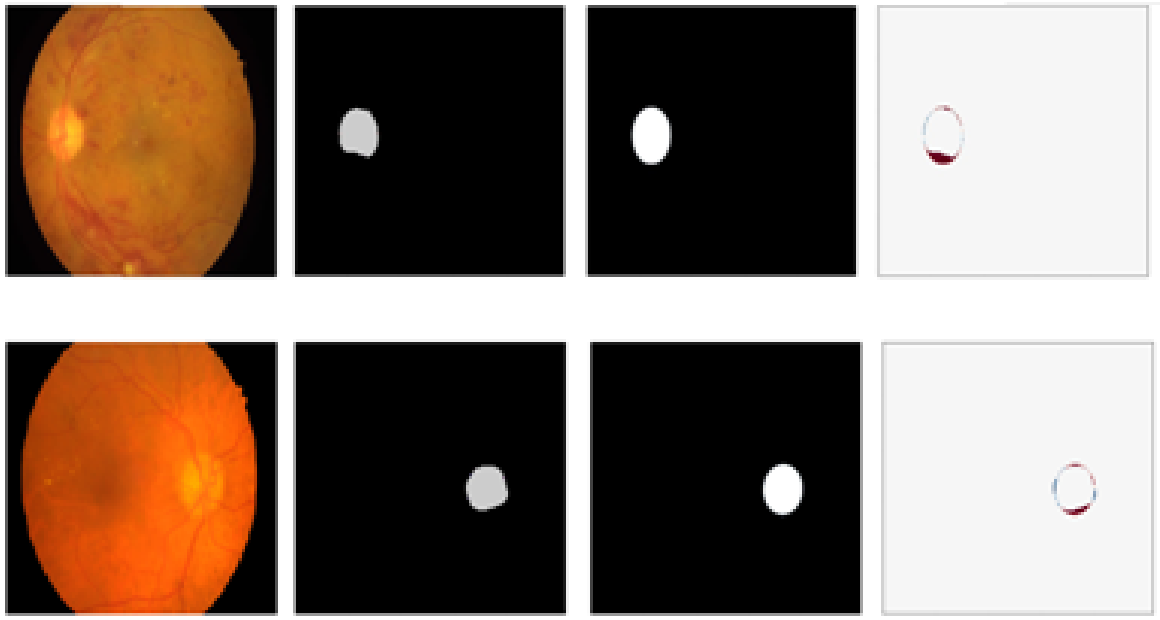
Fig. 3.8. (a) Fundus photograph, (b) predicted optic disc mask by the proposed framework, (c) ground truth mask of the optic disc as identified by the experts and (d) error in the prediction while considering CHASE-DB1dataset.

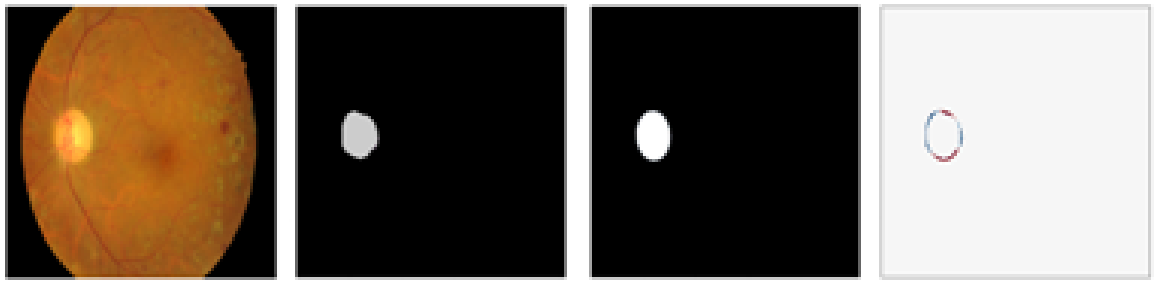




(a) (b) (c) (d)

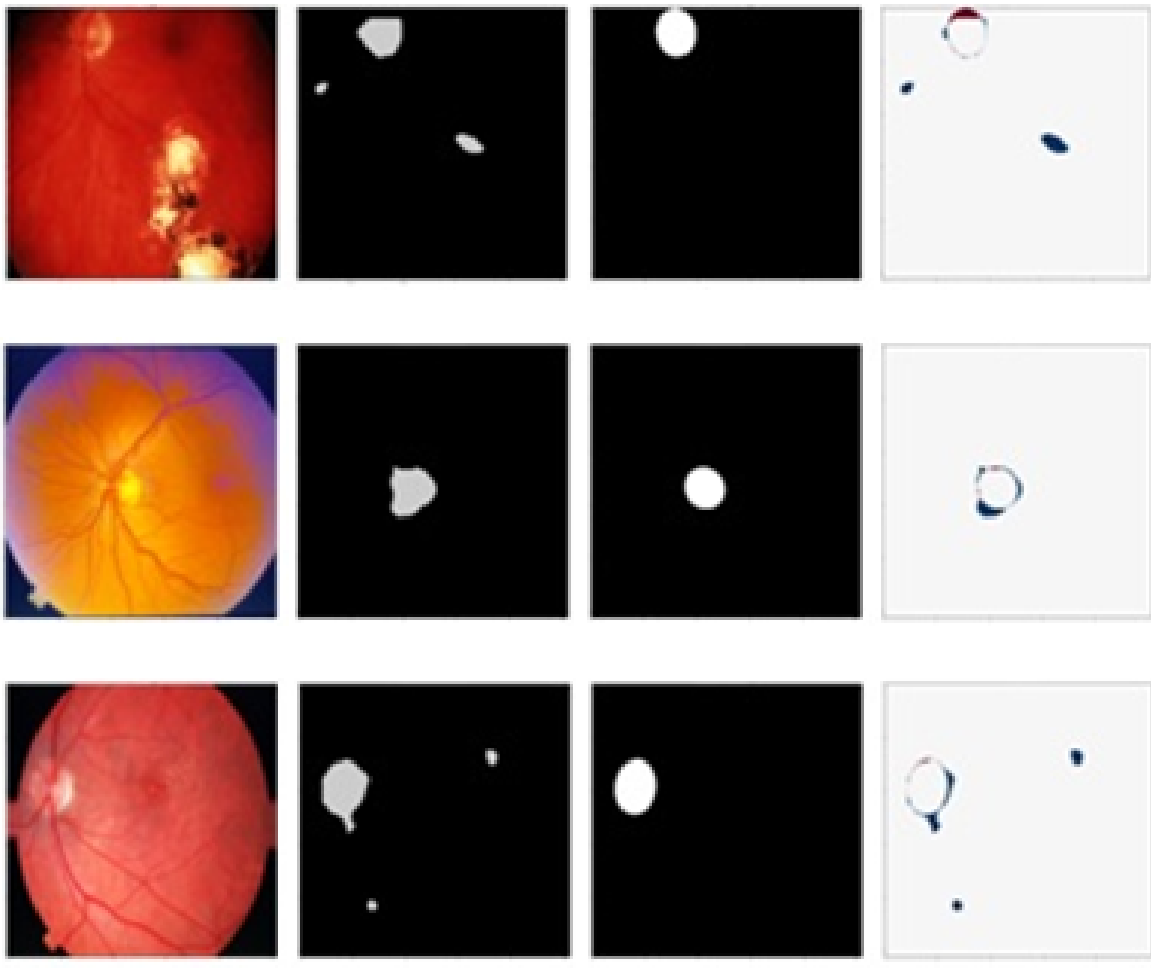
Fig. 3.9. (a) Fundus photograph, (b) predicted optic disc mask by the proposed framework, (c) ground truth mask of the optic disc as identified by the experts and (d) error in the prediction while considering DIARETDB1 dataset.





(a) (b) (c) (d)

Fig. 3.10. (a) Fundus photograph, (b) predicted optic disc mask by the proposed framework, (c) ground truth mask of the optic disc as identified by the experts and (d) error in the prediction while considering IDRiD dataset.



(a) (b) (c) (d)

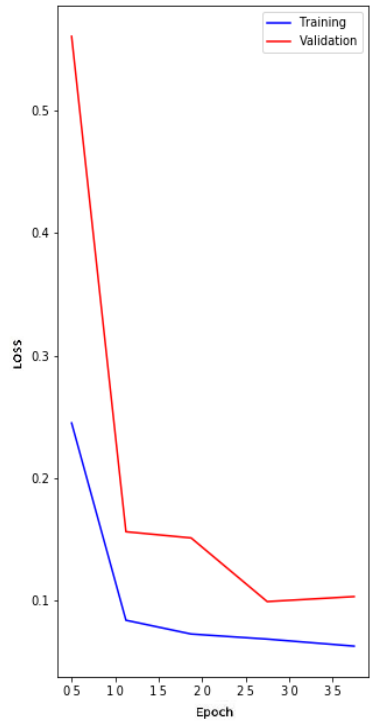
Fig. 3.11. (a) Fundus photograph, (b) predicted optic disc mask by the proposed framework, (c) ground truth mask of the optic disc as identified by the experts and (d) error in the prediction while considering STARE dataset.

A few typical sample of the OD segmentation result obtained while experimenting with MESSIDOR, DRIVE, CHASE-DB1, DIARETDB1, IDRiD and STARE dataset are shown in fig. 3.6, 3.7, 3.8, 3.9, 3.10 and 3.11 respectively. The areas that have been mistakenly identified as OD area and that which have not been tracked by the proposed methodology are shown in fig. 3.6(d), 3.7(d), 3.8(d), 3.9(d), 3.10(d) and 3.11(d). The robustness of the proposed work is confirmed by analyzing retinal images of various resolutions, intensity and pathologies.

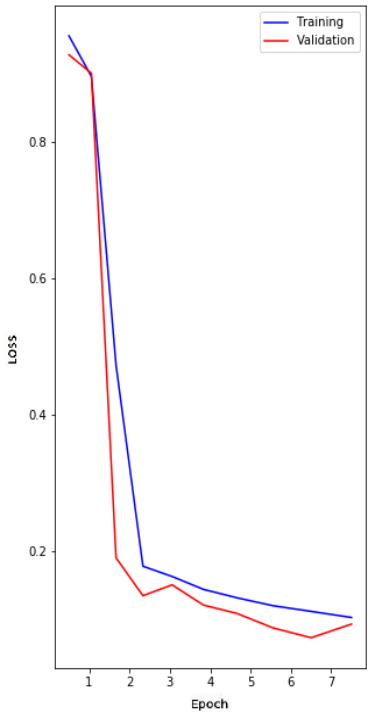
Table 3.1. Performance measure of the proposed algorithm

Algorithm	**Metric**	**Dataset**						
		MESSI-DOR	**DRIVE**	**CHASE-DB1**	**DIARET-DB0**	**DIARET-DB1**	**IDRiD**	**STARE**
CLSTM Enriched Proposed Network	DC	0.9373	0.9138	0.9110	0.9105	0.8809	0.9214	0.7776
	Ac	0.9992	0.9984	0.9937	0.9929	0.9906	0.9972	0.9890
	Se	0.9659	0.9526	0.9452	0.9487	0.9405	0.9518	0.8563
	Loss	0.0627	0.0862	0.089	0.0895	0.1191	0.0786	0.2224
Proposed Network without CLSTM	DC	0.9028	0.9075	0.8921	0.8978	0.8793	0.8952	0.7492
	Ac	0.9896	0.9907	0.9768	0.9795	0.9712	0.9849	0.9483
	Se	0.9274	0.9351	0.9237	0.9128	0.9053	0.9187	0.7846
	Loss	0.0972	0.0925	0.1079	0.1022	0.1207	0.1048	0.2508
FCN with Resnet-50 architecture	DC	0.8831	0.8715	0.8589	0.8504	0.8463	0.8791	0.7132
	Ac	0.9435	0.9387	0.9198	0.9082	0.9016	0.9326	0.8853
	Se	0.9186	0.9043	0.9017	0.8977	0.8865	0.9072	0.7594
	Loss	0.1169	0.1285	0.1411	0.1496	0.1537	0.1209	0.2868

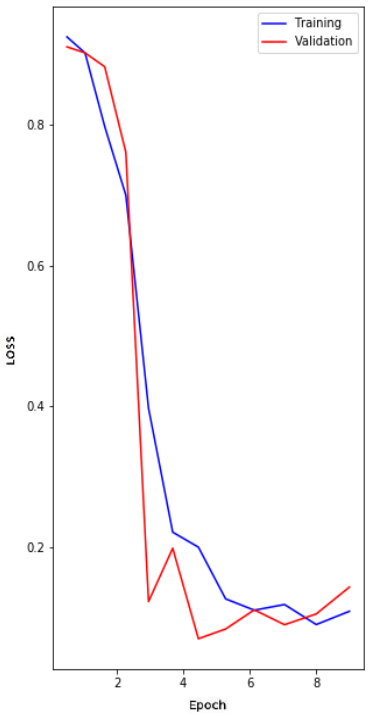
The Dice Coefficient (DC), Accuracy (Ac) and Sensitivity (Se) for each database recorded in table 3.1 assess the efficacy of the model. This table confirms the superiority of the proposed framework by depicting higher DC, Ac and Se values over other models.



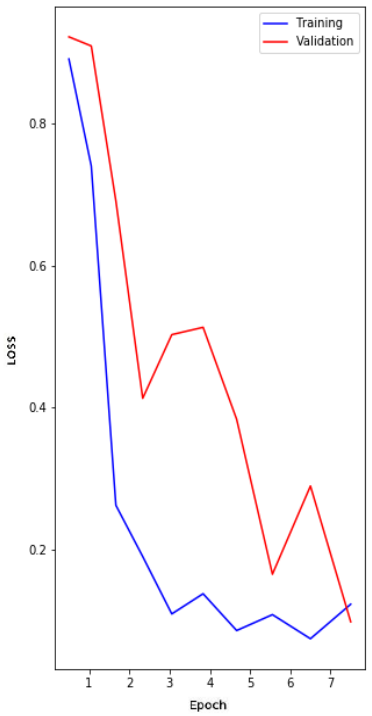
(a)



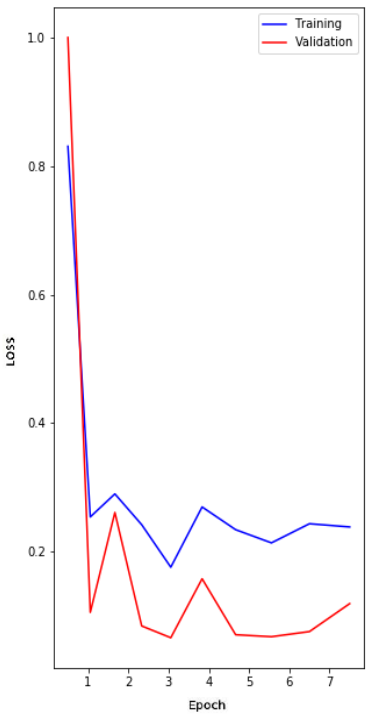
(b)



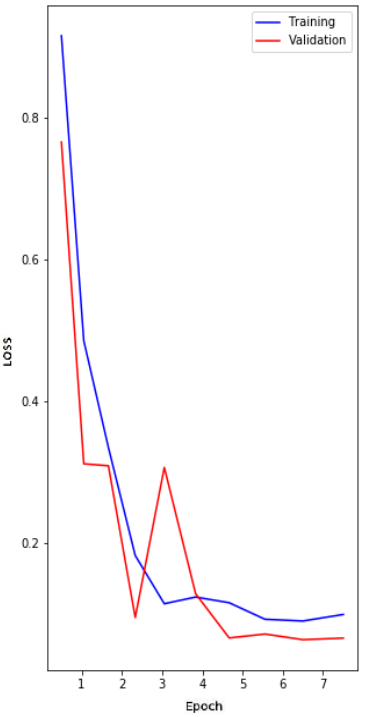
(c)



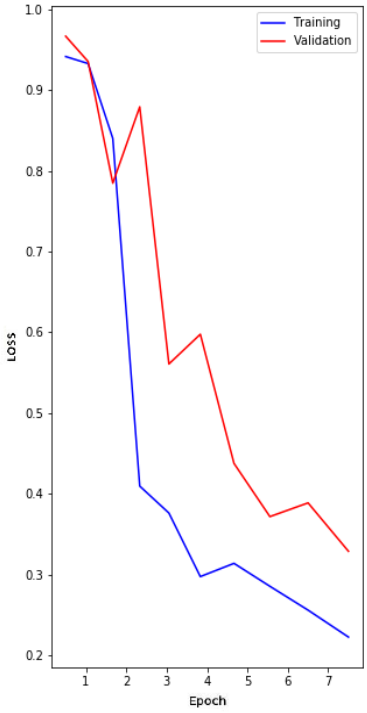
(d)



(e)



(f)



(g)

Fig. 3.12. Training loss curves (denoted by blue colour) and validation loss curves (denoted by red colour) of the proposed network considering (a) MESSIDOR (b) DRIVE (c) CHASE-DB1 (d) DIARETDB0 (e) DIARETDB1 (f) IDRiD and (g) STARE dataset.

The training and the validation loss curves are presented in fig. 3.12. A comparative study carried out between the suggested technique and the other existing approaches considering the dice coefficient, accuracy and sensitivity values as parameters is provided in table 2. The dice coefficient and sensitivity are considered more valuable metrics compared to accuracy in case when segmentation task is evaluated. Table 3.2 illustrates that the proposed algorithm outperforms the benchmark techniques when both the dice coefficient and sensitivity values are assessed. Although in some dataset the segmentation accuracy is slightly less than that of the existing works, but it is comparable and considerably high value for all the datasets. A comparison between the dice coefficient and dice loss values recorded while experimenting on the seven encoder architectures namely VGG11, VGG13, VGG16, VGG19, Resnet34, Densenet121 and InceptionV3 is presented in table 3. Among these architectures, VGG16 has been found to provide the best segmentation results by delivering minimum loss on all the test datasets. A graphical representation of the loss values revealed in table 3.3 is depicted in fig. 3.13. The study outcomes authenticate that the suggested technique for OD segmentation is more reliable and efficient in processing retinal images.

Table 3.2. A comparative study between the suggested network and the other existing approaches

Dataset	Algorithm	DC	Ac	Se
MESSIDOR	Rehman et al. [112]	0.8510	0.9880	0.9530
	Morales et al. [113]	0.8950	0.9949	-
	Roychowdhury et al. [114]	-	0.9956	0.9043
	Abdullah et al. [115]	0.9339	0.9989	0.8954
	Fan et al. [116]	0.9196	0.9770	-
	Zahoor et al. [117]	-	0.9918	0.8891
	Nija et al. [118]	-	0.9993	0.9043
	Proposed	0.9373	0.9992	0.9659
DRIVE	Morales et al. [113]	0.8169	0.9903	-
	Roychowdhury et al. [114]	-	0.9910	0.8780
	Abdullah et al. [115]	0.8720	0.9672	0.8187
	Zahoor et al. [117]	-	0.9980	0.8309
	Hasan et al. [23]	-	0.9990	-
	Proposed	0.9138	0.9984	0.9518
CHASE-DB1	Roychowdhury et al. [114]	-	0.9914	0.8962
	Abdullah et al. [115]	0.9050	0.9579	0.8313
	Ramani et al. [120]	0.8211	0.9900	-
	Proposed	0.9110	0.9937	0.9405
DIARETDB0	Roychowdhury et al. [114]	-	0.9956	0.8660
	Abdullah et al. [119]	-	0.9965	0.8745
	Nija et al. [118]	-	0.9969	0.8946
	Proposed	0.9105	0.9929	0.9487
DIARETDB1	Roychowdhury et al. [114]	-	0.9963	0.8815
	Abdullah et al. [119]	-	0.9768	0.8463
	Nija et al. [118]	-	0.9968	0.9374
	Proposed	0.8809	0.9906	0.9452
STARE	Roychowdhury et al. [114]	-	0.9854	0.8380
	Proposed	0.7776	0.9890	0.8563

Table 3.3. A comparison between the different encoder frameworks

Metric	Encoder Network	Dataset						
		MESSI-DOR	**DRIVE**	**CHASE-DB1**	**DIARET-DB0**	**DIARET-DB1**	**IDRiD**	**STARE**
Dice Co-efficient (DC)	VGG11	0.8521	0.8632	0.8576	0.8697	0.8504	0.8498	0.7195
	VGG13	0.9023	0.8947	0.8809	0.8932	0.8721	0.9007	0.7428
	VGG16	0.9373	0.9138	0.911	0.9105	0.8809	0.9214	0.7776
	VGG19	0.8653	0.8892	0.8684	0.8786	0.8623	0.8575	0.7329
	Resnet 34	0.8905	0.9004	0.8892	0.8974	0.8748	0.8993	0.7581
	Densenet 121	0.8831	0.8894	0.8745	0.8755	0.8679	0.8781	0.7387
	Inception V3	0.8773	0.8931	0.8718	0.8827	0.8605	0.8652	0.7264
Loss	VGG11	0.1479	0.1368	0.1424	0.1303	0.1496	0.1502	0.2805
	VGG13	0.0977	0.1053	0.1191	0.1068	0.1279	0.0993	0.2572
	VGG16	0.0627	0.0862	0.089	0.0895	0.1191	0.0786	0.2224
	VGG19	0.1347	0.1108	0.1316	0.1214	0.1377	0.1425	0.2671
	Resnet 34	0.1095	0.0996	0.1108	0.1026	0.1252	0.1007	0.2419
	Densenet 121	0.1169	0.1106	0.1255	0.1245	0.1321	0.1219	0.2613
	Inception V3	0.1227	0.1069	0.1282	0.1173	0.1395	0.1348	0.2736

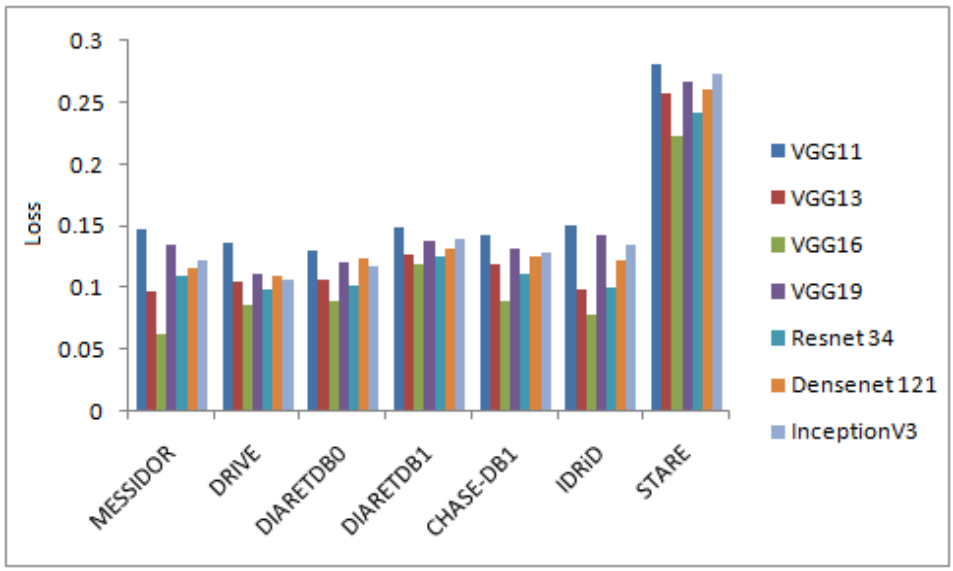


Fig. 3.13. Graphical representation of the dice loss values obtained while dealing with different encoder frameworks.

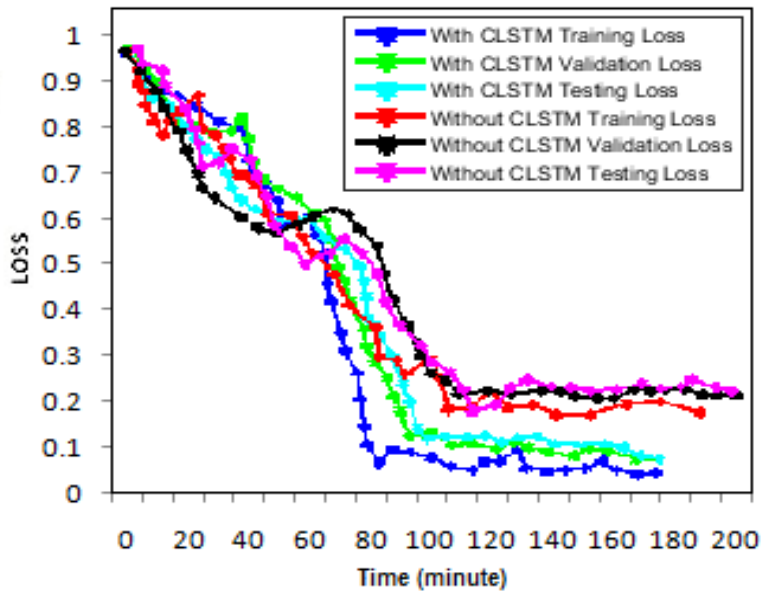


Fig. 3.14. Training loss, validation loss, and testing loss curves of the proposed CLSTM incorporated network and for the same model without CLSTM with 40 epochs.

A comparison between the training loss, validation loss and testing loss curves of the proposed CLSTM incorporated network and the same model without CLSTM is shown in fig. 3.14. The time in minutes is represented along the horizontal axis and loss value is

represented along the vertical axis. The models are trained for 40 epochs. The loss value and the time for each iteration are recorded. The obtained curves show that the CLSTM incorporated architecture completes all the 40 iterations faster than the other model without CLSTM. Fig. 3.14 also depict that the loss curves of CLSTM enriched model decreases faster than the other, inferring that CLSTM incorporated network converges faster.

3.3 Summary

This study illustrates a deep-learning technique for detecting and segmenting the OD in fundus photographs automatically. The proposed framework has been found to deliver promising result in case where there is a lack of annotated samples. Incorporation of spatiotemporal information and use of pre-trained weights in the encoder of the framework showed improvement in convergence process. It has been noticed that models using initial weights obtained from pre-trained networks helps in saving the training time and also prevents over-fitting compared to that using randomly initialized weights. In this work, efficiency of the proposed network has been enhanced by fine-tuning the initial weights, in the encoder of the network. The task of fine-tuning a network is presumed to be highly challenging when segmentation of biomedical images comes into account due to insufficient availability of labeled training data. However, the incorporation of CLSTM has enriched the architecture of model by identifying the spatial characteristics. Due to the structural supremacy, the CLSTM incorporated architecture provides superior performance compared to the traditional architectures. It has also been noticed that CLSTM provides better segmentation accuracy by considering the information of neighborhood pixels.

Chapter 4

Segmentation of the Exudates

The determination of the exact dimension of exudates in the fundus photographs is a vital factor in the treatment of DR, and the position and the count of the lesion helps to assess the stage of the DR. Three strategies for automatically identifying and segmenting exudates in retinal pictures have been developed in this study. The first and second approach deals with individual spatial attention technique and channel attention technique for segmenting exudates, respectively, while the third approach employ the CLSTM methodology, RES unit, and the combined spatial and channel attention mechanism. These models can be utilized as a clinical aid for diagnosing DR in mass screening programs. The outcomes of the proposed models revealed gradual improvements with the evolution of the segmentation techniques and indicated their reliability when assessed on several public retinal databases. Thus, the developed algorithms can assist the ophthalmologist in making the decisions.

4.1 Methodology Adopted for the Segmentation of Exudates Utilizing Spatial Attention Mechanism

An encoder-decoder framework along with spatial attention techniques is being developed in this study for improving the segmentation operation. This model has been designed to retrieve salient features by recovering spatial information in fundus images. The main characteristic of attention mechanism is to stress on the important features while discarding

unnecessary information. The attention module is combined with convolutional neural network to speed up the learning process, to extract more crucial features and adapt to small training datasets.

The layout of the suggested framework has been presented in fig. 4.1. The decoder and encoder section of the framework comprises of convolutional block, pooling block, up-sampling block and attention block. Every convolutional block contains convolution layer, batch normalization layer, and rectified unit (ReLU). A spatial attention mechanism is implemented in the decoder section. The count of feature channels doubles in the encoder section, while in the decoder section it gets half. The skip connections link the decoder and encoder through feature maps of the subsequent layers. The last layer generates the segmentation maps, by performing 1x1 convolution and applying sigmoid activation function.

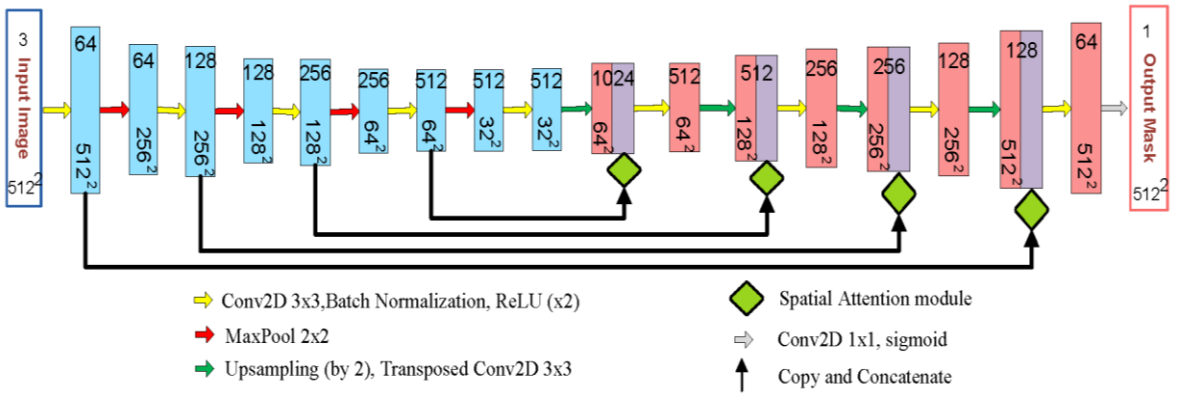


Fig. 4.1. The proposed spatial attention enhanced network

The up-sampling step combines the feature maps retrieved from the encoder block via the skip connections, with the spatial information from the low-resolution decoder block. This paper proposes a spatial attention module which plays a vital role to recover several valuable features. The attention module focuses on the various shapes of the object under consideration. In a fundus image, this architecture along with attention mechanism learns to emphasize important features required for proper recognition of the object while suppressing irrelevant regions. The attention module creates a spatial attention map (m) by processing the features (y_e) from the encoder and the corresponding features (y_d) from the decoder. The refined encoder features (y^*_e) is given by equation 4.1.

$$y^*_e = y_e \otimes m(y_e, y_d) \tag{4.1}$$

where, element wise multiplication is represented by ⊗.

4.1.1 Spatial Attention Module

This attention module which focuses on position based attributes in images, consider the spatial correlation between the input features. As the encoder feature contains a lot of position information, it can be used to concentrate on the regions that are most useful for determining the object's location and target structure. Fig. 4.2 shows the layout of the spatial attention module. The spatial attention map is developed by merging the spatial attributes from both the decoder and the encoder features. Average and max pooling operation, inspired by [78], and 1×1 convolution operation, inspired by [71], is applied to generate the spatial map. The maps from every decoder and encoder feature are then concatenated to produce a significant feature representation. A convolution operation is then implemented on the merged features for developing a final map m(y) which highlights the salient region.

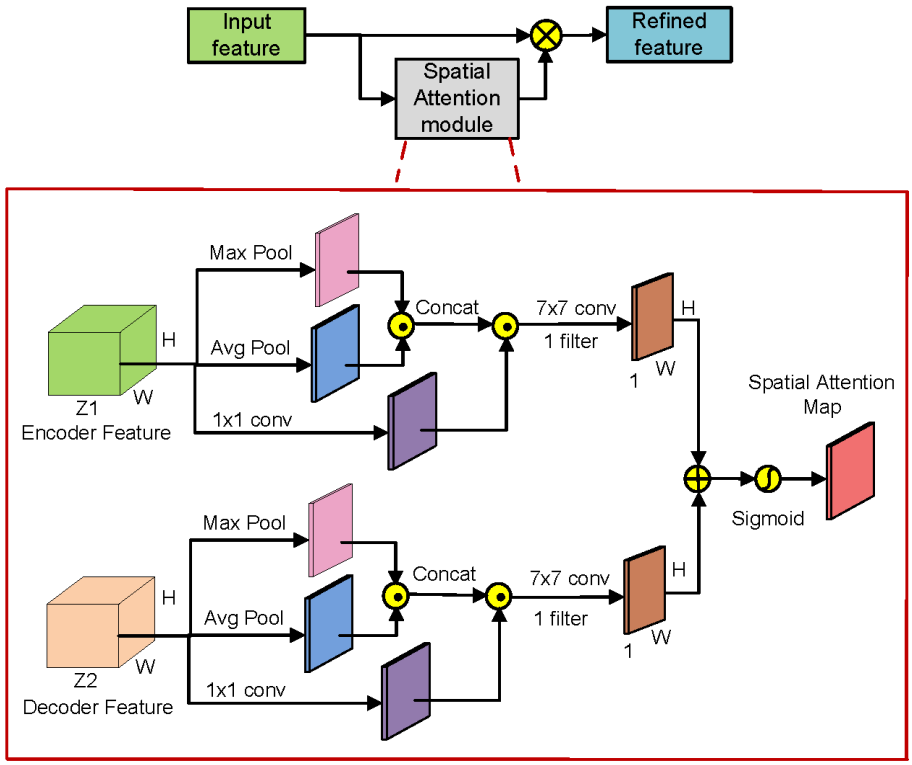


Fig. 4.2. Structure of spatial attention mechanism

The maps obtained from average pooling, max pooling and 1×1 convolution operations are represented by y_{avg}, y_{max}, and $y_{1\times1}$ respectively. The final spatial map is created by

concatenating the obtained maps and using a convolution operation with a large kernel size 7×7 for effectively capturing long-range contextual information, inspired by [78]. The spatial maps of the decoder features and the encoder features are represented by $m_d(y_d)$ and $m_e(y_e)$ respectively. A sigmoid operation is applied to the sum of $m_d(y_d)$ and $m_e(y_e)$ to finalize the development of a spatial map $m(y_e, y_d)$. The corresponding operations for computing spatial attention map are given by the equations 4.2-4.4.

$$\begin{aligned} m_e(y_e) &= y_1^{7\times7}\left(\left[\text{avg-pool}(y_e), \text{max-pool}(y_e), y_1^{1\times1}(y_e)\right]\right) \\ &= y_1^{7\times7}\left(\left[(y_e)_{avg}, (y_e)_{max}, (y_e)_{1\times1}\right]\right) \end{aligned} \tag{4.2}$$

$$\begin{aligned} m_d(y_d) &= y_1^{7\times7}\left(\left[\text{avg-pool}(y_d), \text{max-pool}(y_d), y_1^{1\times1}(y_d)\right]\right) \\ &= y_1^{7\times7}\left(\left[(y_d)_{avg}, (y_d)_{max}, (y_d)_{1\times1}\right]\right) \end{aligned} \tag{4.3}$$

$$m(y_e, y_d) = \sigma\left(m_e(y_e) + m_d(y_d)\right) \tag{4.4}$$

where, σ signifies sigmoid operation and $y_q^{p\times p}$ represents q filters of p×p convolution operation.

4.1.2 Result

4.1.2.1 Dataset Used

To determine the robustness of the proposed methodology, it is tested on four datasets namely DIARETDB0 [6], DIARETDB1 [7], MESSIDOR [4], and IDRiD [9]. The images of each dataset are resized to 512×512 pixel.

4.1.2.2 Performance Evaluation Metric

The efficiency of the developed approach is appraised using accuracy (Ac), specificity (Sp), and sensitivity (Se) as evaluation metrics. The metrics are expressed as:

$$Ac = \frac{TP+TN}{TP+FP+TN+FN} \tag{4.5}$$

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

$$Se = \frac{TP}{TP+FN} \tag{4.7}$$

where, the correctly recognized object and background pixels have been indicated by TP and TN respectively, and the incorrectly identified object and background pixels have been indicated by FP and FN respectively.

4.1.2.3 Network Implementation

The proposed framework is accomplished utilizing Python and OpenCV 3.6 is used to carry out the training and test procedures. The Adam optimizer is considered in this work while training. The batch size and learning rate are fixed to 8 and 0.0001 respectively. The Dice Loss (DL) given by equation 4.9 is employed as loss function in this model. The training and validation loss curve obtained in this study is shown in fig. 4.3.

$$\text{Dice Score (DC)} = \frac{2\times TP}{2\times TP+FN+FP} \tag{4.8}$$

$$\text{Dice Loss (DL)} = 1\text{- Dice Score} \tag{4.9}$$

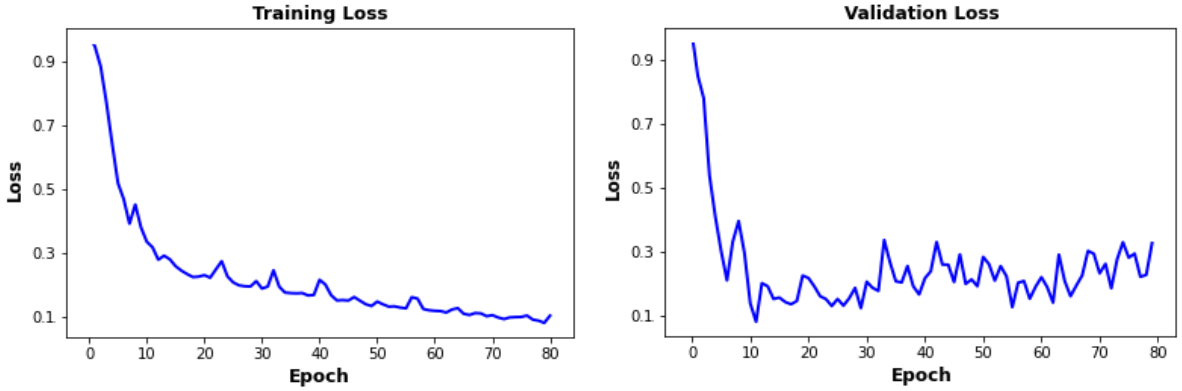


Fig. 4.3. Training and validation loss curve of the spatial attention enhanced model

4.1.2.4 Experimental Outcomes

The exudate segmentation outcomes achieved from the suggested approach is presented in fig. 4.4. The experimental findings depict that the predicted object is very identical to the ground truth. This proves the efficiency of the work and its acceptability as a computer application for identifying the shape, size and position of exudates in fundus images.

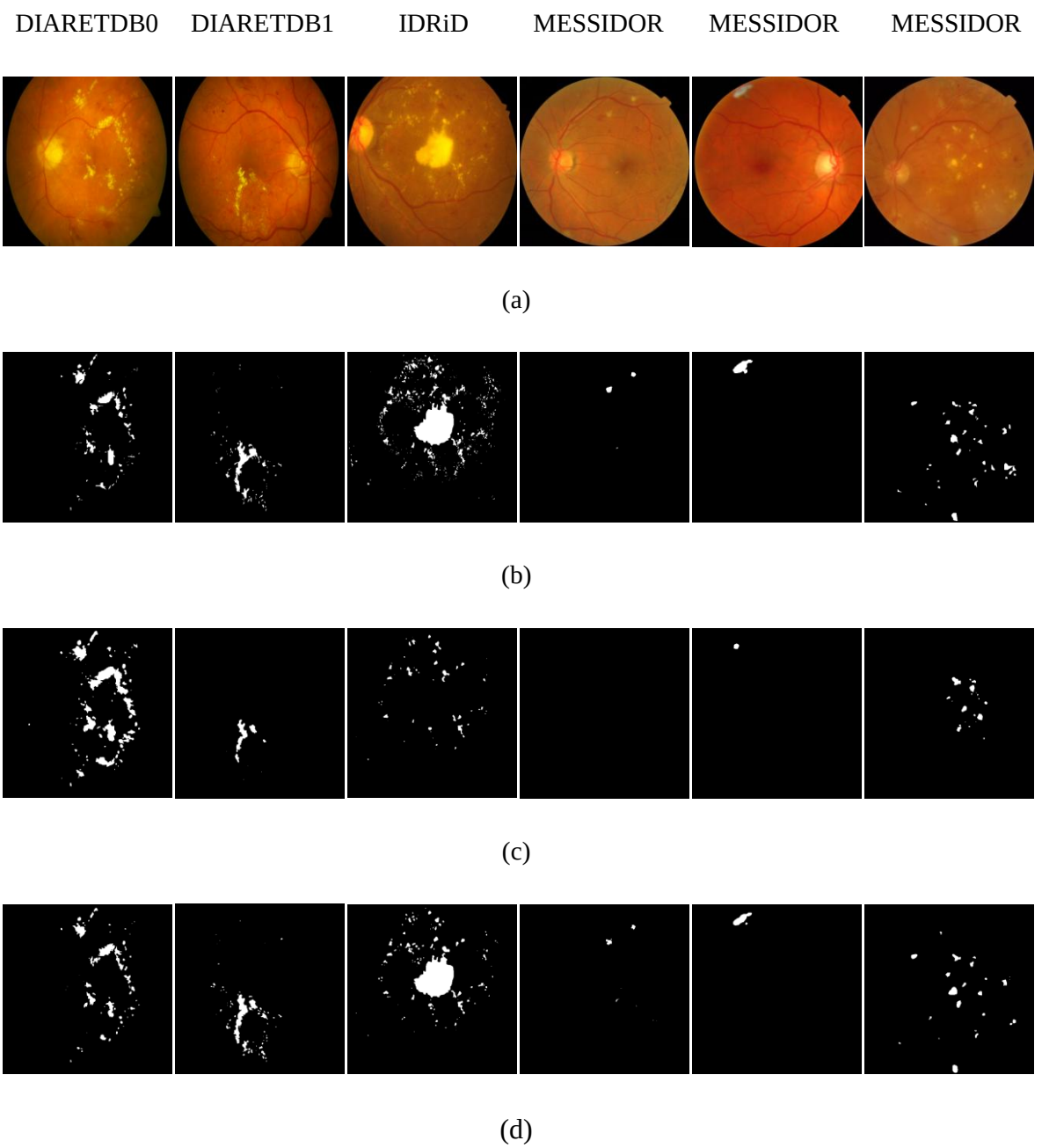


Fig. 4.4. (a) Fundus photograph, (b) Labelled image for exudate, (c) U-Net segmented image and (d) Prediction done by the suggested approach utilizing spatial attention mechanism

Table 4.1. Comparison of performance of the proposed model utilizing spatial attention mechanism

Dataset	Algorithm	Ac (%)	Sp (%)	Se (%)
DIARETDB0	Akram et al. [121]	96.48	98.38	93.7
	Lokuarach-chi et al. [122]	-	90	93.64
	Proposed	97.23	98.52	95.26
DIARETDB1	Akram et al. [121]	98.92	98.35	99.65
	Lokuarach-chi et al. [122]	-	88.46	94.36
	Yazid et al. [123]	-	97.4	98.2
	Welfer et al. [124]	-	98.84	70.48
	Liu et al. [125]	79	75	83
	Mateen et al. [126]	98.72	-	97
	Proposed	98.87	99.16	98.92
MESSIDOR	Zhang et al. [127]	-	-	62.3
	Kaur et al. [128]	86.00	90.47	81.32
	Proposed	91.14	90.78	91. 45
IDRiD	Proposed	92.69	93.27	92.21

Table 4.2. Performance comparison based on dice score

Algorithm	DIARETDB0	DIARETDB1	MESSIDOR	IDRiD
U-Net [21]	81.93%	82.65%	79.21%	85.37%
Proposed	90.15%	89.71%	91.03%	90.64%

The outcomes reported in table 4.1 show the potentiality of this technique utilizing spatial attention mechanism in surpassing the existing methodologies considering metrics namely accuracy, specificity, and sensitivity with an overall accuracy of 94.98%. A comparison of dice score achieved in this study which utilizes spatial attention mechanism for segmenting exudates, with U-Net is depicted in table 4.2.

4.2 Methodology Adopted for the Segmentation of Exudates Utilizing Channel Attention Mechanism

The proposed network, which involves a decoder-encoder architecture with a channel attention technique, is displayed in fig. 4.5. The framework includes convolutional, max-pooling, up-sampling and channel attention modules. All the convolution operation is succeeded by batch normalization and ReLU activation, with an exception in the last convolutional layer, where sigmoid activation is being used. The skip connection facilitates the framework to transmit the fine details learned in the encoder section to the decoder. The concatenation procedure used in the architecture enables both the low and the high level characteristics to be present in the final feature map. The channel attention block serves a crucial function in developing the channel attention map (m) considering the encoder characteristics (z_e) and the corresponding decoder characteristics (z_d). The equation 4.10 gives the refined encoder characteristics (z^*_e).

$$z^*_e = z_e \otimes m(z_e, z_d) \tag{4.10}$$

where, $\otimes$ signifies element-by-element multiplication process.

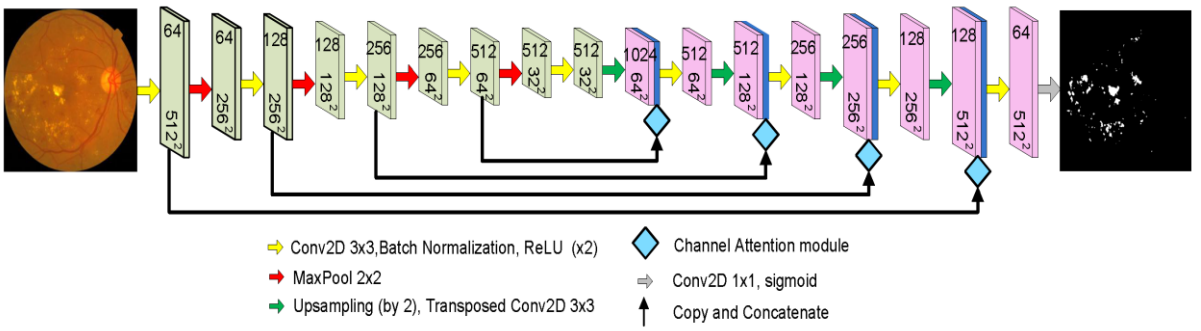


Fig. 4.5. The proposed channel attention incorporated deep network

4.2.1 Channel Attention Block

The low level characteristics acquired from the encoder have limited semantic details, whereas the high level characteristics acquired from the decoder include lots of information which can be employed to assist the low level characteristics of the encoder to collect semantic relationships. The collection can be improved by effective integration of decoder and encoder features which can be assured by enhancing the contextual details in the low level feature of the encoder. Each encoder and decoder feature is subjected to max and

average pooling, inspired by [78], to produce two feature descriptors for every channel. Then, N numbers of 1×1 convolution are applied on the feature descriptors, such that each of the 1×1 convolutions is responsible for collecting the channel relationships required to create the squeeze channel attention map. In order to minimize parameter overhead, N is considered to be equal to 1/16 of the channel counts of encoder feature. The ultimate attention map is designed by using R number of 1×1 convolution operation on the decoder and encoder squeeze channel map. R denotes encoder feature's channel count. The maps obtained from the average and max pooling operations are portrayed as z_{avg} and z_{max} respectively. Channel attention block is presented in fig. 4.6.

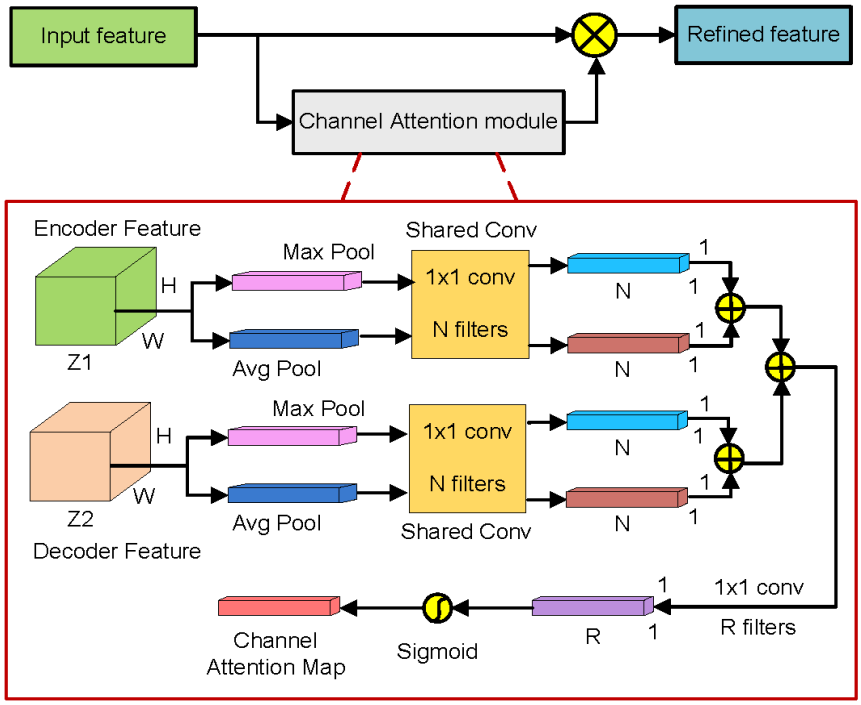


Fig. 4.6. Framework of channel attention block

The maps created as a consequence are fed through N, 1×1 convolutions to create a squeeze channel attention map. Element-wise summation is then performed to combine the obtained output maps. The squeeze channel attention map of the decoder and encoder characteristics are portrayed as $m_d(z_d)$ and $m_e(z_e)$ respectively. The summed map is conveyed through R, 1×1 convolutions, succeeded by sigmoid operation to generate the ultimate channel attention map $m(z_e, z_d)$. The subsequent steps adopted for evaluating channel attention map are provided in equations 4.11-4.13.

$$\begin{aligned} m_e(z_e) &= z_N^{1\times1}\left(\text{avg-pool}(z_e)\right) + z_N^{1\times1}\left(\text{max-pool}(z_e)\right) \\ &= z_N^{1\times1}\left((z_e)_{avg}\right) + z_N^{1\times1}\left((z_e)_{max}\right) \end{aligned} \tag{4.11}$$

$$m_d(z_d) = z_N^{1\times1}\left(\text{avg-pool}(z_d)\right) + z_N^{1\times1}\left(\text{max-pool}(z_d)\right) = z_N^{1\times1}\left((z_d)_{avg}\right) + z_N^{1\times1}\left((z_d)_{max}\right) \quad (4.12)$$

$$m(z_e, z_d) = \sigma\left(z_R^{1\times1}\left(m_e(z_e) + m_d(z_d)\right)\right) \quad (4.13)$$

where σ denotes sigmoid operator and $f_p^{q\times q}$ denotes p filters of q×q convolution operation

4.2.2 Result

4.2.2.1 Database Used and Performance Assessment Metric

The suggested approach is evaluated on four databases namely DIARETDB1 [7], IDRiD [9], MESSIDOR [4] and DIARETDB0 [6] to ascertain its robustness. The assessment metrics namely the Accuracy (Ac), Specificity (Sp), and Sensitivity (Se), given by equations 4.5-4.7, are employed to ascertain the potency of the suggested approach.

4.2.2.2 Network Implementation

Python has been chosen as the programming language to complete the work. The learning rate and batch size are specified as 0.0001 and 8 respectively. During the network training, the Dice Loss, given by equation 4.9 and the Adam has been adopted as a loss function and as an optimizer respectively. All the images have been resized to 512×512 pixels. Fig. 4.7 presents the loss curves achieved during the training and the validation process.

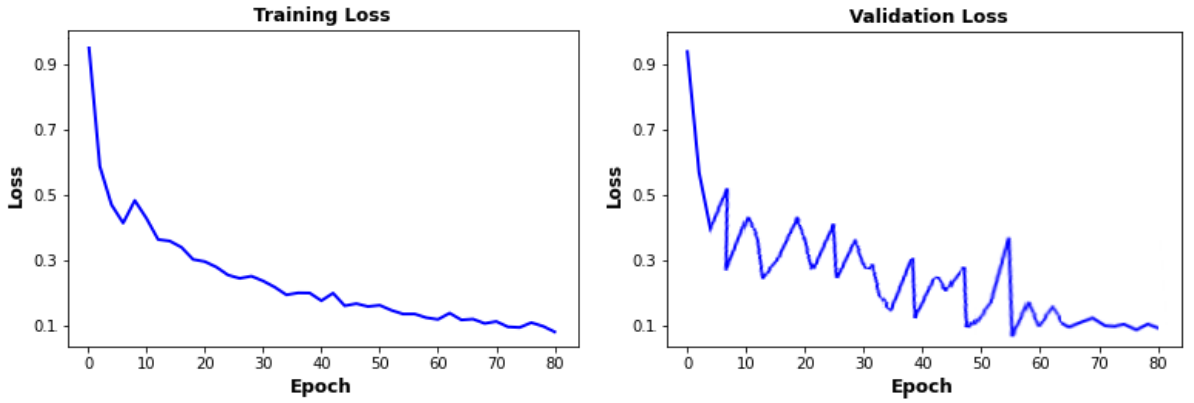
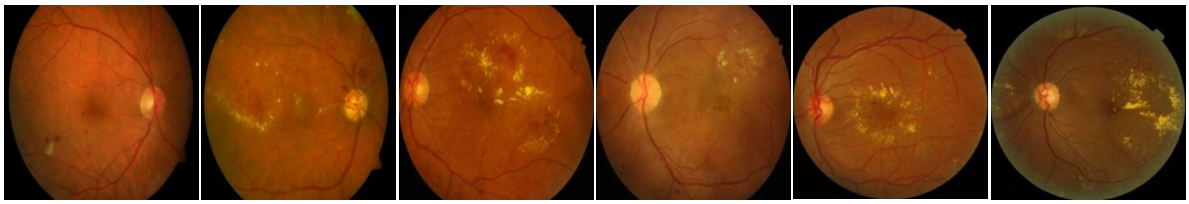


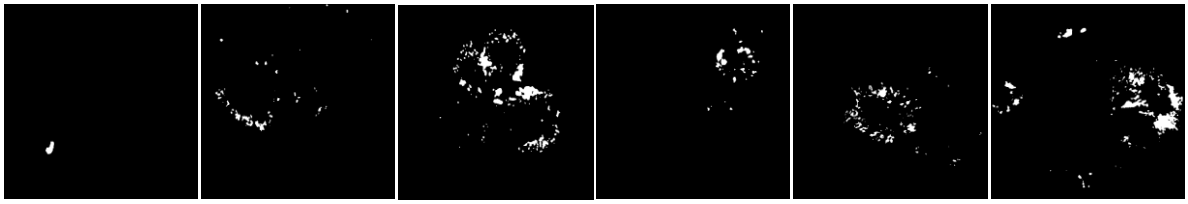
Fig. 4.7. Training and validation loss curve of the channel attention enhanced model

4.2.2.3 Experimental Outcomes

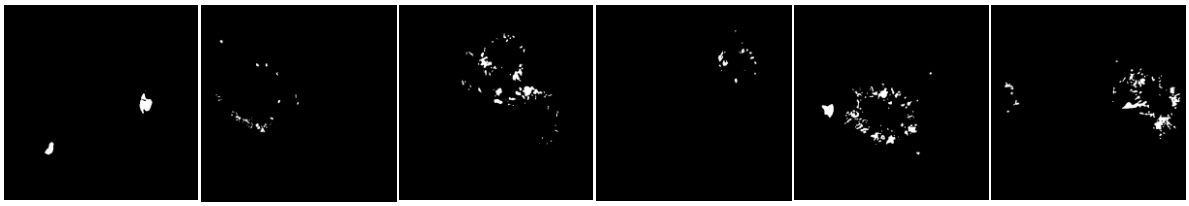
The outcomes of the proposed exudate segmenting approach displayed in fig. 4.8 reveal that the segmented object and its labeled data are remarkably similar. The significant performance enhancements illustrate the viability of the proposed methodology's potential for adoption as a computer algorithm for revealing the exert possie and dimension of pathologies in fundus images.



(a)



(b)



(c)

(d)

Fig. 4.8. (a) Fundus photograph, (b) Labelled image for exudate, (c) U-Net segmented image and (d) Prediction done by the suggested approach utilizing channel attention mechanism

Table 4.3. Performance evaluation of the proposed model utilizing channel attention mechanism

Dataset	Algorithm	Ac (%)	Sp (%)	Se (%)
DIARETDB1	Liu et al. [125]	79	75	83
	Yazid et al. [123]	-	97.4	98.2
	Akram et al. [121]	98.92	98.35	99.65
	Welfer et al. [124]	-	98.84	70.48
	Proposed	98.73	99.38	98.87
IDRiD	Proposed	93.25	92.13	92.64
MESSIDOR	Kaur et al. [128]	86	90.47	81.32
	Zhang et al. [127]	-	-	62.3
	Proposed	90.91	91.15	91.26
DIARETDB0	Akram et al. [121]	96.48	98.38	93.7
	Proposed	97.64	98.69	94.53

Table 4.4. Comparison of performances considering the dice scores

Algorithm	DIARETDB1	IDRiD	MESSIDOR	DIARETDB0
U-Net [21]	82.06 %	84.93 %	80.27 %	83.51%
Proposed	90.34 %	91.18 %	91.65 %	90.87 %

With 95.13 % overall accuracy, the experimental outcome depicted in table 4.3 demonstrates that this strategy utilizing channel attention mechanism, surpasses the prevailing techniques considering metrics like accuracy, sensitivity and specificity. Table 4.4 compares the dice score attained in this work which utilizes channel attention mechanism for segmenting exudates, with that of the U-Net architecture.

4.3 Methodology Adopted for the Segmentation of Exudates Utilizing Combined Channel and Spatial Attention Mechanism

An attention method works in a manner identical to that of human attention mechanism for emphasizing on the high-value details from the overall information received in an ongoing work. The ability of attention method to emphasize significant and key input components through learning method has made them a popular element in deep neural networks. Although the attention mechanism has gained much importance in segmenting natural scenes, but there has been less progress in the application these techniques to segment the biomedical pathologies.

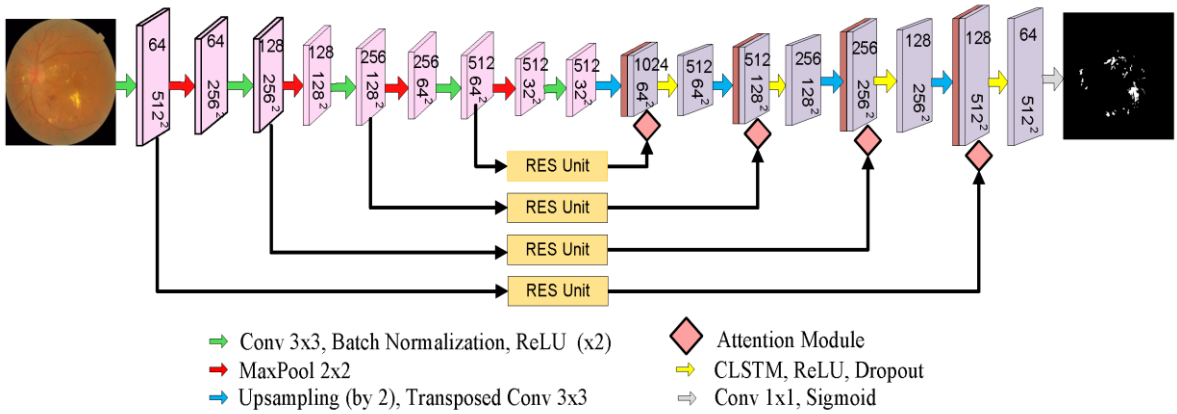


Fig. 4.9. The layout of the proposed framework utilizing combined channel and spatial attention mechanism

In an effort to retrieve the peculiarities of spatial information which are dropped as a result of the max pooling procedure, the decoder-encoder framework conveys the locational details from encoder side into the decoder module. Superior segmentation outcomes can be achieved with the utilization of combined Channel and Spatial Attention Mechanism (CSAM) which provides the location and semantic details in the skip connection. This study fuses the spatial and channel attention features, as in [129], to segment the exudate efficiently in fundus image. The channel attention mechanism incorporates more contextual details in the low-level encoder features and thus reducing the semantic discrepancy among the decoder and the encoder attributes, while the spatial attention process directs the framework to concentrate on spatial information of important region. The suggested framework involving Channel and Spatial attention mechanism is depicted in fig. 4.9.

4.3.1 Channel Attention Mechanism

The channel attention mechanism (CAM) has gained significant interest and exhibited remarkable prospects in enhancing the neural network performance. The main aspect is the automatic identification of the weights for every channel, such that the feature maps having inaccurate or invalid information are allocated lesser weights, whereas feature maps having information that is more crucial for predicting the eventual outcome, are allocated higher weights. The channel attention module has been implemented by analyzing the relationships among the various feature channels. The fig. 4.10 (a) displays the proposed CAM framework. The feature map (M) learnt in the decoder side is combined for yielding a channel index by utilizing average pooling operation, which is expressed as:

$$n_q = \frac{1}{X \times Y} \sum_{x=1}^{X} \sum_{y=1}^{Y} m_q(x,y) \tag{4.14}$$

where, n_q signifies the feature map's q^{th} channel information and $m_q(x,y)$ indicates the feature score at location (x,y) for channel q of the feature map in M. Subsequently, fully connected (FC) blocks have been utilized to analyze the channel corelationships. The channel information vector $n = [n_1, n_2, \ldots, n_k]^T$ is transformed by the initial FC block into a vector with diminished dimension. The second FC block helps in restoring the original channel dimension, k, as a channel attention vector P_{CAM} given as

$$P_{CAM} = X_2(X_1 n) \tag{4.15}$$

where, $X_1 \in \Re^{\frac{k}{r} \times k}$ and $X_2 \in \Re^{k \times \frac{k}{r}}$ signifies the FC layers parameter respectively and r denotes for the scaling factor. In this work r=8 has been considered. The sigmoid function (σ) is then employed to produce attention maps in [0, 1] range. A value closer to 1 denotes more significant features.

$$Q_{CAM} = \sigma(P_{CAM}) \tag{4.16}$$

4.3.2 Spatial Attention Mechanism

The upsampling procedure utilized in the decoding side of the conventional neural network framework brings about losses in valuable spatial details. To resolve this challenge,

skip connections are employed to integrate the encoder feature map containing important spatial information with that of the decoder path. However, the straightforward integration of the encoder decoder feature map adds a lot of redundant unnecessary low-level features. As a result spatial attention mechanism (SAM) has been incorporated in the decoder to efficiently inhibit the activation zones with insufficient discriminant information so that the redundant features get diminished. The fig. 4.10 (b) depicts the suggested SAM architecture. A feature map (M) retrieved from the decoder section is given as $M \in \Re^{X \times Y \times C}$. The spatial attention method is applied initially using a convolutional unit with kernel of dimension 1×1 and an output channel having size of 1. The mathematical expression is given as:

$$P_{SAM} = f_{conv1\times1}(M) \tag{4.17}$$

where $P_{SAM} \in \Re^{X \times Y}$ share identical spatial dimension as that of M. The sigmoid activation (σ) is then performed to develop attention maps in [0, 1] range. The mathematical expression is given as

$$Q_{SAM} = \sigma\,(P_{SAM}) \tag{4.18}$$

4.3.3 Combined Channel and Spatial Attention Mechanism

To avail the merits the aforementioned two attention models, the characteristics of the channel and spatial attention map are combined, as expressed in equation 4.19 to generate the ultimate refined features. These obtained features are further integrated with decoder features to perform the subsequent computations. The integration of the two attention models is shown in fig. 4.10 (c).

$$Q_{CSAM} = f_{Ext}^{Spa}(Q_{CAM}) + f_{Ext}^{Ch}(Q_{SAM}) \tag{4.19}$$

where, $f_{Ext}^{Spa}()$ and $f_{Ext}^{Ch}()$ expands the channel and the spatial maps in spatial and channel direction to the identical size as that of M respectively. Thereafter, it feeds to the mainstream. The obtained attention maps are gradually added with the original feature maps for developing the final attention modulated maps as given in equation 4.20.

$$M_{CSAM} = M \otimes Q_{CSAM} \tag{4.20}$$

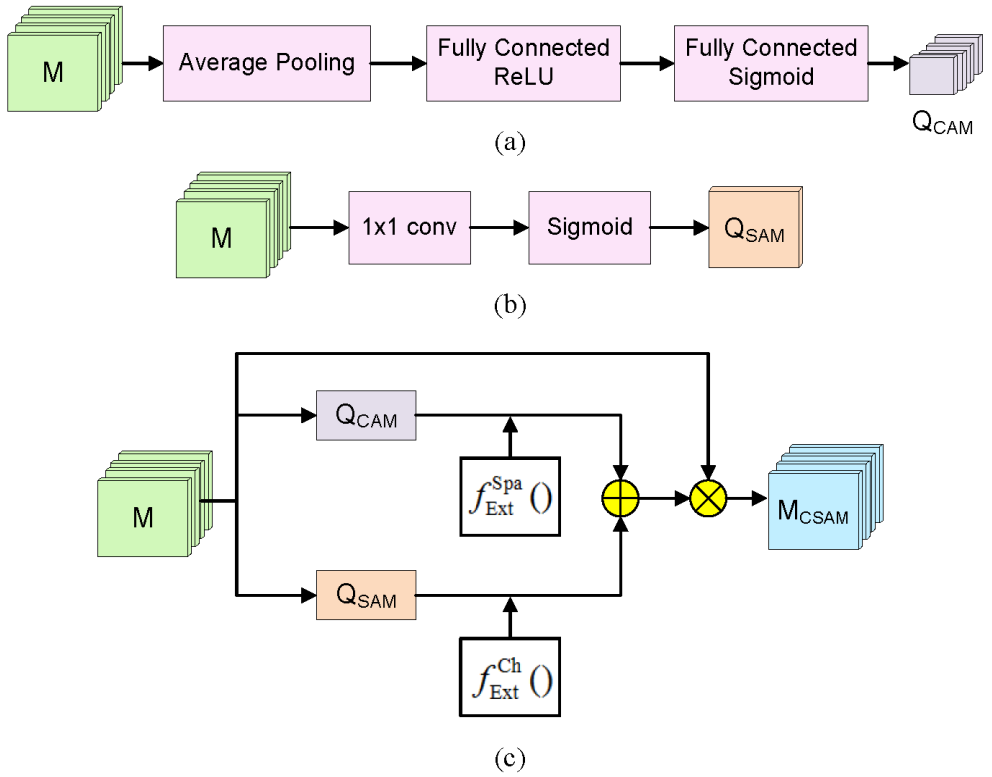


Fig. 4.10. (a) The module for channel attention (b) the module for spatial attention (c) the module for attention feature fusion

4.3.4 Convolutional Long Short Term Memory

Although the Long Short Term Memory (LSTM) [28] framework is well-known for executing sequential activities, but the conventional LSTM misses the spatial information as it requires a one-dimensional vectorized input while processing. The spatial information contributes a vital part in boosting the network's ability. CNN algorithm helps to identify and preserve only the spatial information. In order to preserve both the temporal and the spatial information, the Convolutional Long Short Term Memory (CLSTM) [39] is introduced into the architecture. In the CLSTM, the progression from state-to-state and from input-to-state is carried out using the convolution operation. The CLSTM records the spatiotemporal characteristics during the convolution operation. The working of CLSTM architecture is governed by the equations 4.21–4.25.

$$i_t = \sigma\left(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \times C_{t-1} + b_i\right) \quad (4.21)$$

$$f_t = \sigma\left(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \times C_{t-1} + b_f\right) \quad (4.22)$$

$$C_t = f_t \times C_{t-1} + i_t \times \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \quad (4.23)$$

$$o_t = \sigma\left(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \times C_t + b_o\right) \quad (4.24)$$

$$H_t = o_t \times \tanh(C_t) \quad (4.25)$$

where, C_t, X_t, H_t, f_t, i_t, o_t, $*$, σ, $\times$, W and tanh represents the cell state, the input data, the hidden state, the forget gate, the input gate, the output gate, convolution operation, the sigmoid function, Hadamard product, learnable weight and the hyperbolic tangent function respectively. To identify the spatiotemporal relationship, the CLSTM layer is used so as to retrieve the hidden state details stepwise for optimizing the sequence and to preserve the internal arrangement of the sequential details. The fig. 4.11 displays the layout of the CLSTM unit.

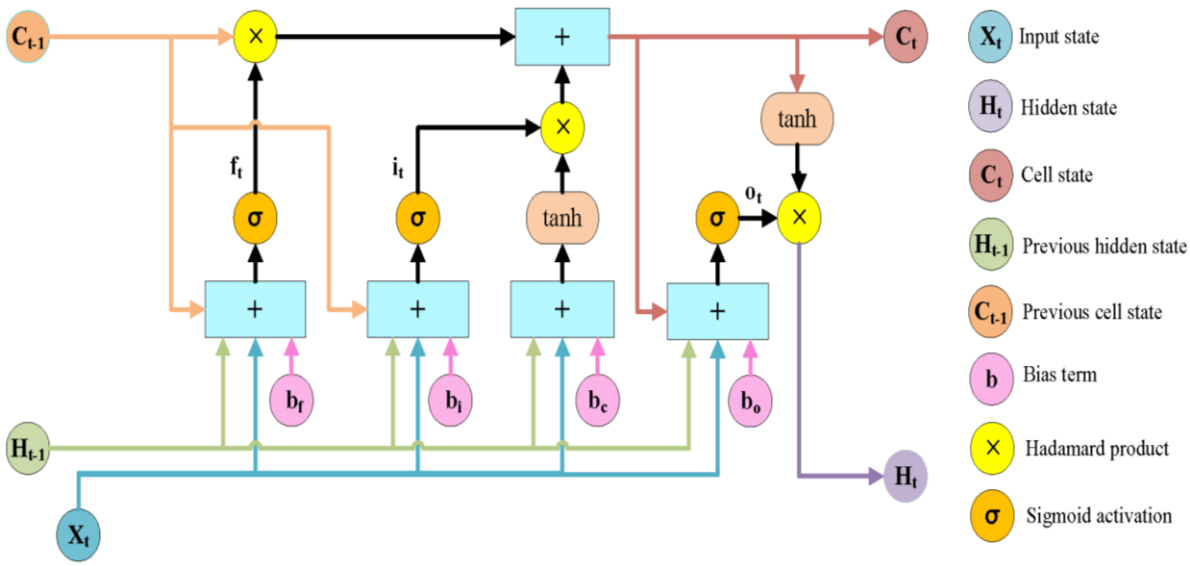


Fig. 4.11. Layout of the CLSTM unit

4.3.5 Residual Extended Skip

The Residual Extended Skip (RES) block, inspired by [130] is introduced into the architecture for transforming the lower grade information to mid-level information and thus helps in regulating the loss of information. The RES widens the network's receptive zone, thus strengthening the model's segmentation efficacy. The capability of RES block in

conducting contextual aggregation at various scales also makes it scale-invariant. The fig. 4.12 displays the structure of the RES unit. The framework's input is transmitted over five alternative paths. Each of the initial four paths employs two convolutional blocks to reduce the dimensions of the framework's parameters. The last path forwards the input received as it is for further processing. The outcomes from all five paths are combined to generate a single output. The combined output is thereafter fed through three convolutional layers to generate the final outcome.

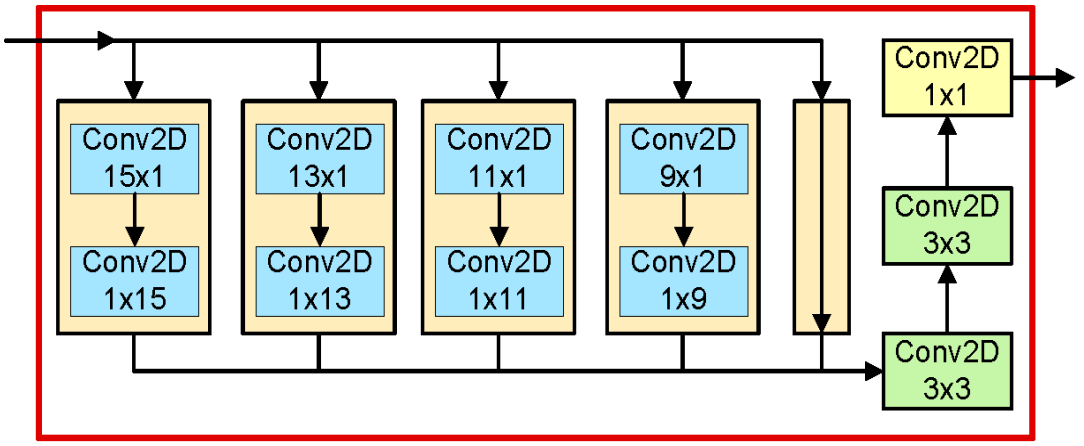


Fig. 4.12. The RES block

The proposed model incorporates the attention mechanism, the CLSTM technique and the RES unit. The attention mechanism focuses on the significant areas by assessing the relationship among the most appropriate features. The CLSTM technique used in the decoder helps to retain the spatiotemporal relationships. The RES unit enlarges the valid receptive field and thus enhances the model's overall effectiveness. At each and every stage of the decoder, the loss is computed and the overall loss is determined by adding all of the losses. The images from all the datasets are resized to 512×512 so as to accomplish the processing of the available data without any intervention. The techniques for data augmentation, namely rotating and flipping**,** are adapted to create training data. A workstation equipped with an i5 processing unit, 8 GB of RAM and a NVIDIA GeForce GTX 1060 graphics card is utilized to carry out the work.

4.3.6 Results

4.3.6.1 Dataset Used

The suggested method is assessed on DIARETDB0 [6], IDRiD [9], DIARETDB1 [7], and MESSIDOR [4]. All the images are scaled to 512×512 pixels.

4.3.6.2 Performance Evaluation Metric

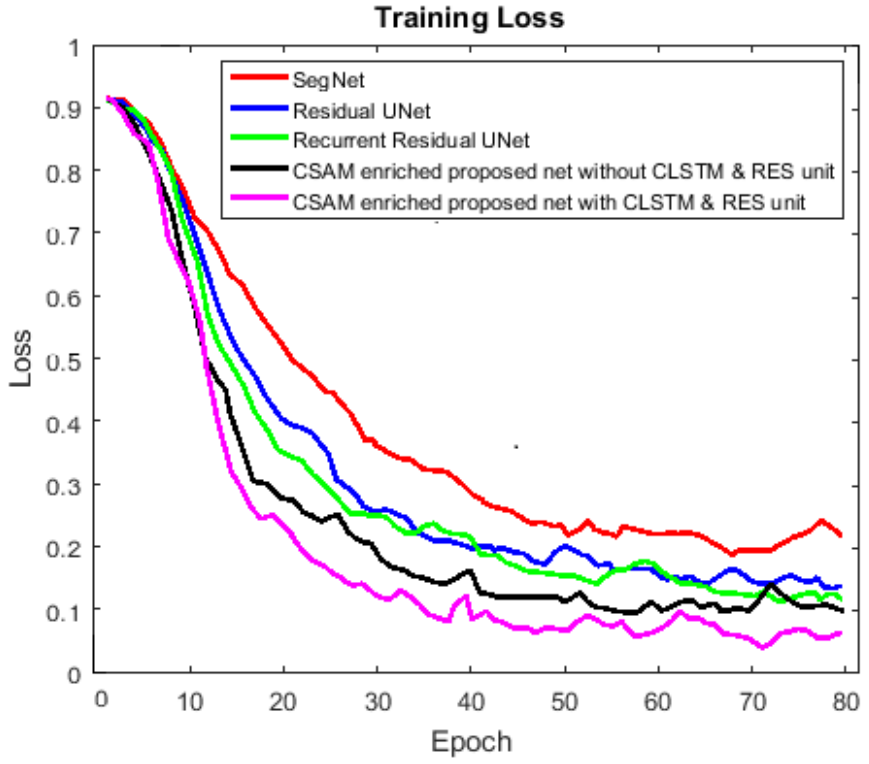
The performance evaluation metrics particularly accuracy (Ac), specificity (Sp), and sensitivity (Se), given by equations 4.5-4.7, and F1-score (F1), given by equation 4.26, are used to assess the effectiveness of the study. The F1 value monitors the ability of a model by computing the mean score of sensitivity and precision and mathematically represented as

$$\mathrm{F1} = \frac{2\times(\mathrm{Sensitivity}\times\mathrm{Precision})}{(\mathrm{Sensitivity}+\mathrm{Precision})} \tag{4.26}$$

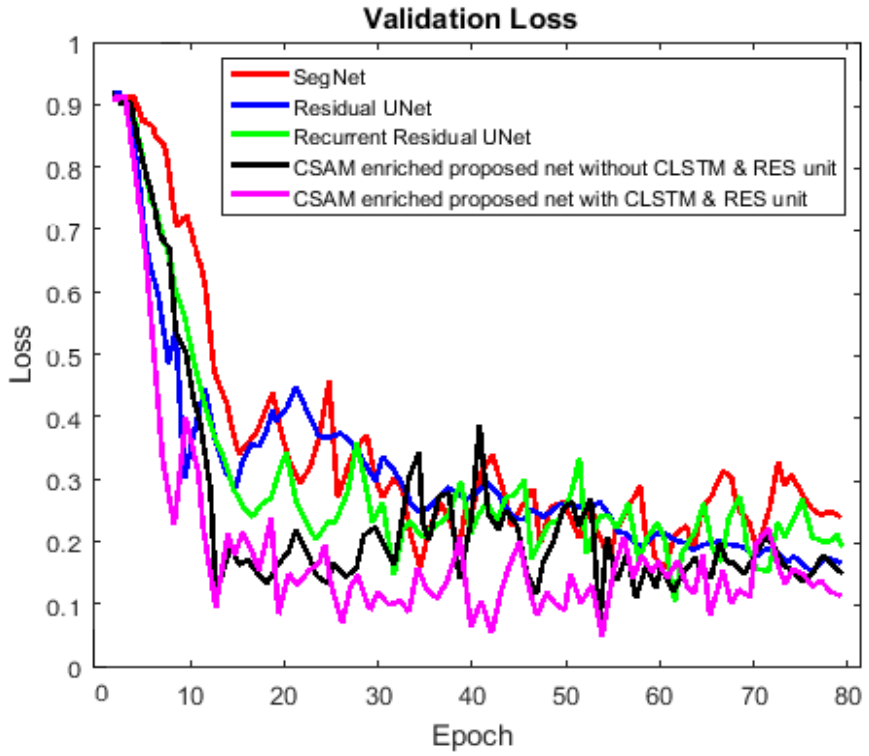
where, $\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$, and TN and TP signify the successfully identified background pixels and exudate samples respectively, and FN and FP signify the mistakenly identified background pixels and exudate samples respectively.

4.3.6.3 Network Implementation

Python with OpenCV 3.6 has been used for implementing of the proposed model. The batch size is fixed at 8 and the learning rate is preset to the value 0.0001. During the network training, the Adam has been chosen to serve an optimizer. The training loss curve and the validation loss curve are depicted in fig. 4.13 (a) and (b) respectively.



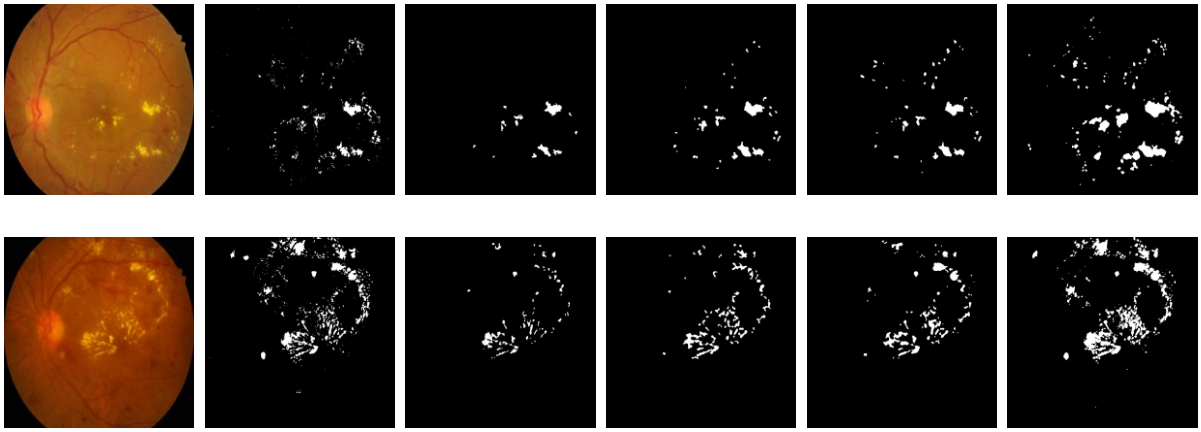
(a)



(b)

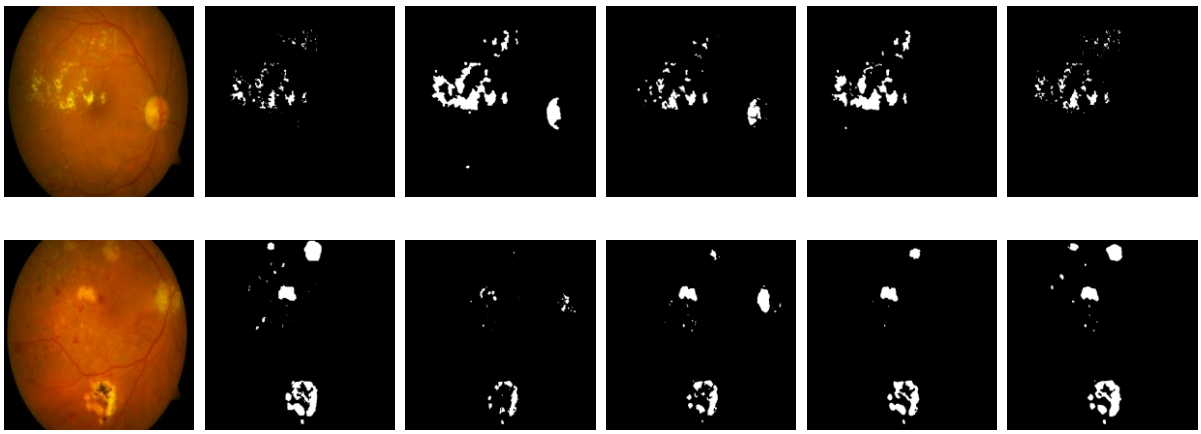
Fig. 4.13. (a) The training loss curve (b) the validation loss curve of the combined channel and spatial attention enhanced model

4.3.6.4 Experimental Outcomes



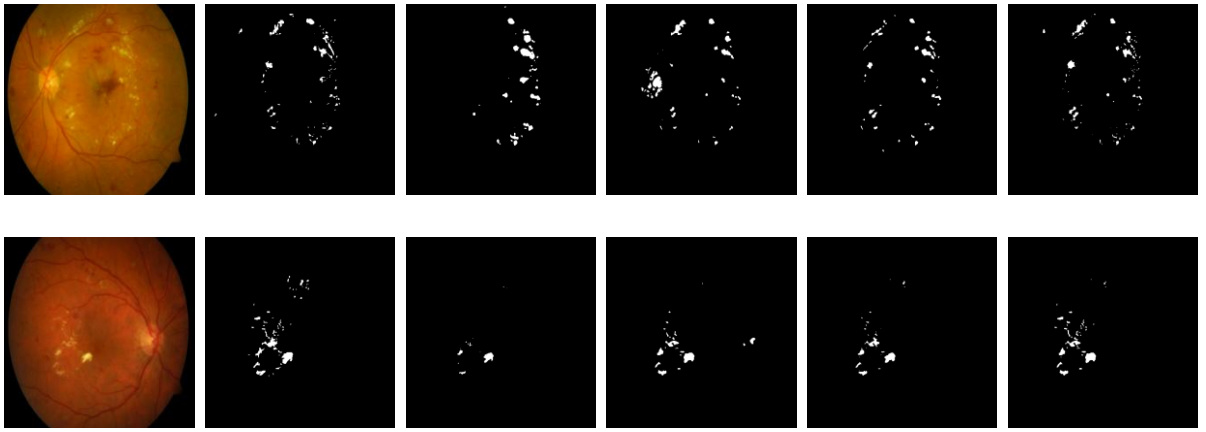
(a) (b) (c) (d) (e) (f)

Fig. 4.14. (a) IDRiD dataset retinal photograph, (b) labeled image, (c)-(e) segmented image by SegNet, Residual UNet, and Recurrent Residual UNet respectively (f) prediction done by the suggested approach utilizing combined channel and spatial attention mechanism



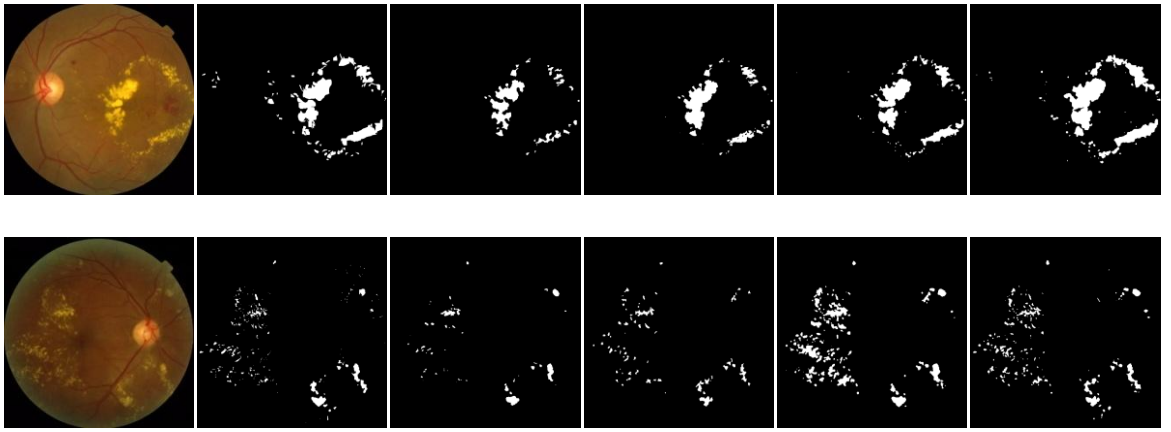
(a) (b) (c) (d) (e) (f)

Fig. 4.15. (a) DIARETDB0 dataset retinal photograph, (b) labeled image, (c)-(e) segmented image by SegNet, Residual UNet, and Recurrent Residual UNet respectively (f) prediction done by the suggested approach utilizing combined channel and spatial attention mechanism



(a) (b) (c) (d) (e) (f)

Fig. 4.16. (a) DIARETDB1 dataset retinal photograph, (b) labeled image, (c)-(e) segmented image by SegNet, Residual UNet, and Recurrent Residual UNet respectively (f) prediction done by the suggested approach utilizing combined channel and spatial attention mechanism



(a) (b) (c) (d) (e) (f)

Fig. 4.17. (a) MESSIDOR dataset retinal photograph, (b) labeled image, (c)-(e) segmented image by SegNet, Residual UNet, and Recurrent Residual UNet respectively (f) prediction done by the suggested approach utilizing combined channel and spatial attention mechanism

The experimental outcomes of the suggested exudate segmenting method are depicted in fig. 4.14-4.17. The outcome shows that the segmented object and the labeled data are very similar. The table 4.5 illustrates that, the combined channel and spatial attention mechanism enhanced proposed architecture with CLSTM and RES unit provides significantly better segmentation performance in comparison to other strategies.

Table 4.5. A comparative analysis considering the different models

Algorithm	**Dataset**	**Ac (%)**	**Se (%)**	**Sp (%)**	**F1 (%)**
SegNet [131]	IDRiD	92.57	91.12	92.28	90.87
	DIARETDB0	93.65	91.74	92.51	90.44
	DIARETDB1	94.42	92.97	94.35	92.01
	MESSIDOR	93.26	90.83	91.93	89.97
Residual UNet [132]	IDRiD	93.69	92.47	92.76	91.94
	DIARETDB0	94.81	93.58	94.17	92.9
	DIARETDB1	95.23	94.76	95.82	93.55
	MESSIDOR	94.38	91.64	93.41	91.21
Recurrent Residual UNet [68]	IDRiD	95.15	94.36	93.83	93.71
	DIARETDB0	95.73	94.91	96.64	94.52
	DIARETDB1	95.87	95.92	96.06	94.81
	MESSIDOR	94.92	92.15	95.24	92.79
CSAM enriched proposed network without CLSTM and RES unit	IDRiD	95.93	94.88	95.12	94.62
	DIARETDB0	97.06	95.28	98.21	95.54
	DIARETDB1	97.54	98.29	97.14	96.42
	MESSIDOR	96.49	93.61	96.57	94.63
CSAM enriched proposed network with CLSTM and RES unit	IDRiD	96.74	96.52	96.88	95.83
	DIARETDB0	97.89	96.34	98.73	96.98
	DIARETDB1	98.27	99.13	97.79	97.61
	MESSIDOR	98.03	95.87	99.11	97.02

Table 4.6. Performance evaluation of the proposed approach utilizing combined channel and spatial attention mechanism with the existing methodologies

Dataset	Algorithm	Ac (%)	Se (%)	Sp (%)	F1 (%)
IDRiD	Zabihollahy et al. [133]	88.46	96.15	80.77	67.23
	Guo et al. [134]	-	95.74	-	95.57
	Proposed	96.74	96.52	96.88	95.83
DIARETDB0	Lokuarachchi et al. [122]	-	93.64	90	-
	Akram et al. [121]	96.48	93.7	98.38	-
	Proposed	97.89	96.34	98.73	96.98
DIARETDB1	Lokuarachchi et al. [122]	-	94.59	88.46	-
	Kaur et al. [128]	87	91	94	-
	Khojasteh et al. [135]	98.2	99	96	-
	Liu et al. [125]	79	83	75	-
	Yazid et al. [123]	-	98.2	97.4	-
	Fraz et al. [49]	87.72	92.42	81.25	-
	Proposed	98.27	99.13	97.79	97.61
MESSIDOR	Fraz et al. [49]	98.36	92.31	99.03	-
	Kaur et al. [128]	93	88	98	-
	Agurto et al. [136]	93	79	92	-
	Zhang et al. [127]	-	62.3	-	-
	Proposed	98.03	95.87	99.11	97.02

The results presented in table 4.6 demonstrate that the designed strategy outpaces the existing techniques in context to accuracy, specificity, sensitivity and F1 score. The method's ability to take care of numerous kinds of retinal pictures from several public datasets indicate the robustness of the algorithm and viability as an automated computerized application for diagnosing the DR.

4.4 Summary

In this work, CNN based algorithms have been presented to segment exudates automatically and thus facilitate in early diagnosis of DR. The illumination irregularity in fundus images, the imbalances in data distribution and the erratic shape and size of the pathologies of DR cause a serious hindrance in the analysis process. Over-fitting is another vital concern in deep learning models particularly in cases where the availability of labeled dataset is limited. Data augmentation technique provides a valuable approach to lessen the severity of the over-fitting by expanding the training dataset image count. The exudate segmentation outcome illustrates that the network provides superior output in several datasets than the earlier approaches. This indicates the reliability of the study and its eligibility as a computer-aided diagnostic system.

Chapter 5

Segmentation of the Macula Region

5.1 Methodology Adopted for the Segmentation of the Macula

In this work an automated macula segmentation algorithm that relies on a deep convolutional network has been developed. It comprises of an expanding and a contracting section. Skip connections which are found to improve network performance [20], have been utilized to communicate between the levels. The proposed method involves stages to maintain illumination equilibrium and to filter out the noises from the fundus images. For enhancing the global and the local network features, the RES unit, inspired by inception net [137], has been introduced in the framework, as shown in fig. 5.1, to transmit the contextual information efficiently between the deep and the shallow layers. This work also indulges an attention module for enabling the decoder to consider specific sections in the input sequence at various levels of decoding. Attention, which is a concept of memory, is achieved by attending several inputs over a certain period of time. The attention mechanism not only solves the vanishing gradient issue by establishing linkages between the decoder and the encoder blocks, but also helps in handling the bottleneck issue.

The model accepts image of dimension 512×512 and produces an image of identical shape as an output. Each block in the encoder comprises of two convolutional layers, and one max-pooling layer. In decoder, each block begins with a transposed-convolution operation. The outcome of the transposed-convolutional block is fused along with the corresponding RES block's output, followed by two convolutional and one dropout layer. Another convolutional layer having 1×1 filter size is incorporated in the decoder's final block. Every convolution

operation is succeeded by batch normalization and ReLU activation with an exception in the last convolutional layer, where sigmoid activation is being used. The convolution operation produces feature maps by extracting the different features from the fundus photograph. The dimension of the feature map is decreased by the pooling technique. The max-pooling technique used in this work selects the maximum value out of each kernel and thus aids in decreasing the feature map dimension. The dropout strategy randomly discards units as well as their connections from the network while training and thereby helps to prevent the over-fitting issue.

5.1.1 Network Implementation

With a batch size of 8, learning rate of 0.001 and early stopping with patience level set to 10, the framework utilized the dice coefficient loss function and Adam optimizer. The experiments are executed in a system equipped with 8 GB RAM, Intel i5 processor, and 6 GB NVIDIA GeForce GTX 1060 GPU. The proposed strategy has been enacted employing Python and OpenCV 3.6. The mathematical expression for the sigmoid, σ(x) and ReLU, R(x) activation are given as follows.

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{5.1}$$

$$R(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{5.2}$$

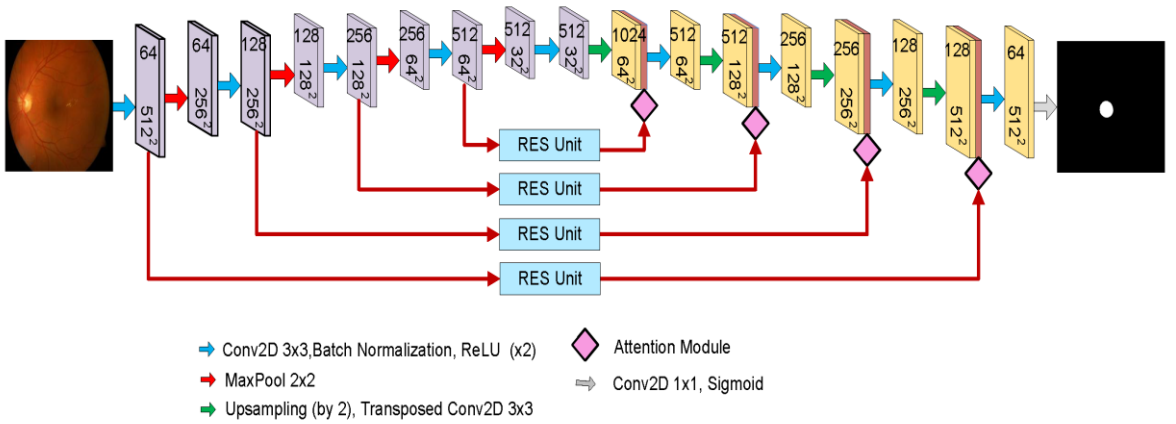


Fig. 5.1. The proposed Residual Extended Skip (RES) unit integrated CNN architecture

The convolutional networks use loss function for estimating error values from every training iteration. The selection of a loss function, which serves a crucial part by adjusting the internal weights at the time of back-propagation, substantially impacts the performance of a model. Dice Coefficient is a popularly known metric that determines how similar two images are. In 2016, Milletari et al. [138] introduced a cost function named the Dice Loss in computer vision by deriving it from the Dice Coefficient for segmenting medical images. In this study, the dice loss has been accepted as a loss function. The computational equation of the dice loss is expressed as:

$$\text{Dice Coefficient (DC)} = \frac{2\sum_j^n p_j g_j + 1}{\sum_j^n p_j^2 + \sum_j^n g_j^2 + 1} \tag{5.3}$$

$$\text{Dice Loss} = 1 - \text{Dice Coefficient} \tag{5.4}$$

where, g and p signifies the expert annotated and prediction pixels respectively. The denominator and the numerator are both increased by a numerical value of 1 in an effort to avoid the loss function from becoming undefined in extreme cases when g = p = 0.

5.1.2 Residual Extended Skip

The Residual Extended Skip (RES) unit prevents the degradation of information by producing a middle-level feature. In RES unit, the input is fed to five blocks connected in parallel. Each of the initial four blocks employs two convolutional layers while the last block is skip connected. Instead of employing one convolutional layer of N×N filter size, two cascaded convolutional layers having filter size of N×1 and 1×N respectively are introduced to reduce the size of network parameters. Also, it has been found through experiments that the cascaded convolutional layers having fewer parameters perform better than a single convolutional layer with larger parameter size. Finally the output is generated by adding the individual outcome obtained from each of the five blocks, followed by three convolutional layers with 3×3, 3×3, and 1×1 filter size respectively. The framework of RES unit is depicted in fig. 5.2. The RES unit retrieves contextual information, which enables the network to segment object more effectively. The unit also carries out information integration at the transitional phases, resulting in a more accurate reformation of images.

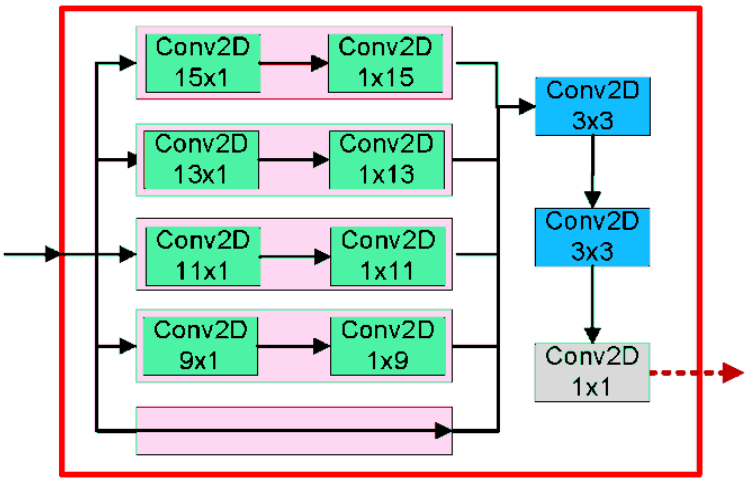
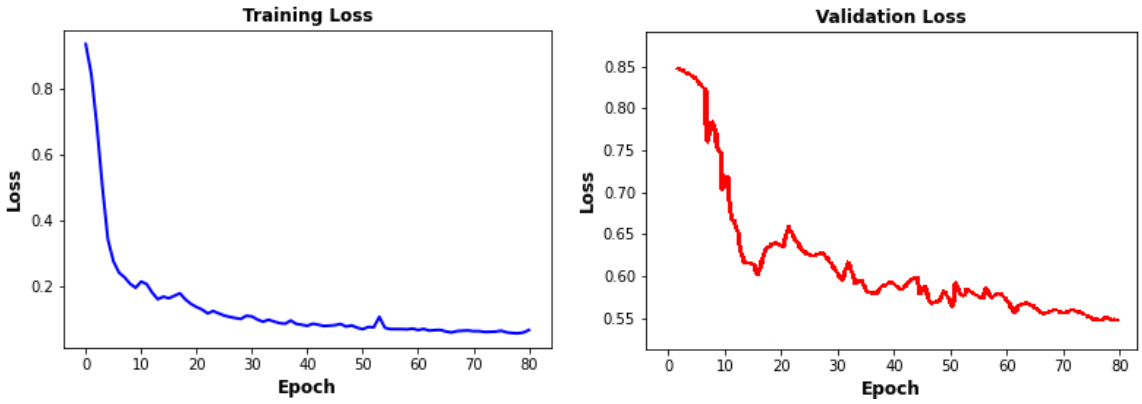


Fig. 5.2. The layout of Residual Extended Skip (RES) unit

5.2 Result

5.2.1 Database Used

The proposed technique is assessed on three retinal databases namely MESSIDOR [4], DIARETDB1 [7] and DIARETDB0 [6]. The MESSIDOR, DIARETDB1 and DIARETDB0 databases include 1200 images (400 without pupil dilation and 800 with dilation), 89 images (5 normal and 84 symptomatic) and 130 images (20 normal and 110 symptomatic) respectively, hence, providing 1419 images in total for analysis. The framework is trained with images of MESSIDOR 2 [139] database which contains 1748 images. The images from all the databases are reduced to 512×512 pixels. Medical professionals have offered their assistance in labeling the macula region on these pictures. The training and the validation loss curves are depicted in fig. 5.3.



(a) (b)

Fig. 5.3. (a) The training loss curve and (b) the validation loss curve

5.2.2 Performance Metric

The accuracy (Ac), dice coefficient (DC), specificity (Sp) and sensitivity (Se) are considered as the metrics for assessing the potency of the suggested network. Accuracy signifies the percentage of correctly predicted occurrences. The dice score records the alikeness between the expert annotated and the prediction image. Specificity ascertains the probability to accurately identify the healthy individuals whereas the sensitivity reveals the probability to accurately identify the sick patients. The mathematical expressions for Ac, Sp and Se are given as:

$$Ac = \frac{TP+TN}{TP+FP+TN+FN} \tag{5.5}$$

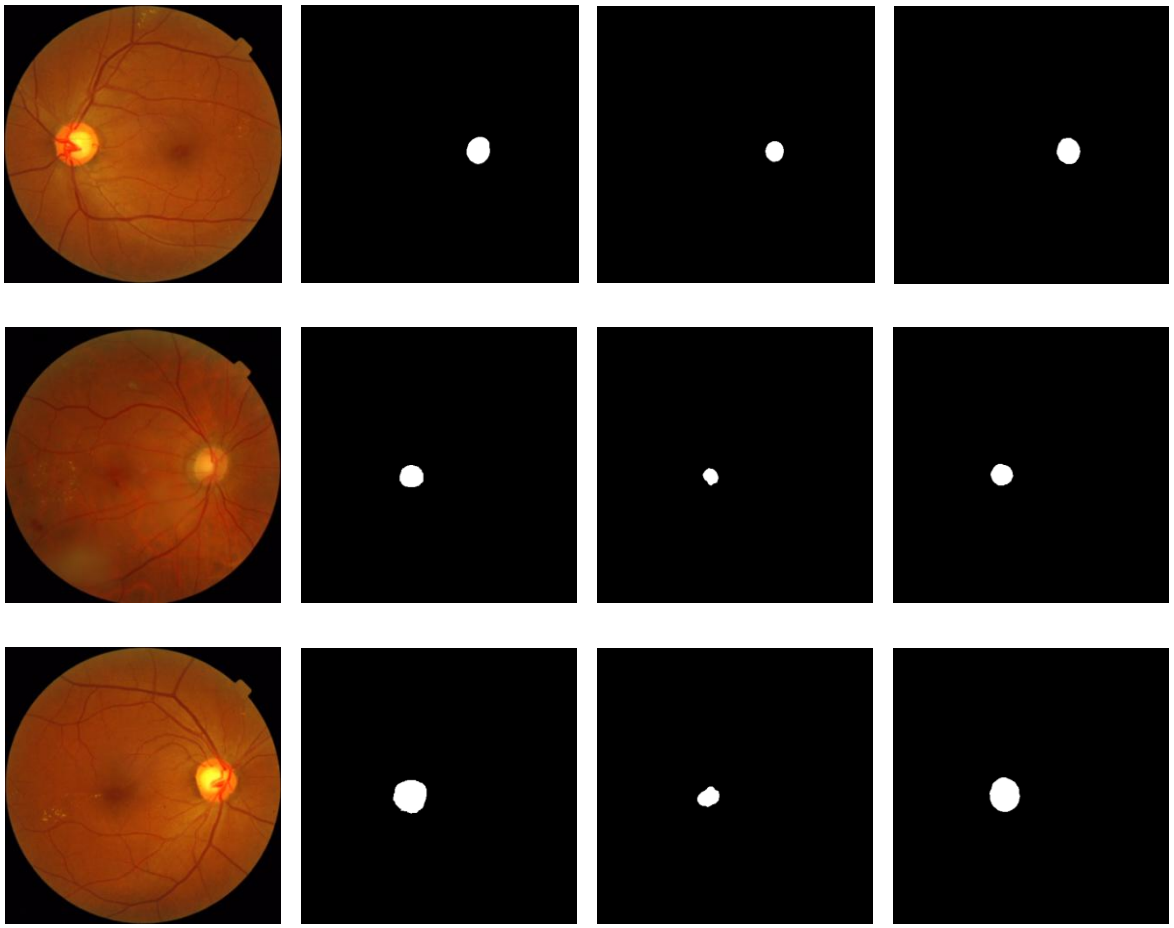
$$Sp = \frac{TN}{TN+FP} \tag{5.6}$$

$$Se = \frac{TP}{TP+FN} \tag{5.7}$$

where, TN and TP stand for the background and macula pixels respectively that were correctly recognized, whereas FN and FP stand for the incorrectly recognized background pixels and macula pixels.

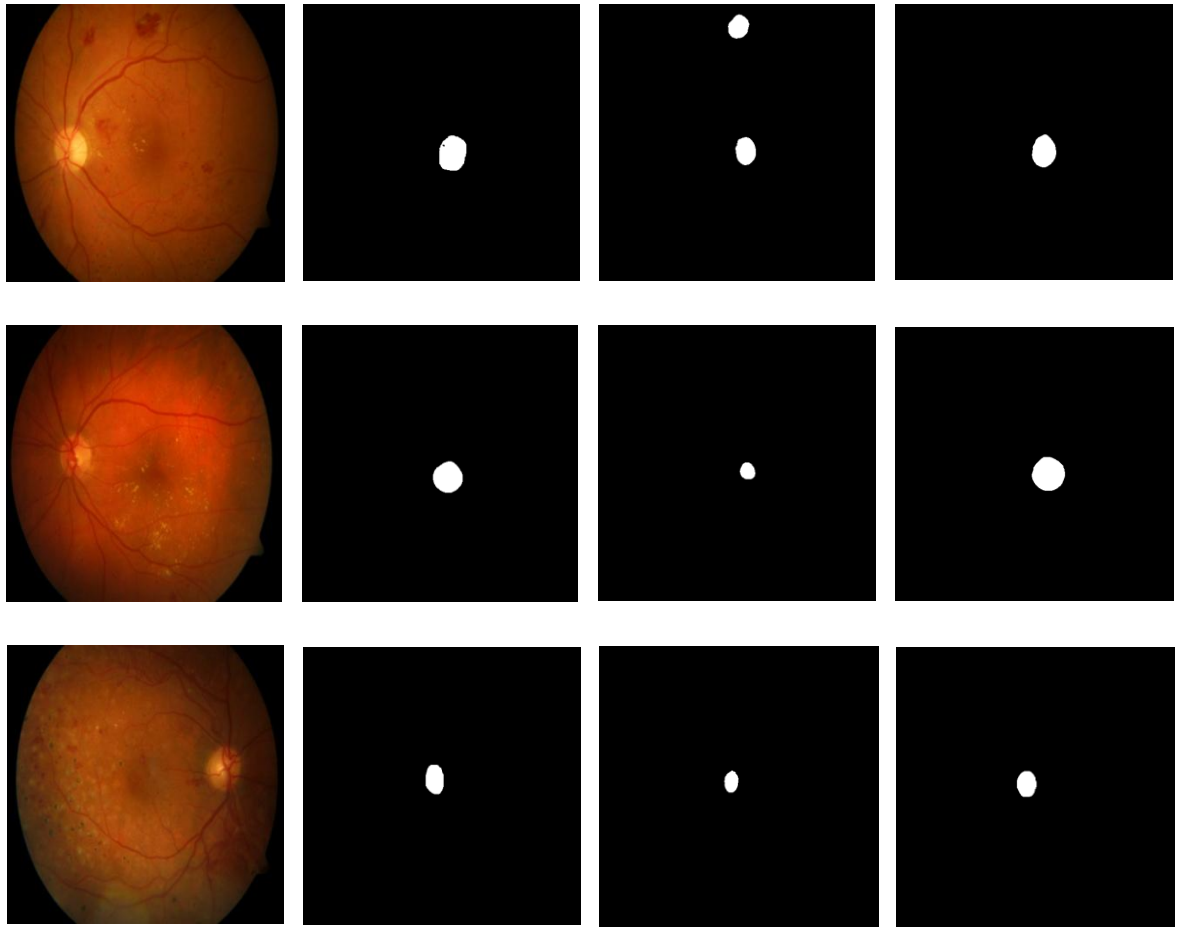
5.2.3 Experimental Outcomes

According to the majority of publications, an acquired fovea center position is considered accurate if the separation between it and the actual site of fovea centre is lesser than the length of the radius of OD. This postulate is referred as the 1R criterion. Two more criterions, quarter and half of the OD radius are also taken into account, referred as 0.25R and 0.5R criterion respectively, in several works [87], [140] to provide a more accurate evaluation. The acquired fovea centre position is regarded to be 'excellent', if it meets the 0.25R requirement. The results presented in table 5.1, 5.2, and 5.3 demonstrate the improved efficacy of the suggested strategy for performing macula segmentation in contrast to the previous segmentation approaches. The outcomes illustrated in table 5.4 also ascertain the reliability of the suggested methodology. The experimental outcomes shown in fig. 5.4, 5.5, and 5.6 demonstrate the improvements of the suggested technique against the widely utilized U-Net.



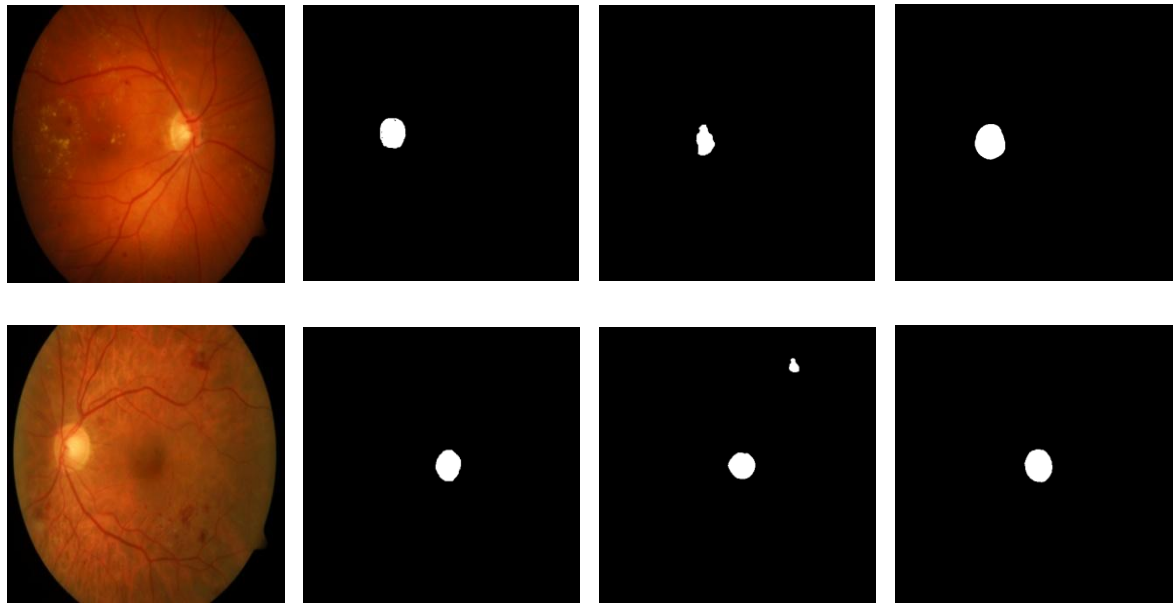
(a) (b) (c) (d)

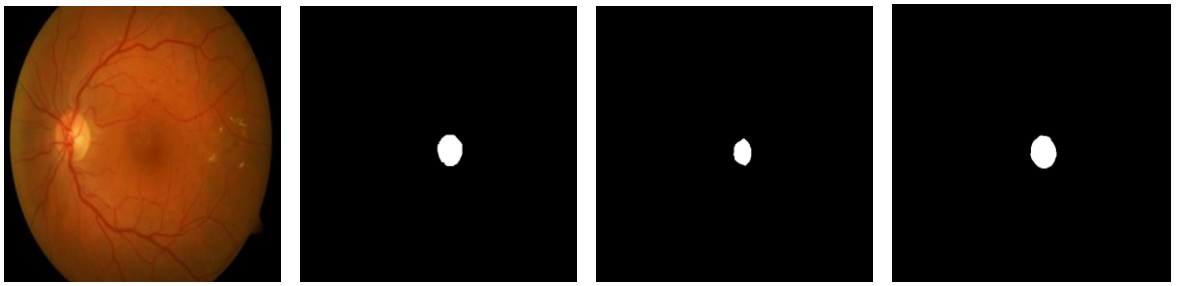
Fig. 5.4. (a) Picture of retina, (b) Labeled macula region, (c) U-Net segmented macula region and (d) Proposed technique predicted macula region considering MESSIDOR database



(a) (b) (c) (d)

Fig. 5.5. (a) Picture of retina, (b) Labeled macula region, (c) U-Net segmented macula region and (d) Proposed technique predicted macula region considering DIARETDB1 database





(a) (b) (c) (d)

Fig. 5.6. (a) Picture of retina, (b) Labeled macula region, (c) U-Net segmented macula region and (d) Proposed technique predicted macula region considering DIARETDB0 database

Table 5.1. Comparison of performances of the suggested technique with other prevailing approaches on MESSIDOR database

Algorithm	**Count of pictures**	**0.25R criterion (%)**	**0.5R criterion (%)**	**1R criterion (%)**
Gegundez-Arias et al. [84]	1136	76.32	93.84	98.24
Tewari et al. [86]	1200	-	-	97
Aquino [87]	1136	83.01	91.28	98.24
GeethaRamani et al. [140]	1200	85	94.08	99.33
Proposed	1200	87.83	95.5	99.58

Table 5.2. Comparison of performances of the suggested technique with other prevailing approaches on DIARETDB1 database

Algorithm	**Count of pictures**	**0.25R criterion (%)**	**0.5R criterion (%)**	**1R criterion (%)**
Antal et al. [141]	89	-	-	92
Tewari et al. [86]	89	-	-	96.9
Aquino [87]	89	-	-	94.38
GeethaRamani et al. [140]	89	-	-	97.75
Proposed	89	83.15	92.13	98.87

Table 5.3. Comparison of performances of the suggested technique with other prevailing approaches on DIARETDB0 database

Algorithm	Count of pictures	0.25R criterion (%)	0.5R criterion (%)	1R criterion (%)
Antal et al. [141]	130	-	-	86
Tewari et al. [86]	130	-	-	96.6
GeethaRamani et al. [140]	130	-	-	96.92
Proposed	130	82.31	93.08	98.46

Table 5.4. Segmentation results attained by proposed technique

Database	Ac (%)	DC (%)	Sp (%)	Se (%)
MESSIDOR	92.62	91.54	93.73	92.43
DIARETDB1	95.39	91.28	97.06	94.69
DIARETDB0	94.16	89.95	95.14	93.87

The algorithm has proved its robustness by effectively accepting the low quality fundus images obtained due to the illumination imbalance condition and heterogeneous composition. The experimental outcome endorses the adoption of this method to assist specialists by precisely identifying the macula in real time.

5.3 Summary

In this study, a fast, accurate and fully automated deep neural methodology to segment the macula region in fundus images has been presented. It is observed that even when dealing with a wide range of input images of various appearances, the proposed framework maintained its stable performance. The results obtained indicate the viability of the approach for using in retinal screening. This automated model which is capable of handling various qualities of retinal images, possesses the potential to segment the macula region from fundus photographs even in clinics with limited resources or employees.

Chapter 6

Conclusion and the Work's Future Aspect

6.1 Conclusion

DR is a complication which affects the retina and gradually impairs eyesight. Digital image processing aids in identifying the symptoms of DR at an earlier stage, preventing vision loss among patients with diabetes. The key purpose of this work is to introduce a computerized system for recognizing the OD, exudate, and macula region in fundus photos. Medical professionals are capable of identifying the pathologies manually, but the manual analysis of the symptoms is tiresome and not always accurate. Hence an automated screening technique must be designed to identify lesions in the captured retinal photographs. Numerous approaches have been established to uncover the anomalies in fundus images considering texture based information. Publicly available retinal datasets are employed for conducting the experiments in this study. The statistical indicators namely specificity, accuracy, F1-score and sensitivity are evaluated in the prescribed approaches. The suggested methodologies have been proved to be significantly effective in localizing and segmenting the lesions in retinal pictures.

Exudates are the most common indication of DR in the early phase, and their recognition is crucial for retinopathy diagnosis. The localizing of exudates close to the OD is difficult because of their similarity in colour. In this study, new techniques have been investigated and explored for reliably segmenting the exudate, OD and macula area in retinal pictures of DR patients. The accuracy of the segmentation methodologies improved as these techniques utilized colour information as a crucial feature to recognize various lesions.

In this study, various digital imaging approaches have been investigated to help medical experts identify the symptoms of DR at the earliest. The retinopathy has emerged as a primary threat to eye sight, because of the rise of patients with diabetes. Although diabetes cannot be completely healed, the associated medical issues particularly in regard to vision can be reduced through proper treatment. The diagnosis of DR is considered effectual only if the pathologies are identified at the initial phase. Contrarily, the symptoms of DR are not evident in the initial phase, which causes many individuals to go untreated unless severe visual impairment has occurred.

To assure that the patients receive timely treatment, routine screening programmes are organized to check the retina of patients with diabetes. The mass screening procedure used nowadays for DR recognition undergoes manual evaluation of coloured fundus pictures. This encourages the need for experienced retinal specialists to investigate the lesions in the retinal photographs. As the DR patients are substantially growing, the number of fundus photos of individuals suffering from visual issues is dramatically rising. This condition enhances the workload of ophthalmologists significantly, necessitating the employment of skilled professionals, and eventually increasing the expense of healthcare. In this instance, an efficient, accurate, quick, reliable and affordable strategy for the automated identification of pathologies in fundus pictures is essential. The suggested automated technology, designed to provide superior diagnosis of retinopathy, offers the potential for rapidly reviewing a significant number of fundus pictures, with high reliability. Moreover, the automated screening methodology provides a beneficial assistance in routine treatment by highlighting the lesions available in retinal pictures of individuals affected with DR, consequently lowering the workload of the specialists.

The proposed techniques are relatively less complex and thus can perform efficiently even in systems with a low configuration. The methodologies will be extremely beneficial in emerging nations where there is lesser number of medical professionals available to treat the significantly growing retinopathy patients.

Ophthalmologists can employ the exudate localization technology in the retinopathy screening process as a primary diagnostic tool to arrive at treatment-related decisions early. Moreover, the integration of the efficient OD and macula region recognition algorithms can increase the accuracy and the capability to determine the severity of DR. The outcome of

comparing the relevant approaches demonstrates that a computerized fundus picture assessment technique is the best strategy for rapid diagnosis of DR.

To assess the applicability of the proposed automated methodologies in clinics, the techniques developed in this study have been extensively evaluated on datasets of fundus pictures of varying contrast and quality. The experimental analysis has provided excellent outcomes, and this reveals that the methodologies can help ophthalmologists in the regular medical treatment by assisting them in identifying the OD, exudate, and macula region. The results indicate the reliability of the suggested neural network-based automated techniques for segmenting the OD, exudate, and macula region in fundus photographs. The designed frameworks represent a significant advancement in the direction of the mission of creating a technique for automatically screening the eye to identify the early symptoms of DR and thus aids in the prevention of sight impairment.

6.2 Scope for Future Work

The future scope includes the introduction of a software application that incorporates multiple automated retinal picture processing methodologies such as image registration, vessel segmentation, vessel tortuosity measurement, DR pathology identification, and crossover detection. The integration of the various pathology recognition algorithms in a single platform which inputs a fundus photograph and generates a range of mathematical indices to represent the severity of DR will be an area to explore. Such works will eventually help in better understanding of the retinal disorders and the eye fundus anatomy. Moreover, it is necessary to develop a resolution for retinopathy screening programmes where a collaboration between retinal professionals and healthcare facilities to evaluate large number of fundus photographs is crucial. In future, the OD, exudate, and macula region detection and segmentation algorithm's performance can be boosted by adopting improved approaches which will be capable of handling noisy fundus photographs obtained due to low illumination during picture acquisition. The segmentation accuracy also suffers from the limited availability of training images. An increase in the collection of annotated retinal images, in future, will help in improving the segmentation accuracy.

6.3 Summary

The concluding statement and the future scope for more research in the area of DR are discussed in this chapter. It is observed that the proposed segmentation networks have achieved superior segmentation performance. Even with challenging pictures having multiple complications, the models proved their superiority by detecting finer details. Nevertheless there are numerous improvements which can be implemented to increase the effectiveness of algorithms for detecting DR.

Bibliography

[1] H. E. Wiley and F. L. Ferris III, “Nonproliferative diabetic retinopathy and diabetic macular edema,” in *Retina*, Elsevier, 2013, pp. 940–968.

[2] A. R. Santiago, R. Boia, I. D. Aires, A. F. Ambrósio, and R. Fernandes, “Deep retinal image segmentation: a FCN-based architecture with short and long skip connections for retinal image segmentation,” *Front. Physiol.*, vol. 9, p. 820, 2018.

[3] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, “Ridge-based vessel segmentation in color images of the retina,” *IEEE Trans. Med. Imaging*, vol. 23, no. 4, pp. 501–509, 2004.

[4] E. Decencière *et al.*, “Feedback on a publicly distributed image database: the Messidor database,” *Image Anal. Stereol.*, vol. 33, no. 3, pp. 231–234, 2014.

[5] A. Hoover and M. Goldbaum, “Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels,” *IEEE Trans. Med. Imaging*, vol. 22, no. 8, pp. 951–958, 2003.

[6] T. Kauppi *et al.*, “DIARETDB0: Evaluation database and methodology for diabetic retinopathy algorithms,” *Mach. Vis. Pattern Recognit. Res. Group, Lappeenranta Univ. Technol. Finl.*, vol. 73, pp. 1–17, 2006.

[7] R. Kälviäinen and H. Uusitalo, “DIARETDB1 diabetic retinopathy database and evaluation protocol,” in *Medical Image Understanding and Analysis*, 2007, vol. 2007, p. 61.

[8] C. G. Owen *et al.*, “Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (CAIAR) program,” *Invest. Ophthalmol. Vis. Sci.*, vol. 50, no. 5, pp. 2004–2010, 2009.

[9] P. Porwal *et al.*, “Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research,” *Data*, vol. 3, no. 3, p. 25, 2018.

[10] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew, “A review of semantic segmentation using deep neural networks,” *Int. J. Multimed. Inf. Retr.*, vol. 7, no. 2, pp. 87–93, 2018.

[11] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. Garcia-Rodriguez, "A survey on deep learning techniques for image and video semantic segmentation," *Appl. Soft Comput.*, vol. 70, pp. 41–65, 2018.

[12] F. Garcia-Lamont, J. Cervantes, A. López, and L. Rodriguez, "Segmentation of images by color features: A survey," *Neurocomputing*, vol. 292, pp. 1–27, 2018.

[13] O. Erkaymaz, M. Ozer, and M. Perc, "Performance of small-world feedforward neural networks for the diagnosis of diabetes," *Appl. Math. Comput.*, vol. 311, pp. 22–28, 2017.

[14] M. Surucu, Y. Isler, M. Perc, and R. Kara, "Convolutional neural networks predict the onset of paroxysmal atrial fibrillation: Theory and applications," *Chaos An Interdiscip. J. Nonlinear Sci.*, vol. 31, no. 11, p. 113119, 2021.

[15] S.-H. Wang, Q. Zhou, M. Yang, and Y.-D. Zhang, "ADVIAN: Alzheimer's disease VGG-inspired attention network based on convolutional block attention module and multiple way data augmentation," *Front. Aging Neurosci.*, vol. 13, p. 313, 2021.

[16] S.-H. Wang, M. A. Khan, and Y.-D. Zhang, "VISPNN: VGG-Inspired Stochastic Pooling Neural Network," *C. Mater. Contin.*, vol. 70, no. 2, pp. 3081–3097, 2022.

[17] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[18] Y. Zhou, L. Xie, W. Shen, Y. Wang, E. K. Fishman, and A. L. Yuille, "A fixed-point model for pancreas segmentation in abdominal CT scans," in *International conference on medical image computing and computer-assisted intervention*, 2017, pp. 693–701.

[19] Y. Li, H. Qi, J. Dai, X. Ji, and Y. Wei, "Fully convolutional instance-aware semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2359–2367.

[20] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation," in *Deep learning and data labeling for medical applications*, Springer, 2016, pp. 179–187.

[21] 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*, 2015, pp. 234–241.

[22] V. I. Iglovikov, A. Rakhlin, A. A. Kalinin, and A. A. Shvets, "Paediatric bone age assessment using deep convolutional neural networks," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*, Springer, 2018, pp. 300–308.

[23] M. K. Hasan, M. A. Alam, M. T. E. Elahi, S. Roy, and R. Martí, "DRNet: Segmentation and localization of optic disc and Fovea from diabetic retinopathy image," *Artif. Intell. Med.*, vol. 111, p. 102001, 2021.

[24] K. A. Korznikov, D. E. Kislov, J. Altman, J. Doležal, A. S. Vozmishcheva, and P. V Krestov, "Using U-Net-Like Deep Convolutional Neural Networks for Precise Tree Recognition in Very High Resolution RGB (Red, Green, Blue) Satellite Images," *Forests*, vol. 12, no. 1, p. 66, 2021.

[25] W. Yao, Z. Zeng, C. Lian, and H. Tang, "Pixel-wise regression using U-Net and its application on pansharpening," *Neurocomputing*, vol. 312, pp. 364–371, 2018.

[26] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*, 2016, pp. 424–432.

[27] V. Iglovikov, S. Seferbekov, A. Buslaev, and A. Shvets, "Ternausnetv2: Fully convolutional network for instance segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 233–237.

[28] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Phys. D Nonlinear Phenom.*, vol. 404, p. 132306, 2020.

[29] J. Li, X. Lin, H. Che, H. Li, and X. Qian, "Pancreas segmentation with probabilistic map guided bi-directional recurrent UNet," *Phys. Med. Biol.*, vol. 66, no. 11, p. 115010, 2021.

[30] T. Zeng, B. Wu, J. Zhou, I. Davidson, and S. Ji, "Recurrent encoder-decoder networks for time-varying dense prediction," in *2017 IEEE International Conference on Data*

Mining (ICDM), 2017, pp. 1165–1170.

[31] W. Bai *et al.*, "Recurrent neural networks for aortic image sequence segmentation with sparse annotations," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018, pp. 586–594.

[32] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[33] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2016.

[34] M. Z. Alom, C. Yakopcic, T. M. Taha, and V. K. Asari, "Nuclei segmentation with recurrent residual convolutional neural networks based U-Net (R2U-Net)," in *NAECON 2018-IEEE National Aerospace and Electronics Conference*, 2018, pp. 228–233.

[35] K. Xu *et al.*, "Show, attend and tell: Neural image caption generation with visual attention," in *International conference on machine learning*, 2015, pp. 2048–2057.

[36] O. Vinyals, Ł. Kaiser, T. Koo, S. Petrov, I. Sutskever, and G. Hinton, "Grammar as a foreign language," *Adv. Neural Inf. Process. Syst.*, vol. 28, pp. 2773–2781, 2015.

[37] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition," *arXiv Prepr. arXiv1402.1128*, 2014.

[38] J. Zhao, F. Deng, Y. Cai, and J. Chen, "Long short-term memory-Fully connected (LSTM-FC) neural network for PM2. 5 concentration prediction," *Chemosphere*, vol. 220, pp. 486–492, 2019.

[39] Y. Liu, H. Zheng, X. Feng, and Z. Chen, "Short-term traffic flow prediction with Conv-LSTM," in *2017 9th International Conference on Wireless Communications and Signal Processing (WCSP)*, 2017, pp. 1–6.

[40] W. Lotter, G. Kreiman, and D. Cox, "Deep predictive coding networks for video prediction and unsupervised learning," *arXiv Prepr. arXiv1605.08104*, 2016.

[41] M. F. Stollenga, W. Byeon, M. Liwicki, and J. Schmidhuber, "Parallel multi-

dimensional LSTM, with application to fast biomedical volumetric image segmentation," *Adv. Neural Inf. Process. Syst.*, vol. 28, pp. 2998–3006, 2015.

[42] H. F. Jaafar, A. K. Nandi, and W. Al-Nuaimy, "Detection of exudates in retinal images using a pure splitting technique," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 2010, pp. 6745–6748.

[43] S. Ali *et al.*, "Statistical atlas based exudate segmentation," *Comput. Med. Imaging Graph.*, vol. 37, no. 5–6, pp. 358–368, 2013.

[44] B. Harangi and A. Hajdu, "Detection of exudates in fundus images using a Markovian segmentation model," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 130–133.

[45] C. Pereira, L. Gonçalves, and M. Ferreira, "Exudate segmentation in fundus images using an ant colony optimization approach," *Inf. Sci. (Ny).*, vol. 296, pp. 14–24, 2015.

[46] A. Sopharak, B. Uyyanonvara, S. Barman, and T. H. Williamson, "Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods," *Comput. Med. imaging Graph.*, vol. 32, no. 8, pp. 720–727, 2008.

[47] B. Harangi and A. Hajdu, "Automatic exudate detection by fusing multiple active contours and regionwise classification," *Comput. Biol. Med.*, vol. 54, pp. 156–171, 2014.

[48] E. Imani and H.-R. Pourreza, "A novel method for retinal exudate segmentation using signal separation algorithm," *Comput. Methods Programs Biomed.*, vol. 133, pp. 195–205, 2016.

[49] M. M. Fraz, W. Jahangir, S. Zahid, M. M. Hamayun, and S. A. Barman, "Multiscale segmentation of exudates in retinal images using contextual cues and ensemble classification," *Biomed. Signal Process. Control*, vol. 35, pp. 50–62, 2017.

[50] H. Li and O. Chutatape, "Automated feature extraction in color retinal images by a model based approach," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 2, pp. 246–254, 2004.

[51] I. N. Figueiredo, S. Kumar, C. M. Oliveira, J. D. Ramos, and B. Engquist, "Automated lesion detectors in retinal fundus images," *Comput. Biol. Med.*, vol. 66, pp. 47–65,

2015.

[52] H. Yazid, H. Arof, and H. M. Isa, “Automated identification of exudates and optic disc based on inverse surface thresholding,” *J. Med. Syst.*, vol. 36, no. 3, pp. 1997–2004, 2012.

[53] K. Wisaeng, N. Hiransakolwong, and E. Pothiruk, “Automatic detection of exudates in retinal images based on threshold moving average models,” *Biophysics (Oxf).*, vol. 60, no. 2, pp. 288–297, 2015.

[54] C. I. Sánchez, M. Niemeijer, M. S. A. S. Schulten, M. Abràmoff, and B. van Ginneken, “Improving hard exudate detection in retinal images through a combination of local and contextual information,” in *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2010, pp. 5–8.

[55] L. Giancardo *et al.*, “Exudate-based diabetic macular edema detection in fundus images using publicly available datasets,” *Med. Image Anal.*, vol. 16, no. 1, pp. 216–226, 2012.

[56] M. García, C. I. Sánchez, M. I. López, D. Abásolo, and R. Hornero, “Neural network based detection of hard exudates in retinal images,” *Comput. Methods Programs Biomed.*, vol. 93, no. 1, pp. 9–19, 2009.

[57] K. C. Santosh, N. Das, and S. Ghosh, *Deep Learning Models for Medical Imaging*. Academic Press, 2021.

[58] S. Ghosh, N. Das, I. Das, and U. Maulik, “Understanding deep learning techniques for image segmentation,” *ACM Comput. Surv.*, vol. 52, no. 4, pp. 1–35, 2019.

[59] Y. Zong *et al.*, “U-net based method for automatic hard exudates segm1. Zong, Y. et al. U-net based method for automatic hard exudates segmentation in fundus images using inception module and residual connection. IEEE Access 8, 167225–167235 (2020).entation in fundus image,” *IEEE Access*, vol. 8, pp. 167225–167235, 2020.

[60] Q. Liu, H. Liu, Y. Zhao, and Y. Liang, “Dual-Branch Network With Dual-Sampling Modulated Dice Loss for Hard Exudate Segmentation in Color Fundus Images,” *IEEE J. Biomed. Heal. Informatics*, vol. 26, no. 3, pp. 1091–1102, 2021.

[61] Z. Feng, J. Yang, L. Yao, Y. Qiao, Q. Yu, and X. Xu, “Deep retinal image

segmentation: a FCN-based architecture with short and long skip connections for retinal image segmentation," in *International conference on neural information processing*, 2017, pp. 713–722.

[62] P. Chudzik, B. Al-Diri, F. Calivá, G. Ometto, and A. Hunter, "Exudates segmentation using fully convolutional neural network and auxiliary codebook," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 770–773.

[63] S. K. Saha, B. Fernando, D. Xiao, M.-L. Tay-Kearney, and Y. Kanagasingam, "Deep learning for automatic detection and classification of microaneurysms, hard and soft exudates, and hemorrhages for diabetic retinopathy diagnosis," *Invest. Ophthalmol. Vis. Sci.*, vol. 57, no. 12, p. 5962, 2016.

[64] J. H. Tan *et al.*, "Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network," *Inf. Sci. (Ny).*, vol. 420, pp. 66–76, 2017.

[65] P. Prentašić and S. Lončarić, "Detection of exudates in fundus photographs using deep neural networks and anatomical landmark detection fusion," *Comput. Methods Programs Biomed.*, vol. 137, pp. 281–292, 2016.

[66] S. Yu, D. Xiao, and Y. Kanagasingam, "Exudate detection for diabetic retinopathy with convolutional neural networks," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 1744–1747.

[67] J. Mo, L. Zhang, and Y. Feng, "Exudate-based diabetic macular edema recognition in retinal images using cascaded deep residual networks," *Neurocomputing*, vol. 290, pp. 161–171, 2018.

[68] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual U-Net for medical image segmentation," *J. Med. Imaging*, vol. 6, no. 1, p. 14006, 2019.

[69] W. Chen *et al.*, "Prostate segmentation using 2D bridged U-net," in *2019 International Joint Conference on Neural Networks (IJCNN)*, 2019, pp. 1–7.

[70] D. Nie, Y. Gao, L. Wang, and D. Shen, "ASDNet: attention based semi-supervised deep networks for medical image segmentation," in *International conference on*

medical image computing and computer-assisted intervention, 2018, pp. 370–378.

[71] A. G. Roy, N. Navab, and C. Wachinger, "Concurrent spatial and channel 'squeeze & excitation'in fully convolutional networks," in *International conference on medical image computing and computer-assisted intervention*, 2018, pp. 421–429.

[72] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv Prepr. arXiv1409.0473*, 2014.

[73] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7794–7803.

[74] X. Cheng, F. Lu, X. Han, Y. Yuan, G. Tian, and H. Wu, "Channel Attention Module for Efficient Action Recognition," in *Journal of Physics: Conference Series*, 2022, vol. 2216, no. 1, p. 12090.

[75] A. Rao, J. Park, S. Woo, J.-Y. Lee, and O. Aalami, "Studying the effects of self-attention for medical image analysis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3416–3425.

[76] P. Zhao, J. Zhang, W. Fang, and S. Deng, "SCAU-net: spatial-channel attention U-net for gland segmentation," *Front. Bioeng. Biotechnol.*, vol. 8, p. 670, 2020.

[77] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[78] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[79] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *International conference on machine learning*, 2019, pp. 7354–7363.

[80] J. Zhao, X. Mao, and L. Chen, "Speech emotion recognition using deep 1D & 2D CNN LSTM networks," *Biomed. Signal Process. Control*, vol. 47, pp. 312–323, 2019.

[81] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Bi-directional ConvLSTM U-Net with densley connected convolutions," in *Proceedings of the IEEE/CVF*

international conference on computer vision workshops, 2019, p. 0.

[82] D. Welfer, J. Scharcanski, and D. R. Marinho, "Fovea center detection based on the retina anatomy and mathematical morphology," *Comput. Methods Programs Biomed.*, vol. 104, no. 3, pp. 397–409, 2011.

[83] K. M. Asim, A. Basit, and A. Jalil, "Detection and localization of fovea in human retinal fundus images," in *2012 International Conference on Emerging Technologies*, 2012, pp. 1–5.

[84] M. E. Gegundez-Arias, D. Marin, J. M. Bravo, and A. Suero, "Locating the fovea center position in digital fundus images using thresholding and feature extraction techniques," *Comput. Med. Imaging Graph.*, vol. 37, no. 5–6, pp. 386–393, 2013.

[85] A. Giachetti, L. Ballerini, E. Trucco, and P. J. Wilson, "The use of radial symmetry to localize retinal landmarks," *Comput. Med. Imaging Graph.*, vol. 37, no. 5–6, pp. 369–376, 2013.

[86] A. Tewari, D. Gupta, and J. Sivaswamy, "Bilateral symmetry based approach for joint detection of landmarks in retinal images," in *2014 International Conference on Signal Processing and Communications (SPCOM)*, 2014, pp. 1–6.

[87] A. Aquino, "Establishing the macular grading grid by means of fovea centre detection using anatomical-based and visual-based features," *Comput. Biol. Med.*, vol. 55, pp. 61–73, 2014.

[88] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.

[89] P. Burlina, D. E. Freund, N. Joshi, Y. Wolfson, and N. M. Bressler, "Detection of age-related macular degeneration via deep learning," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016, pp. 184–188.

[90] P. M. Burlina, N. Joshi, M. Pekala, K. D. Pacheco, D. E. Freund, and N. M. Bressler, "Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks," *JAMA Ophthalmol.*, vol. 135, no. 11, pp. 1170–1176, 2017.

[91] J. H. Tan *et al.*, "Age-related macular degeneration detection using deep convolutional

neural network," *Futur. Gener. Comput. Syst.*, vol. 87, pp. 127–135, 2018.

[92] D. W. K. Wong *et al.*, "Automatic detection of the macula in retinal fundus images using seeded mode tracking approach," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012, pp. 4950–4953.

[93] C. Aravindan, V. Sharma, A. Thaarik Ahamed, M. Yadav, and S. Chandran, "Fundus Image-Based Macular Edema Detection Using Convolutional Neural Network," in *Advances in Systems, Control and Automations*, Springer, 2021, pp. 143–153.

[94] Q. Yang *et al.*, "Automated layer segmentation of macular OCT images using dual-scale gradient information," *Opt. Express*, vol. 18, no. 20, pp. 21293–21307, 2010.

[95] A. Fuller, R. Zawadzki, S. Choi, D. Wiley, J. Werner, and B. Hamann, "Segmentation of three-dimensional retinal image data," *IEEE Trans. Vis. Comput. Graph.*, vol. 13, no. 6, pp. 1719–1726, 2007.

[96] H. Ishikawa, D. M. Stein, G. Wollstein, S. Beaton, J. G. Fujimoto, and J. S. Schuman, "Macular segmentation with optical coherence tomography," *Invest. Ophthalmol. Vis. Sci.*, vol. 46, no. 6, pp. 2012–2017, 2005.

[97] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv Prepr. arXiv1409.1556*, 2014.

[98] C. Balakrishna, S. Dadashzadeh, and S. Soltaninejad, "Automatic detection of lumen and media in the IVUS images using U-Net with VGG16 Encoder," *arXiv Prepr. arXiv1806.07554*, 2018.

[99] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[100] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[101] A. Demir, F. Yilmaz, and O. Kose, "Early detection of skin cancer using deep learning architectures: resnet-101 and inception-v3," in *2019 Medical Technologies Congress (TIPTEKNO)*, 2019, pp. 1–4.

[102] S. Kwon, "CLSTM: Deep feature-based speech emotion recognition using the hierarchical ConvLSTM network," *Mathematics*, vol. 8, no. 12, p. 2133, 2020.

[103] A. Arbelle and T. R. Raviv, "Microscopy cell segmentation via convolutional LSTM networks," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 1008–1012.

[104] B. S. Zapata-Impata, P. Gil, and F. Torres, "Learning spatio temporal tactile features with a ConvLSTM for the direction of slip detection," *Sensors*, vol. 19, no. 3, p. 523, 2019.

[105] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[106] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*, Springer, 2017, pp. 240–248.

[107] B. Zoph, V. Vasudevan, J. Shlens, and Q. V Le, "Learning transferable architectures for scalable image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8697–8710.

[108] X. Zhang, Z. Li, C. Change Loy, and D. Lin, "Polynet: A pursuit of structural diversity in very deep networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 718–726.

[109] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V Le, "Autoaugment: Learning augmentation policies from data," *arXiv Prepr. arXiv1805.09501*, 2018.

[110] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, 2019, pp. 6105–6114.

[111] Y. Huang *et al.*, "Gpipe: Efficient training of giant neural networks using pipeline parallelism," *Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.

[112] Z. U. Rehman, S. S. Naqvi, T. M. Khan, M. Arsalan, M. A. Khan, and M. A. Khalil, "Multi-parametric optic disc segmentation using superpixel based feature classification," *Expert Syst. Appl.*, vol. 120, pp. 461–473, 2019.

[113] S. Morales, V. Naranjo, J. Angulo, and M. Alcañiz, "Automatic detection of optic disc based on PCA and mathematical morphology," *IEEE Trans. Med. Imaging*, vol. 32, no. 4, pp. 786–796, 2013.

[114] S. Roychowdhury, D. D. Koozekanani, S. N. Kuchinka, and K. K. Parhi, "Optic disc boundary and vessel origin segmentation of fundus images," *IEEE J. Biomed. Heal. informatics*, vol. 20, no. 6, pp. 1562–1574, 2015.

[115] M. Abdullah, M. M. Fraz, and S. A. Barman, "Localization and segmentation of optic disc in retinal images using circular Hough transform and grow-cut algorithm," *PeerJ*, vol. 4, p. e2003, 2016.

[116] Z. Fan *et al.*, "Optic disk detection in fundus image based on structured learning," *IEEE J. Biomed. Heal. informatics*, vol. 22, no. 1, pp. 224–234, 2017.

[117] M. N. Zahoor and M. M. Fraz, "Fast optic disc segmentation in retina using polar transform," *IEEE Access*, vol. 5, pp. 12293–12300, 2017.

[118] K. S. Nija, C. P. Anupama, V. P. Gopi, and V. S. Anitha, "Automated segmentation of optic disc using statistical region merging and morphological operations," *Phys. Eng. Sci. Med.*, vol. 43, no. 3, pp. 857–869, 2020.

[119] A. S. Abdullah, Y. E. Özok, and J. Rahebi, "A novel method for retinal optic disc detection using bat meta-heuristic algorithm," *Med. Biol. Eng. Comput.*, vol. 56, no. 11, pp. 2015–2024, 2018.

[120] R. G. Ramani and J. J. Shanthamalar, "Improved image processing techniques for optic disc segmentation in retinal fundus images," *Biomed. Signal Process. Control*, vol. 58, p. 101832, 2020.

[121] M. U. Akram, A. Tariq, M. A. Anjum, and M. Y. Javed, "Automated detection of exudates in colored retinal images for diagnosis of diabetic retinopathy," *Appl. Opt.*, vol. 51, no. 20, pp. 4858–4866, 2012.

[122] D. Lokuarachchi, K. Gunarathna, L. Muthumal, and T. Gamage, "Automated detection of exudates in retinal images," in *2019 IEEE 15th International Colloquium on Signal Processing & Its Applications (CSPA)*, 2019, pp. 43–47.

[123] H. Yazid, H. Arof, and H. M. Isa, "Exudates segmentation using inverse surface

adaptive thresholding," *Measurement*, vol. 45, no. 6, pp. 1599–1608, 2012.

[124] D. Welfer, J. Scharcanski, and D. R. Marinho, "A coarse-to-fine strategy for automatically detecting exudates in color eye fundus images," *Comput. Med. imaging Graph.*, vol. 34, no. 3, pp. 228–235, 2010.

[125] Q. Liu *et al.*, "A location-to-segmentation strategy for automatic exudate segmentation in colour retinal fundus images," *Comput. Med. imaging Graph.*, vol. 55, pp. 78–86, 2017.

[126] M. Mateen, J. Wen, N. Nasrullah, S. Sun, and S. Hayat, "Exudate detection for diabetic retinopathy using pretrained convolutional neural networks," *Complexity*, vol. 2020, 2020.

[127] X. Zhang *et al.*, "Exudate detection in color retinal images for mass screening of diabetic retinopathy," *Med. Image Anal.*, vol. 18, no. 7, pp. 1026–1043, 2014.

[128] J. Kaur and D. Mittal, "A generalized method for the segmentation of exudates from pathological retinal fundus images," *Biocybern. Biomed. Eng.*, vol. 38, no. 1, pp. 27–53, 2018.

[129] W. Fang and X. Han, "Spatial and channel attention modulated network for medical image segmentation," 2020.

[130] M. U. Rehman, S. Cho, J. H. Kim, and K. T. Chong, "Bu-net: Brain tumor segmentation using modified u-net architecture," *Electronics*, vol. 9, no. 12, p. 2203, 2020.

[131] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, 2017.

[132] S. Kolhar and J. Jagtap, "Convolutional neural network based encoder-decoder architectures for semantic segmentation of plants," *Ecol. Inform.*, vol. 64, p. 101373, 2021.

[133] F. Zabihollahy, A. Lochbihler, and E. Ukwatta, "Deep learning based approach for fully automated detection and segmentation of hard exudate from retinal images," in *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and*

Functional Imaging, 2019, vol. 10953, pp. 17–22.

[134] S. Guo, K. Wang, H. Kang, T. Liu, Y. Gao, and T. Li, "Bin loss for hard exudates segmentation in fundus images," *Neurocomputing*, vol. 392, pp. 314–324, 2020.

[135] P. Khojasteh *et al.*, "Exudate detection in fundus images using deeply-learnable features," *Comput. Biol. Med.*, vol. 104, pp. 62–69, 2019.

[136] C. Agurto *et al.*, "A multiscale optimization approach to detect exudates in the macula," *IEEE J. Biomed. Heal. informatics*, vol. 18, no. 4, pp. 1328–1336, 2014.

[137] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," 2017.

[138] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*, 2016, pp. 565–571.

[139] M. D. Abràmoff *et al.*, "Automated analysis of retinal images for detection of referable diabetic retinopathy," *JAMA Ophthalmol.*, vol. 131, no. 3, pp. 351–357, 2013.

[140] R. GeethaRamani and L. Balasubramanian, "Macula segmentation and fovea localization employing image processing and heuristic based clustering for automated retinal screening," *Comput. Methods Programs Biomed.*, vol. 160, pp. 153–163, 2018.

[141] B. Antal and A. Hajdu, "A stochastic approach to improve macula detection in retinal images," *Acta Cybern.*, vol. 20, no. 1, pp. 5–15, 2011.

Souvik Maiti
20/11/2023