Using Novel Fundus Image Preprocessing to Improve the Classification of Retinopathy of Prematurity (ROP) Using Deep Learning

Using Novel Fundus Image Preprocessing to Improve the Classification of Retinopathy of Prematurity (ROP) Using Deep Learning

By

Sajid Rahim, B.Sc, BCom (Honours), MSc.

a thesis

submitted to the Department of Computing and Software

and the School of Graduate Studies of McMaster University

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Doctor of Philosophy (August 2024) McMaster University
(Software Engineering) Hamilton, ON, Canada

Title:	Using Novel Fundus Image Preprocessing to Improve the Classification of Retinopathy of Prematurity (ROP) Using Deep Learning
Author:	Sajid Rahim B.Sc, BCom (Honours) (Rhodes University (South Africa), North-West University (South Africa)) MSc. (North-West University (South Africa))
Supervisors:	Dr. Kourosh Sabri, Dr. Alan Wassyng, Dr. Mark Lawford
Number of Pages:	xiv, 223

In memory of my long late father, Abdul Rahim - his last words were "Keep studying and make a difference in people's lives"

Preface

This dissertation is original, unpublished, independent work by the author, S. Rahim. Portions of the dissertation have been published as a pre-print paper on Arxiv site: [98]. I was the lead investigator, responsible for all major areas of collating literature review, data collection and analysis, pre-processing and code generation. I was responsible for the creation of this manuscript. Dr Kourosh Sabri was involved with establishment of the ground truth for the Calgary dataset and subsequent reviews of image pre-processing and CNN results. Dr Alan Wassyng helped with review and editing. Dr Lingyang Chu provided insights into heat maps and segmentation methods while Dr Wenbo He provided guidance on tuning CNNs.

Abstract

Retinopathy of Prematurity (ROP) can affect babies born prematurely. It is a potentially blinding eye disorder which can arise from the complications of undeveloped retina. Thus, screening of ROP is essential for early detection and treatment in these infants. An Retcam digital camera is used to capture patient's retinal image. The captured image quality is constrained due to many factors. Effective and accurate image pre-processing methods for ROP Retcam images are required for improving ROP clinical features prior to being used in any CNN based classification system.

We reviewed present literature on image pre-processing pertaining to digital retina images. This included image domain, restoration based methods and the latest machine learning methods. Our first contributions were two improved novel restoration image pre-processing methods. An image domain method was then applied to the output of these improved methods to create new hybrid methods. These new pre-processing methods improved ROP clinical features in ROP Retcam images. The third contribution used a novel approach using deep learning-based segmentation classifier to generate vessel map from an ROP Retcam image. The purpose was to erode the blood vessels from the original image thereby reducing the blood vessel noise. For our fourth contribution, we used transfer learning based CNNs, namely, InceptionResv2,

and ResNet50 to create 3 sets of classifiers representing each ROP condition, namely Plus Disease, Stages and Zones. These CNNs were trained and validated using the improved pre-processing methods and traditional methods independently. The comparative evaluations of all identified pre-processing methods showed that these new pre-processing methods contributed to higher accuracy when classifying ROP using limited training images. With these methods, our results were as equal or better than comparative peer results using limited data. In this research, using the above components, we created a framework, McROP, that deals with key three ROP conditions. This framework can be extended easily to other pediatric ophthalmology conditions. To our knowledge, this is the first known use of restoration-based image pre-processing for ROP Retcam for improving ROP clinical features. These methods demonstrated effectiveness in CNNs based classification for ROP when compared against traditional pre-processing methods.

Acknowledgments

I would like to recognize my co-supervisor, Dr Kourosh Sabri, Pediatric Ophthalmologist/Professor of Medicine at McMaster Children's hospital who brought in specialized insights into this pediatric challenge and Dr Anna Ells, University of Calgary for supporting this study by making available critically needed ROP dataset. I also acknowledge and thank my friend, Alex Gaudio for detailed blackboard deep discussions in fundus image processing. Dr David Wong, Head of Ophthalmology/Professor of Medicine (University of Toronto) at St. Michael's Hospital, Toronto who triggered the idea to study deep learning applicability for retinal diseases.

My departmental co-supervisors, Mark Lawford and specially Alan Wassyng who supported me during this long journey together with excellent insights from Dr Wenbo He and Dr Lingyang. They kept me focused when it mattered the most.

To my family, especially my daughter Aqsa Rahim who inspired me to embark on this journey to make a difference in people's lives. This journey will not be complete without special acknowledgement to my high school physics teacher - Mrs Jyoti Thygarajan. My university professors namely, Dr Dave Sewry (Dean of Science, Rhodes University, South Africa) and Dr Patrick D. Terry (Rhodes University, South Africa), Dr Eboyo Sepharine (North-West

University, South Africa) who oversaw my entire Computer Science studies and research projects. All played a pivotal part of my educational life and prepared me for reaching this final goal. Lastly a tribute to an old friend David Makola, a teacher by profession and a lecturer who was awarded his PhD posthumously, in 2021 in Cape Town. We started our journey together at Rhodes University as undergraduate students in 1989 when our country, South Africa was on the brink of change. We embarked on our PhDs together but in different continents. Alas, he passed away before he could defend his thesis from Covid-19. I note his personal impact towards my studies as a great student and friend. He will be missed.

Terms

Terms and Abbreviations

BCNN Bayesian Convolutional Neural Networks

CA Chronological Age

CAD Computer-Aided Diagnosis

CMIF Collection of Multispectral Images of the Fundus

CLAHE Contrast Limited Adaptive Histogram Equalisation

CNN Convolutional Neural Networks

DPFR Double Pass Fundus Reflection

DR Diabetic Retinopathy

GMM Gaussian Mixture Model

GT Ground Truth

HRF High-Resolution Fundus

kNN k-Nearest Neighbours

PCA Pixel Colour Amplification

PDR Proliferative Diabetic Retinopathy

PMA Postmenstrual Age

ReLU Rectified Linear Unit

REC Recall

ROC Receiver Operating Characteristic (curve)

ROP Retinopathy of Online Challenge (database)

ROP Retinopathy of Prematurity

SN Sensitivity

TNR True Negative Rate

TPR True Positive Rate

VGG Oxford Group's Variable Geometry Group

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Retinopathy of prematurity (ROP) is a disorder of the developing retinal blood vessels in premature infants and is a leading cause of childhood blindness. There are 3 sections within ROP diagnosis. These are Plus Disease, Stages and Zones which are determined individually. Plus Disease is determined by the tortuosity of the blood vessels, Stages are determined by the presence and intensity of the demarcation line and finally Zones determination is based on how far the first two pathologies are from the center of vision namely the macula. Early screening is critical and early treatment initiated if required. While indirect ophthalmoscopy examination is the preferred method, at times RetCam is also used to capture the patient's retina details. ROP RetCam images quality generally suffers from quality issue such as poor contrast, illumination imbalance. In order to have a clinically viable image, image pre-processing is required [129, 126]. This thesis is devoted to the identification and development of improved novel image pre-processing methods that result in improved ROP clinical specific features such as demarcation line. These resulting images can be used with various automated diagnostic tools including Convolutional

Neural Networks (CNN) for detecting and classifying the respective ROP clinical conditions.

The scope of this thesis was focused based on the experience in the Department of Ophthalmology, McMaster Hospital, Hamilton, Ontario with support from Dr Anna Ells at Calgary University.

Development of accurate ROP classification systems, image pre-processing methods are researched and investigated. Images are captured digitally using a standard RetCam [44] image capturing device that is a standard in pediatrics. Image pre-processing can be done either using image domain or restoration methods, and lately includes machine learning models which are described in later sections. In this research, two novel restoration methods are improvised for ROP RetCam and also combined with traditional image domain methods. All derived and traditional methods are used to process ROP RetCam images. The resulting images are tested and evaluated with CNNs for Plus Disease, Stages (0-3) and Zones (I-III). To date, these novel restoration based methods have not been investigated with ROP RetCam images. We demonstrate that improved images are produced using these new methods, do enable better classifications using CNNs. This can aid in earlier detection of these ROP conditions.

We start by presenting the motivation for this research in Section 1.1. Then, Section 1.2 outlines Retinopathy of Prematurity and challenges within the Canadian Health system. Section 1.3 lays out the research questions and objectives. Section 1.4 outlines the contributions which address the research questions. Section 1.5 provides the overview of contributions towards improving ROP RetCam images for better CNN accuracy. The remaining outline for the thesis is described in Section 1.7.

1.1 Motivation

The thesis's motivation lay in part to my own personal visual impairment following life altering retina issues that required multiple surgeries with ever present complications. This journey started in a way to find a solution that could aid diagnosing vision pathology. As a patient and as a computer scientist, I recognised that the problem of detection of a retinal pathology lay in solving the problems with the retina image quality. This would highlight the pathology which then would allow a CNN to be better trained and thereby classify conditions more accurately. Although I planned to eventually study ROP, my first focus was on Diabetic Retinopathy, which had open data sets available to study this problem while waiting for ROP RetCam imaging data to become available.

Access to Dr Sabri and other retina specialists such as Dr David T Wong, allowed further discussions to occur across multiple areas of challenges related to imaging and operational aspects they face daily. Encouraged by Dr Kourosh Sabri and Dr Alan Wassyng, we retained this focus on diagnosing Retinopathy of Prematurity (ROP). In this pathology, the primary problem is the captured ROP RetCam image quality as noted by [129, 126]. This makes it difficult in identifying finer ROP clinical features that are essential for human diagnosis as well as CNN classification. I focused on solving this challenge primarily using novel image pre-processing algorithms and then use CNN based classification to study the improvements. This thesis is my personal contribution towards this body of knowledge.

1.2 Retinopathy of Prematurity Screening and Challenges within the Canadian Health System

Retinopathy of Prematurity (ROP) is a proliferative retinal vascular disease which afflicts approximately two third of premature infants who weigh less than 1250g at birth [46]. It is a leading cause of childhood blindness. In full term infants, ROP does not occur as the retinal vasculature is fully developed. In premature infants, the development of the retinal blood vessels, which proceeds peripherally from the optic nerve during gestation, is incomplete. Hence, the extent and possibility of immature development of the retina depends on the degree of prematurity. [25] [82].

Globally, 19 million children suffer from visual impairment of which 1.85 million are likely to have developed some degree of ROP. For those diagnosed with ROP, 11% develop severe visual impairment, and 7% develop mild/moderate impairment. ROP Incidence rates have been estimated as 9% in developed countries but higher at 11% in developing countries [106].

ROP can be visualized only in the retina. The retina is a photo-sensitive tissue layer that lines the rear surface of the eye. Most ROP cases do not progress to severe and will resolve without treatment in a short period of time. 5% to 10% percent of patients can progress to severe ROP that can result in retinal detachment. This is the primary cause of blindness if untreated. Therefore, effective screening of ROP is critical for early treatment for better patient prognosis.

Timely detection of ROP allows for early intervention to prevent significant vision loss. Early detection of ROP needs to be improved as ROP can easily be misdiagnosed even by experts. Considering the improvement of neonatal care across Canada and the world, ROP screening programs also have further room for improvement. This will mean, more trained personnel will be needed for the overall screening and management of the condition. At present, this is challenged with the limited number of trained specialists.

ROP is diagnosed using direct clinical examination or by interpreting captured retinal images which are also known as fundus images. These two primary modes of diagnosis result in varying sensitivity and specificity. The impact of these on the infants needs to be considered. Training primary care clinicians may be out of reach depending on regulations. Specific training is needed for retinal screening and grading as well as performance assessment such that the interpretation and grading accuracy is improved. This will be comprised of clinical skills, digital image capturing skills, clinical environment situational awareness, and fundus grading grounding.

As noted earlier, with a growing population, and despite improved neonatal care across entire regions, ROP remains an ever present growing challenge including:

- **Inadequate ROP screening programs.**

  To perform successful ROP screening a trained team of health care professionals is needed. In rural hospitals, there is a challenge as screening is only performed at specialised or designated nearest qualified centers.

- **Training of ROP screeners.**

In order to make ROP screening more effective, there is a need to train health providers who do not have access to a specialist. This includes firstly taking retinal images for ROP correctly. Then to accurately interpret and grade it prior to sending the images to the designated ROP pediatric ophthalmologists electronically for further assessment.

Direct ophthalmoscopy is the gold standard for examination. Digital cameras, such as RetCam [44] are increasingly being used due to their high sensitivity and specificity to support the examination. Fundus camera standardisation, cost, and widespread adoption including training on best use remains limited. These challenges combined, can and do result in longer wait list for initial screening and referral to specialised ophthalmologists resulting in potentially serious eye complications.

Imaging tools such as automated diagnosis or further image pre-processing using these digital images are excellent areas of research that allow for early automated screening of ROP in Canada and other ophthalmology related pathologies.

1.3 Using ROP RetCam images for Diagnosis

As highlighted before, clinical diagnosis remains a challenge which requires visual ophthalmological examination. There is variability between experts diagnosing various aspects of ROP. There are very limited highly trained ROP specialists able and willing to manage ROP in many countries. This impacts patients who may be located across a large geographic area such as Canada. This is additionally hampered by the fact that training of ROP specialists is time intensive. Examining patients is itself a challenge given the age of these

premature infants. Providing improved quality of care by enabling specialists to remotely diagnose patients can be very beneficial, especially in remote locations in Canada and many other parts of the world.

Using a captured ROP retinal image, an automated computer-based image analysis/diagnosis of ROP can provide a support system for early detection. The preferred pediatric ophthalmological camera system used globally is RetCam [44]. Lately, some teams have opted to use Smartphone cameras using MilRetCam [43]. Different capture systems have different resolution, depth and field of view quality that requires standardisation for usage in deep learning, especially with intermixed sources.

Other factors to be considered include, image variability due to retinal pigmentation that varies across ethnicity and geographic regions. Premature infants have more visible choroid vessels which also can challenge classification [101, 36]. Meeting these challenges is essential for clinicians who may be considering using machine learning techniques in clinical practice for screening, diagnosis, and monitoring post treatment.

Captured RetCam images of a pediatric patient's retina suffer from quality issues [129, 126]. Issues include blurry images due to motion of the patient or extra pressure on the eye when placing the camera. Aberration and artifacts appear in the captured image due to incorrect placement of camera. When properly captured, the image still suffers from poor contrast and illumination imbalance. These factors reduce the visibility of ROP clinical features in the captured image. These clinical features include the clear presence of demarcation line, blood vessels, macula and optic disc. To enhance these clinical features for use in any automated diagnosis, image pre-processing is required [129, 126]. This thesis focused on solving challenges related to ROP

RetCam image quality by improving ROP clinical features by using improved image pre-processing methods.

Review of relevant literature in Section 3.4.3 showed that CNNs are the most effective methods for an image-based diagnosis/classification task for retinal pathology. Therefore, in this research, we focused only on CNN based methods. When using CNN for ROP classification, researchers primarily used traditional image pre-processing methods to overcome ROP quality challenges. This thesis proposed using novel pre-processing image methods on pediatric fundus images captured by a RetCam imaging camera, held the key to better classification of the conditions by the deep learning classifiers by improving salient ROP features in the fundus image. In this work, we proposed two improved novel pre-processing methods for improving ROP RetCam image quality. The output images from these methods were further enhanced using existing traditional image pre-processing techniques referred to here as hybrid methods. We also employed a CNN based segmentation approach for novel way of eroding blood vessels in the ROP RetCam image. We highlighted the resizing problem uncovered when using existing resizing algorithms with pre-processed ROP RetCam images and provided recommendations. The recommended resizing methods avoided the loss of key ROP clinical features. We then conclusively demonstrated that the improved quality of ROP RetCam images using the improved novel and hybrid methods led to better diagnosis of ROP disease in terms of quality and consistency when used with a CNN based classifier. These pre-processed images can also be used in a clinical setting to improve practical ROP diagnosis and classification.

1.4 Research Aims and Objectives

Artificial intelligence based systems are starting to revolutionize the field of fundus based pathology detection for patients globally. While the success has primarily been focused on adults eye pathologies such as diabetic retinopathy (DR), very little work has been carried out in the field pediatric ophthalmology due to various factors. As ROP cases increase, the demand for pediatric ophthalmologists is increasing.

The primary goal of this research was to identify and resolve the quality problem associated with ROP RetCam images. Then to use resulting ROP clinical feature enriched RetCam images with CNN based models classification to study the respective pre-processing methods effectiveness.

The overall system, we created is referred to as McROP, described in Section 4.1. The project's first step was to remediate the ROP RetCam image quality by identifying novel image pre-processing techniques and associated clinical improvements for ROP RetCam images. Clinically significant ROP features required included better contrast, clear visibility of blood vessels, demarcation line, macula and optic disc. These pre-processed images were then used with 2 CNNs which separately diagnose Plus disease, Stages of ROP and Zones independent of each other. Training, and validation was repeated for each different pre-processed image datasets using these two deep learning classifiers for Plus Disease, Stages and Zones for ROP. The final validation results were compared for each pre-processed dataset. The results demonstrated that for the same pair of ROP RetCam images dataset (training and validation), the improved image pre-processing methods did improve CNN classifications

results for Plus Disease, Stages, and Zones in comparison to traditional methods.

As we will see in Chapter 2, clinical classification of ROP depends on identifying visual pathologies in the infant's retina. This classification is hampered by low quality of retinal images that are possible with premature infants. Since we know what these pathologies are, we can design image pre-processing methods to enhance the images such that relevant pathologies are included in the images used for machine learning clinical classification of ROP. This gives rise the following research questions.

RQ1. Understanding ROP RetCam images features specific to ROP Classification. Which image pre-processing methods could be examined to resolve quality problems? How could we use these methods to improve the image quality such that the appropriate ROP features can be highlighted namely the blood vessels for Plus Disease, and demarcation line for Stages? What were the potential gaps with respect to ROP RetCam we need to solve such that the quality of ROP RetCam images are improved with respect to ROP features?

RQ2 To better improve ROP RetCam images, could this be improved further by reducing artifacts such as blood vessels from the image such that only the demarcation line was present? Could we explore using CNN based segmentation instead of classical morphological tracing by making use of CNN based segmentation?

RQ3 Validation of each pre-processing method for classification. Given that we did not have adequate training, could we use transfer learning to assist us with ROP Classifications? Investigate different CNNs frameworks. Could we find two or more that are suitable for our use?

RQ4 Given lack of ROP data, could we use data augmentation in such a way that each image is unique for training purposes?

Chapters 4, 5, and 6 of this thesis, focused on these research questions.

1.5 Thesis Contributions

The primary contributions of this work sought to answer the prior questions and includes:

- **Identifying novel image pre-processing methods to improving ROP RetCam image for CNN usage.** To date, no restoration methods had been used for any ROP captured digital images. To understand how ROP RetCam images could be pre-processed, it was important to understand the fundamentals of restoration methods. Two existing novel candidates were studied, and their underlying problems were identified. The first method known as Pixel Colour Amplification [34] (PCA) is challenged with illumination uncertainty and red colour saturation. The second method known as Double Pass Fundus Reflection [145] (DPFR) resolved fundus reflection problem but is unsuitable for ROP RetCam images due to grainy images making it unsuitable for Stage disease classification. We first focused on PCA and solved the auto-illumination problem and reduced the red colour saturation. DPFR was improved upon to reduce the grainy nature including colour being attenuated. These contributions in the form of new improved image pre-processing for ROP RetCam images significantly enhanced ROP clinically associated features such as demarcation line, blood vessels, macula and optic

disc while resolving contrast and colour saturation required for use in CNNs.

- **Segmentation approach to eroding blood vessels in an ROP Retcam** To further improve pre-processed images for ROP Stage classification, we attempted to reduce the presence of blood vessels in a ROP RetCam image by eroding the vessels. A novel approach was used that made use of an existing minimalistic U-Net [104] CNN based LWNET [31]. First, LWNET CNN was trained using the DRIVE dataset that contained adult fundus images with ground truth retina vessel maps. The trained LWNET, was then used to predict and generate a premature infant's ROP RetCam segmentation blood vessel map. We then implemented a novel approach for vessel erosion. It eroded the vessels from the original ROP RetCam image by using its related generated segmentation map. The result was an image without blood vessels but replaced by a blended background area. This allowed the visibility of demarcation line without the presence of blood vessel noise.

- **Developing effective CNNs for ROP detection.** Implement and evaluate two deep learning models CNNs namely ResNet50, Inception-ResV2 with transfer learning with layers for Plus, Stages (0-3) and Zones (I-III) classifications. Each was trained and validated separately for each pre-processing method. It was used to demonstrate that the image pre-processing method(s) when used in machine learning techniques can assist classify ROP conditions from pre-processed images. This identified new improved and hybrid methods that yield significantly better comparative results vs traditional methods.

- **Resizing and Augmentation.** Resizing is a standard process prior to usage in CNNs. We examined the different algorithms available and recommend the optimum for ROP RetCam which retains the ROP data well. To date, this is the only research which highlighted this for ROP RetCam images. Augmentation normally is used to assist in generating unique images when we do not have sufficient images for training. The newer derived pre-processing methods can be considered as new ROP data augmentation techniques.

- **Labelled ROP Dataset repository.** Across all literature and peer groups, ROP data is not readily available. We have accurately labelled and sub-categorised the data by Plus disease, Stages, and Zone. This data is available to researchers with prior authorisation from Dr Sabri.

1.6 Research Methodology

The research is focused on image pre-processing that is the most critical component of the entire system which aims to improve the ROP features in an ROP RetCam captured image. Without quality images, the diagnosis and classification methods will not deliver quality results. Figure 1.1 below highlights the components of the system. These components are image pre-processing, augmentation, resizing and CNN based classification. To evaluate our system results, we compared them to published papers with comparative setup and experiments.

Our framework is referred to as McROP shown in Figure 4.2. The 3 types of systems developed are for Plus Disease, Stages 0-3 and Zones classifications.

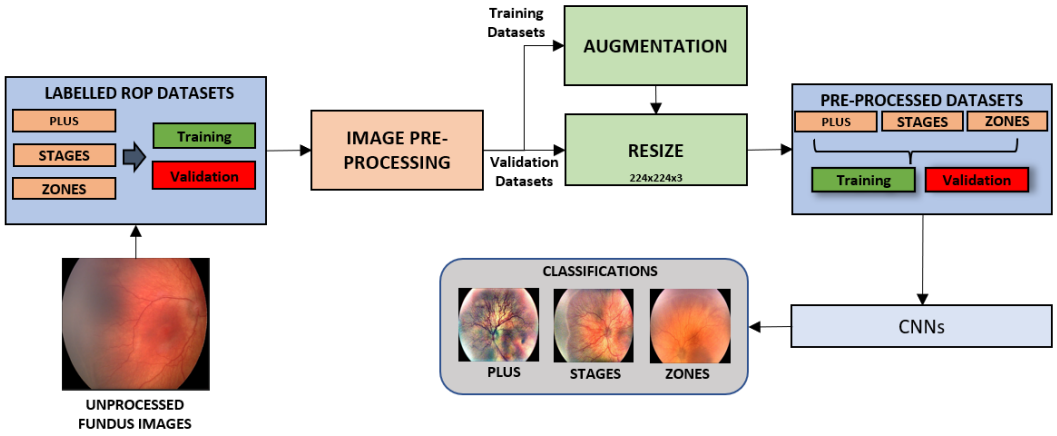


Figure 1.1: Processing Flow

1.7 Thesis Outline

Thesis proposal is organised into six chapters. Each chapter focuses on different aspects of the research work.

- Chapter 1 provides an overview about Retinopathy Of Prematurity and challenges within the Canadian Health system. The core research aims and objectives of this thesis are laid out with respect to ROP. Contributions and methodology are outlined.

- Chapter 2 provides background on Retinopathy of Prematurity and its prevalence. classification, and its screening. A brief overview on computer aided diagnosis and its components which are then explored as part of the overall thesis.

- Chapter 3 discusses image pre-processing as well as the CNN frameworks to be considered. It highlights related work in these areas specific to retinopathy of prematurity as well as references to related diabetic retinopathy.

- Chapter 4 discusses the research approach and data collection. The chapter discusses the guidelines and specific steps followed to develop experimental datasets for Plus disease, Stages and Zones based on medical guidelines are highlighted. Data augmentation requirements and steps followed are described that allow a larger pool of datasets creation for ROP CNN training.

- Chapter 5 provides the thesis's contributions towards enabling better detection of ROP pathologies.

- Chapter 6 tabulates and discusses the results from CNNs for Plus Disease, Stages and Zones using our new pre-processing methods as well as existing pre-processing methods.

- Chapter 7 concludes with a summary of accomplishments and results. It discusses areas of future research that were discovered during this .

Chapter 2

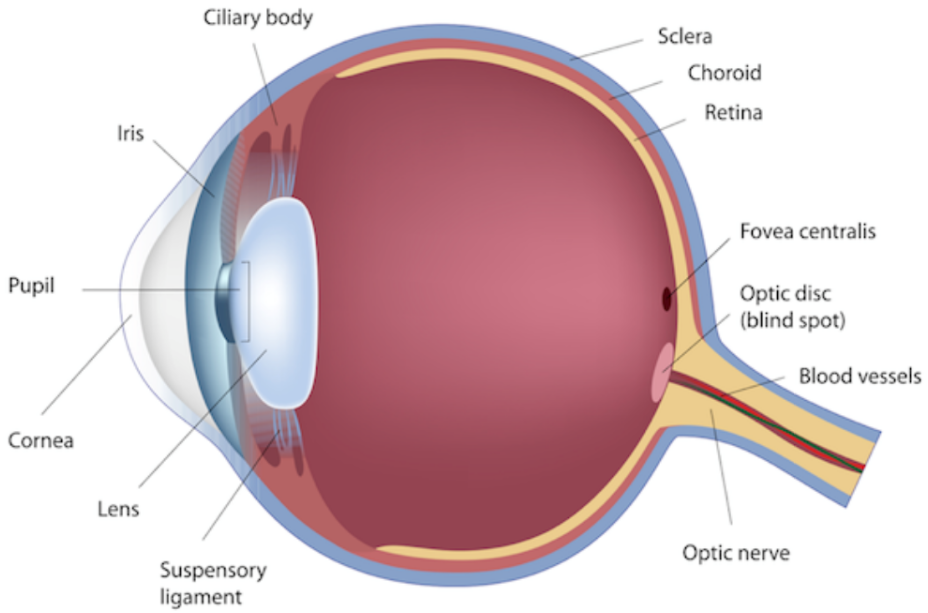
Background

This chapter provides the background for core areas focused on in this research. It first provides insight into the physiology of the eye, classification of ROP, screening for ROP including the use of digital cameras for capturing the retina image (also known as fundus), the usage of computer aided diagnosis for ROP and its core components. It is these core components which are the areas of our interest for techniques that seek to improve the diagnostic outcome. Section 2.1 which describes the physiology of the eye and uses this to provide insight and explanation on ROP. Section 2.2 provides the detailed overview of ROP in terms of Plus Disease, Stages, and Zones. Section 2.3 highlights how screening is performed for ROP and usage of digital cameras for capturing retinal fundus images of a premature infant. Section 2.4 notes the use of computer assisted diagnostics and how it can help with ROP diagnosis. Components that make up a computer assisted diagnostics are explained in Section 2.5 which include image pre-processing, feature extraction and pathology classification. This forms the base of our research contributions. The summary of this chapter is noted in Section 2.6.

2.1 Eye structure and Retinopathy of Prematurity

A human eye is a visual sensory organ that is made up of many parts forming a very complex structure [80]. The outer part of the eye is called sclera that also acts as a protective tissue in which the eye cavity is placed. The eye cavity is filled with an aqueous humour. The inner eye is comprised of the iris, the pupil, the lens, the retinal layer, blood vessels and the optic nerve. The iris is responsible for controlling how much light enters the eye through the pupil by expanding or contracting. The lens which is flexible, has peripheral muscles which change the shape of the lens. This changing the lens allows the light to be focused onto the light sensitive area called the retina or fundus. The structure of a human eye is presented in Figure 2.1.

Figure 2.1: Anatomy of a human eye [135]



The retina which is the layer on the inner rear of the eye, receives light. It converts it into visual signals that are relayed to the brain. In a premature baby, the development of blood vessels beneath the retina can exhibit abnormal growth. This happens during the retina's peripheral development away from the optic nerve. During gestation, this development is incomplete. In full-term infants, the retina as well as associated vessels are completely developed. Therefore, ROP does not occur in full-term infants. In pre-mature infants, most of ROP will resolve without any retinal complications. However, when ROP is severe, retinal detachment can occur. In other words, the retinal layer can separate away from the wall of the eye. This can lead to blindness if not treated promptly. Infants who are born prematurely at the age of 31 weeks of gestation and below the weight of 1250 grams are at the highest

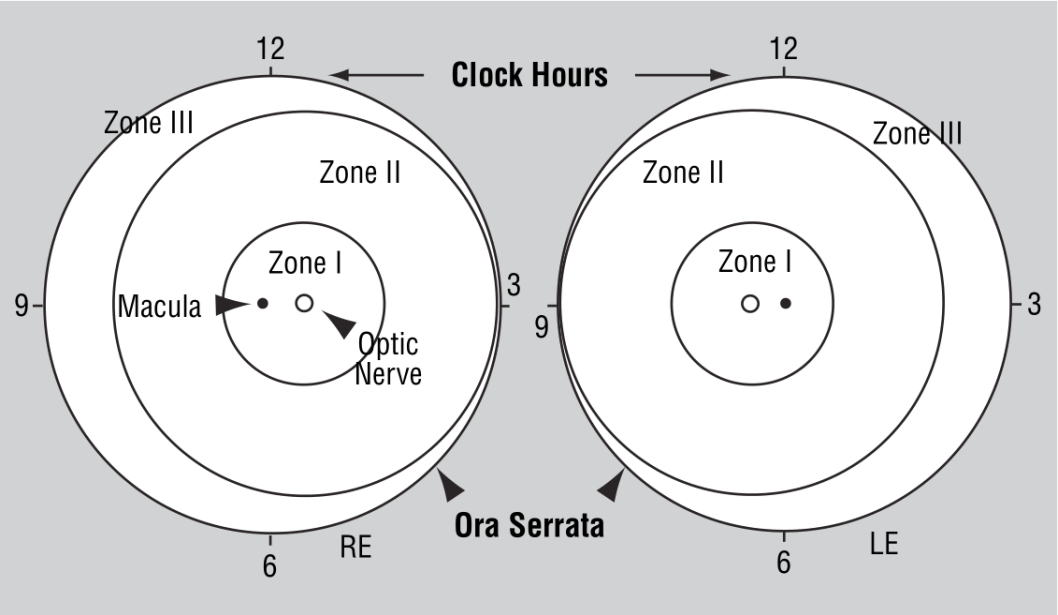
risk. Approximately 50%-60% of premature infants are affected with ROP. Almost 90% will be of the milder category and do not need treatment. Of the remaining, ROP is severe enough to require medical treatment [46, 106, 96].

With advances in neonatal care, premature mortality rate has dropped significantly over the past several decades resulting in increased survivability of youngest born premature infants near to or lower than 31 weeks and 1250 grams in weight and resulting incidents of ROP are proportionally increasing [46, 96].

2.2 Classification of Retinopathy of Prematurity

There are several examination aspects required to determine the classification of ROP. The determination is based off schema of retina in both left and right eye as noted in Figure 2.2.

Figure 2.2: Scheme of retina showing zones/hours to show location and extent of ROP [46]



The optic disc is considered as the center as normal retinal vasculature proceeds peripherally from it towards ora serreta. The outer area from this center, each circular Zone is presented around it. Normal retinal vasculature proceeds outwards from the center of the optic disc towards the ora serrata [102]. *Ora serreta is the peripheral termination of the retina* [102]. Location of Zones is based on the distance between the center of the optic nerve to the center of the macula. Twice the value of this distance from the center of the optic disc is the Location of Zone I radius [1, 46, 117]. Zone II describes the region from the boundary of Zone I to the nasal ora serrata [1, 46, 117]. Zone III is the residual temporal crescent of the retina anterior to Zone II [1, 46, 117].

ROP is classified based on the presence of Plus disease, Stages, and lastly where these pathologies are in terms of Zones defined earlier. Stages are primarily ascertaining the presence of a demarcation line which appears from Stage 1 and gathers in intensity up to Stage 3. If the patient is not treated on time, it progresses to Stage 4, that is a partial detachment leading to Stage 5 that is a full retinal detachment. Plus disease [117] denotes vascular dilation and tortuosity of posterior retinal vessels in at least 2 quadrants of the eye. Determination of Stages, Plus disease and which Zone these pathologies are located in denotes ROP. These can be summarised in the following list:

- **Stage 0**: Absence of any demarcation line features.
- **Stage 1**: Presence of demarcation line separates vascularised and non-vascularised retina. Figure 2.3.A.
- **Stage 2**: Ridge arising in the region of demarcation line. Figure 2.3.B.
- **Stage 3**: Extraretinal fibrovascular proliferation. Figure 2.3.C.
- **Stage 4**: Partial retinal detachment. Figure 2.3.D.
- **Stage 5**: Full retinal detachment.
- **Pre-plus**: Blood vessels exhibit elevated tortuosity than normal.
- **Plus**: Blood vessels show abnormal enlargement and tortuosity indicating worsening of disease. Figure 2.4.

Figure 2.3: RetCam images showing ROP Stages 1, 2, 3 and 4.[25]

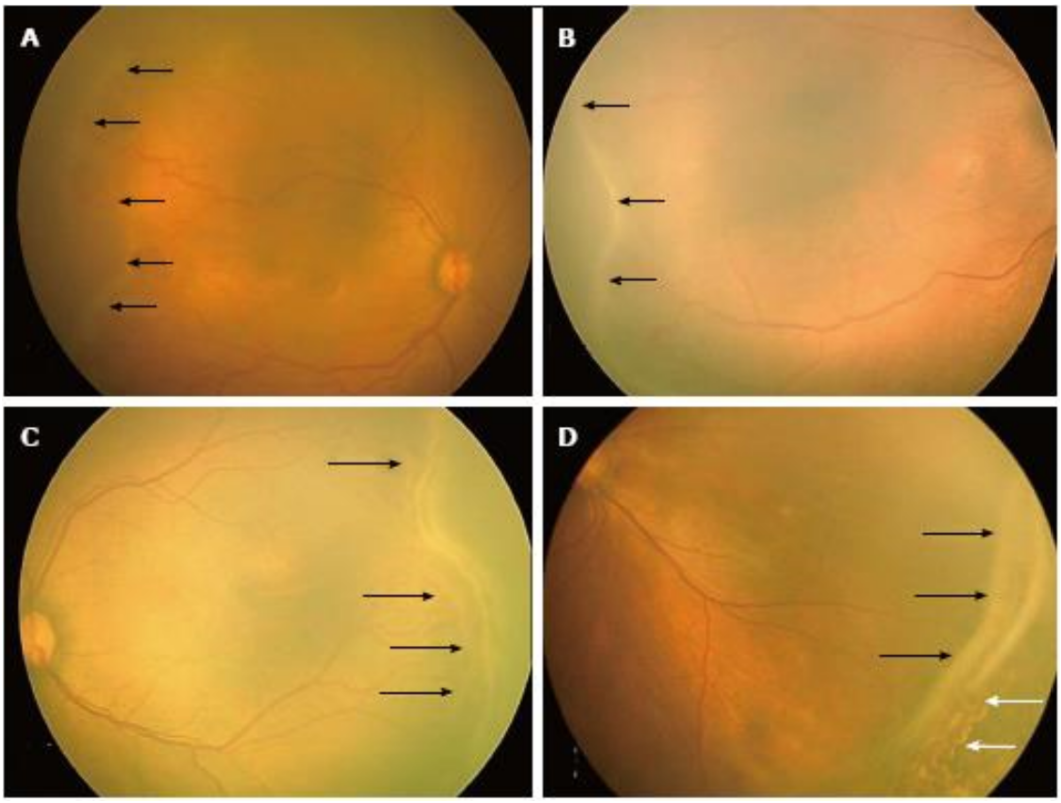
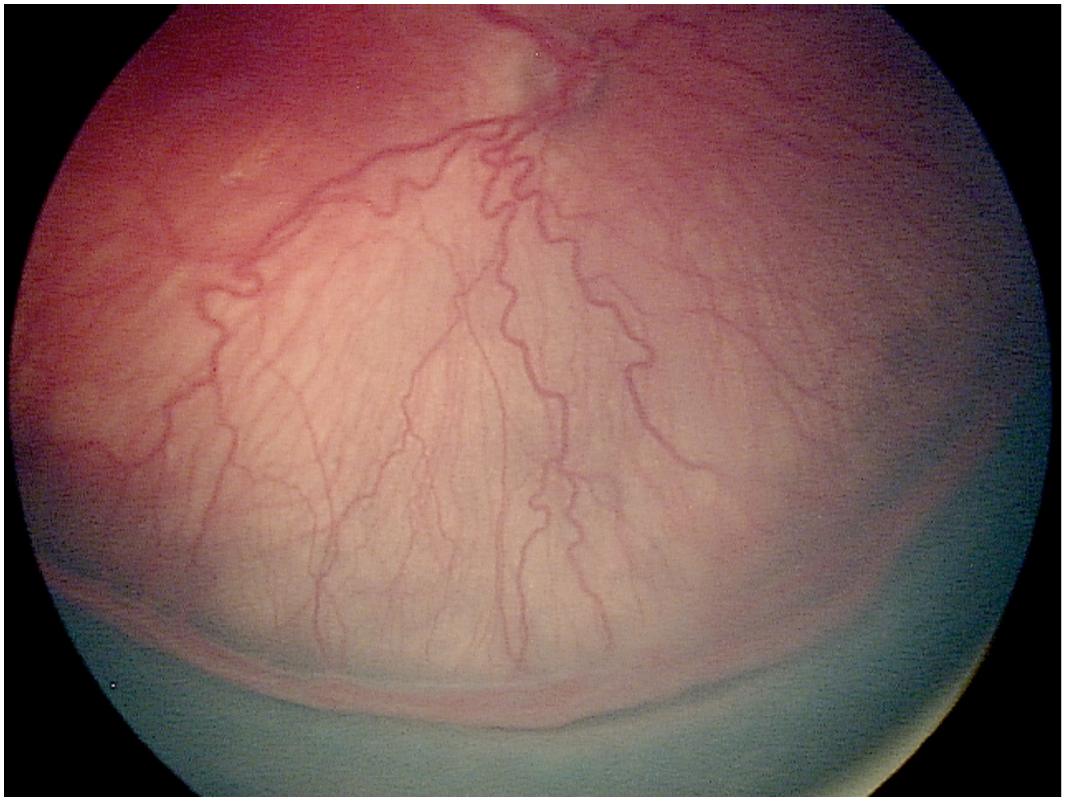


Figure 2.4: RetCam image showing Plus disease with Stage 3.[27]



ROP guidelines [46] note that in most cases the patients have either Stage 1 or 2. However in some cases, the ROP condition may worsen and if not treated on time, can result in severe visual impairment and probability of vision loss. This includes, patient with Stage 4,5 requires urgent medical intervention while those with Stage 3 combined with Plus disease may be considered for treatment.

Figure 2.5: Follow up Schedule for ROP [25]

Zone of retinal findings	*Stage of retinal findings*	*Follow up interval*
Zone 1	Immature vascularization	1-2 weeks
	Stage 1 or 2	1 week or less
	Regressing ROP	1-2 weeks
Zone 2	Immature vascularization	2-3 weeks
	Stage 1	2 weeks
	Stage 2	1-2 weeks
	Stage 3	1 week or less
	Regressing ROP	1-2 weeks
Zone 3	Stage 1 or 2	2-3 weeks
	Regressing ROP	2-3 weeks

2.3 Screening of Retinopathy of Prematurity

There is no general consensus when to screen. The general guidelines is for screening infants born at 30 weeks or less [25, 46]. Their birth weight should be 1250 grams or less [25, 46]. For most immature infants, the risks of ROP are the highest as the condition takes a long time to develop. 99% of pre-threshold ROP develops by 45 weeks PMA [46]. The recommendations is to identify the eyes with highest potential for required treatment at the earliest. This approach also reduces the examinations required for low risk infants. There is a pre-determined timing of initial examination. This is based on the both postmenstrual age (PMA) and chronological age (CA). This allows

99% of infants to be examined that have high risk. Subsequent follow ups are required depending on the ROP stage diagnosis. This ROP examination may be discontinued once there is no longer any risk for development severe ROP.

Treatment for ROP, include Laser photocoagulation and Anti-Vascular Endothelial Growth Factor (Anti-VEGF) eye injections. Anti-VEGF can consist of either Bevacizumab or Ranibizumab or a combination thereof. Surgery is prescribed for most advanced cases which includes vitrectomy (removal of vitreous gel or scleral buckle where a band is placed around the outside of the eye [25, 46].

There are two methods of screening, namely indirect ophthalmoscopy and digital fundus photography. With the indirect ophthalmoscopy method (Figure 2.6) the examiner's interpretations of the clinical findings are recorded manually on a grading sheets. No digital imaging is performed. This method of recording assumes clinician's reading of the findings to be valid. Digital image recording and labelling would allow the confirmation of diagnosis if required using a second opinion. Additionally, this would be beneficial in assessing the therapeutic outcomes of full or partial treatments. With TeleHealth now available in Ontario, there is opportunity for these digital images to be merged with deep learning algorithms which is the focus of this thesis. Nevertheless despite potential benefits, there are significant challenges related to image capture as the event, in itself, is traumatic for the infant. Once the images have been captured, this data is archived for record and can be used to monitor progress.

Figure 2.6: Indirect Ophthalmoscopy [82]

Figure 2.7: Digital Fundus photography [82]

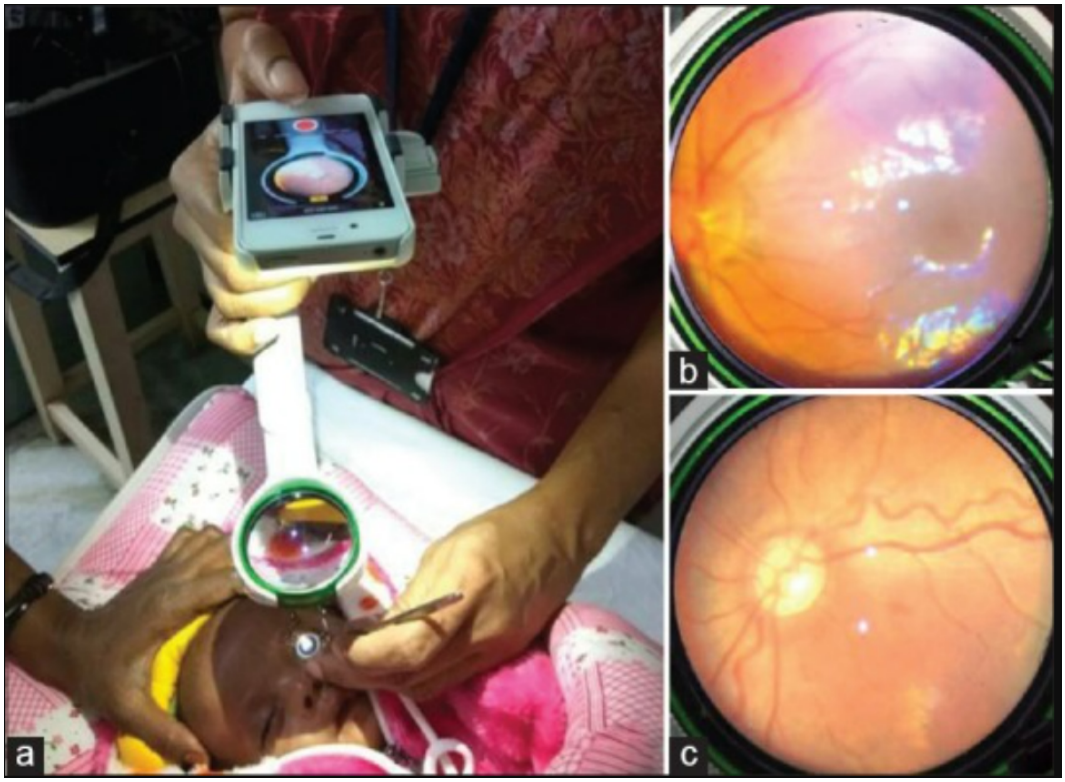
The use of digital wide field cameras to capture retinal images is gaining traction. This allows digital images to be transmitted for remote specialist interpretations. This allows the development of alternative approach to in person ophthalmoscopic ROP screening especially for remote locations within a country or globally. Digital fundus photography is utilised at present as a screening tool for a wide variety of eye diseases in adults. There is merit for

incorporating this for ROP screening and diagnosis (Figure 2.7). Many studies also reported *no difference in overall accuracy between ophthalmoscopy and telemedicine for detection of clinically significant ROP* [7] and in some cases d*igital fundus photography provided for better diagnostic accuracy than the local ophthalmologist for ROP* [136] hence making it possible for Telemedicine to be as accurate as a regular exam.

A wide field imaging camera known as a RetCam [44] is shown in Figure 2.8 is being used in some instances as an alternative to indirect ophthalmoscope for remote locations. It provides variable sensitivity, good specificity, and ease of use. It can provide high quality digital photos of the interior surface such as retina, optic nerve, and peripheral areas. The digital images captured can be used to identify and track disease progression. Quality for our use implies ability to show the areas as identify earlier in Figure 2.2. RetCam is expensive and requires additional training for correct use. Smartphone cameras are now more universally available, and as such have also been adapted for use for similar functions [43] as shown in Figure 2.9. Smartphones can produce excellent quality but have a narrow field of vision unlike RetCam which provides a wider field.

Figure 2.8: Wide Angle Imaging (Retcam)[44]

Figure 2.9: Smartphone based fundus imaging [63]

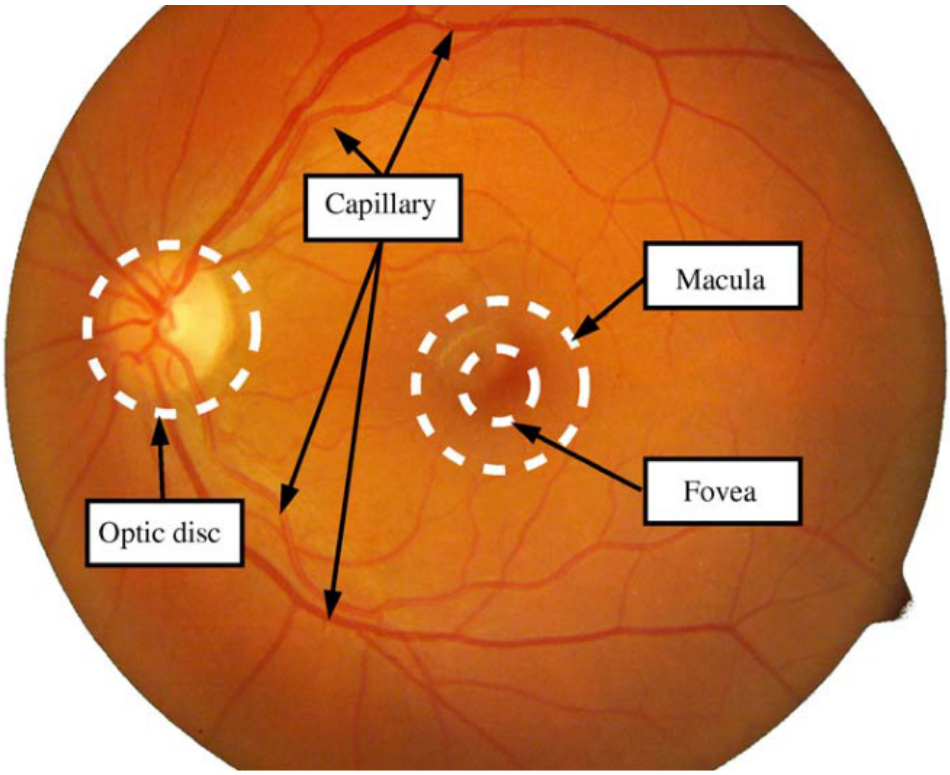


2.4 Using Computer Aided Diagnosis

A human eye's pupil shrinks in the presence of light. This prevents a detailed view of the fundus of the eye. In order to perform a detailed examination, dilation of the pupil is carried by putting dilating drops in the eye. This allows the pupil to be dilated allowing for a thorough examination by the specialist. In order to capture this detail, there are two kinds of cameras for this purpose, namely mydriatic and non-mydriatic. For mydriatic cameras, the pupil of the eye needs to be dilated similar to a specialist examining a patient's eye. This allows a larger view of the retina. While in non-mydriatic,

dilation is not needed but resolution, quality and area coverage can suffer. In either case, special filters can also be used in case of retinopathy. The end result is an image which highlights the retina's surface details including blood vessels, macula and optic disc as shown in Figure 2.10.

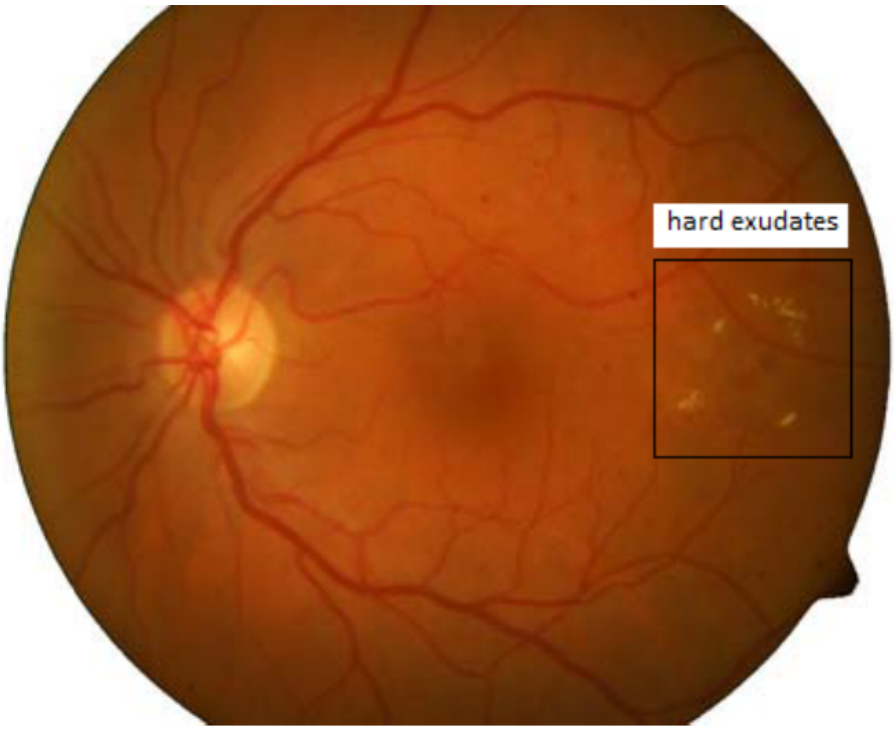
Figure 2.10: Labelled Retina image [52]



By using a fast and efficient way of obtaining a retinal image also known as fundus image, in conjunction with medical expert knowledge, it is feasible to perform diagnosis of various eye conditions. For example, signs of tortuosity and the presence of a demarcation line are markers that indicate potential areas of interest which can diagnose conditions such as ROP. Similarly, in adults, exhudates can show signs of Diabetic Retinopathy (DR) as shown in Figure 2.11. Therefore, by using a digital retinal image having the clinical

specific features, the question arises if we are able to train a CNN to recognize various pathology features we are one step closer to diagnosis. For better accuracy, the fundus image quality must be clear and showing the clinical features desired. This image improvement is possible by using various image pre-processing methods.

Figure 2.11: Example of retina with Exudates [52]



Computer Aided Diagnosing (CAD) is rapidly improving to the level where, at least, initial prediction is possible. Using a family of machine learning techniques that allows a CNN to learn about key features directly from images using large training labelled datasets without specifying rules or features directly. Once the model is built and by presenting it a new patient's image, a pathology can be detected. To date, there have been numerous successful at-

tempts using image pre-processing and classification methods to detect various pathologies [73, 12, 41, 94] such as DR, and breast cancer to name a couple.

In order to model a Computer Aided Diagnosing tool as with DR but ROP, we need to define this task as a multi-class classification. For Plus disease classification, two class method is used so that we determine whether there is Plus disease or not. However we can use multi-class classification by adding various Stages of ROP or Zones as described earlier in Section 2.2. In the case of DR, many classification techniques such as CNNs, support vector machines (SVM) as well as statistical methods have been used with a good success rate. In the case of ROP, there are very limited scope attempts made especially with Deep Learning using CNNs and novel image pre-processing methods.

Figure 2.12: Public Available Databases [50, 70, 116]

Database	Number of images	Manually segmented images	Resolution	Application
EYEPACS-1	12000	n/a	Various kinds	Multiple
MESSIDOR-2	1728	n/a	3 kinds	Blood vessel segmentation
STARE	400	n/a	1280x960	Blood vessel segmentation

Research in various adult eye diseases has been carried out for decades. Consequently, there are large databases available with labelled eye fundus images noted in Figure 2.12. These are limited to adults and no such database is available for ROP in the public domain at the present time. In order to create a new ROP Dataset, several steps were required. These are discussed in Section 4.2. We commenced with a process of collection, filtering, and anonymization from collaborating specialist groups. New images were filtered for acceptance criteria, then categorised, and labelled with diagnosis for ground

truth. Dr Anna Ells was most supportive and assisted us with the Calgary ROP Dataset. During the long wait for the Calgary ROP Dataset, we made use of the Kaggle EYEPACS [50], MESSIDOR [70], and STARE [116] for prototyping image pre-processing and CNN models. Kaggle contains primarily DR images at various stages of pathology including different types and models of cameras under various illumination. Actual ROP images are very different as they show underdeveloped retinas vs fully developed adult retinas.

2.5 Image Pre-Processing of Fundus Images

Computer Aided Diagnosis (CAD) is comprised of three components that work in sequence: Image Pre-processor, Feature Extraction and Classification.

Image Pre-Processing: This is a process when applied to an original image results in enhanced features. These improved features can be useful in the subsequent steps for feature extraction. Pre-processing can include pixel and geometric transformation, local pre-processing, filtering, and image restoration amongst many techniques. There is not a one fit approach but a combination depending on the application area. We discuss various techniques in Chapter 3 including restoration methods which we improved upon for ROP RetCam images.

Feature Extraction: feature extraction methods are used to detect and separate features from the digital image such as vascular segments, optic nerve head, exudates, demarcation line depending on the targeted pathology. Additional details such as area, vessel density and width, as well as intensity of features such as saturation density amongst others to detect pathology severity. We propose a novel way of feature reduction by leveraging an existing

segmentation improvised CNN framework to segment vascular and then use it to remove it from the original image.

Classification: Classification algorithms work on feature rich images to detect pathologies. Classification is categorised into two types namely supervised and unsupervised classifiers. A supervised classifier takes input from the user on a set of samples to determine the data structure from which discriminant functions are derived. An unsupervised classifier uses cluster analysis to derive the data structure composition from the data itself. SVM have been widely used for supervised learning. Examples of their use includes classification and predictions using machine learning theory for maximising prediction accuracy without overfitting data. Neural networks are the newer and more powerful alternative and have shown great promise in medical diagnosis.

A neural network is made of many neurons or nodes that perform two functions: aggregation of its input from other neurons and generation of output from aggregated inputs. A Convolutional Neural Network (CNN) is constructed with an input and output layer with several hidden middle layers. The CNN is able to identify the precise mathematical arrangements that transform the input into the output through using a linear or nonlinear relationship. The network calculates the probability of each output while moving across the various layers. There are many variants of CNN that have been used for different problem areas including retinal pathology detection.

2.6 Chapter Summary

This chapter highlights the challenges posed with increasing cases of premature infants and the impact that ROP will have on their quality of life. It

is critical that these infants be screened for early detection of this condition. It summarised different aspects of classification within ROP namely that of Plus Disease, Stages and Zones. Image pre-processing is reviewed briefly and has been proven to assist in DR as well as other pathologies. The thesis focuses on this opportunity to improve the image quality for Retinopathy of prematurity. Feature extraction and machine learning for classification are briefly discussed. Subsequent chapters demonstrate the use of novel and traditional pre-processing techniques including segmentation based CNNs with deep learning classification techniques using transfer learning.

Chapter 3

Fundus Image Pre-Processing and Related Works

Image pre-processing techniques are used widely for diagnosing various pathologies including adult eye conditions. Computer based imaging tools are increasingly playing a vital role in identifying various ophthalmological pathologies such as diabetes retinopathy. This identification allows for earlier preventative actions that can prevent vision loss. There is very active research being carried out in diagnosis of various eye disorder screening by advocating automated systems for detecting, and classifying various pathologies.

This chapter starts with examination of image pre-processing in section 3.1, especially image restoration methods which are focused upon in this research. It is followed by section 3.2 which discusses the Image segmentation problem which is leveraged for removal of blood vessels. Section 3.3 discusses deep learning and the key convolutional networks we will use later. Section 3.4 examines related works in these fields especially with respect to fundus images for both DR and then in ROP.

3.1 Fundus Image Pre-Processing

In the application of CNNs to the ROP classification problem, we recognised that the crucial factor is the patients' RetCam images quality as also noted by Vinekar et al., [129] and Valikodath et al., [126]. Image pre-processing for paediatric RetCam captured images, a critical part of our work, is now reviewed. Image pre-processing can be defined as either the extraction, enhancing, or removal of features within the image. This includes pixel brightness and geometric transformations [114, 149]. These methods are grouped either in image domain or restoration method categories. Most widely used methods fall in the first category and are popular primarily due to their simplicity. A second class of methods is categorised as restoration and are complex. The restoration class of methods are fairly new with novel techniques based on similar principles. The former set of methods have primarily been used for ROP. The restoration category has not yet been documented in the improvement of ROP fundus image quality. We define ROP fundus image quality as the ability to improve ROP clinical features in the captured image. To the best of our knowledge, this is the first time where restoration methods are used to improve the ROP RetCam quality. Described below are the primary methods used in both classes which are used and/or extended by this research.

3.1.1 Image domain methods

Fundus images are challenged significantly with noise that includes reflection, artefacts and out of focus capturing. As previously noted, Fundus Image pre-processing can be defined as either the extraction, enhancing or removal

of features within the image which can include pixel brightness and geometric transformations [149]. These are needed to correct image brightness, correct for image non-uniformity, and reduce noise or image artefacts such that image clarity is restored. Some conventional feature-based techniques for pre-processing include Grayscale Conversion, Contrast Limiting Adaptive Histogram Equalization (CLAHE) [149], and image filtering [3][2]. It should be noted that RetCam images are in colour. Colour images can be represented in different colour spaces [22]. RetCam colour image contains Red, Green, Blue (RGB) [22] channels with the green channel containing the most data. These are denoted by **R**, **G**, **B** symbols with each symbol denoted in an 8 bit number ranging from 0-255. Together, they can combine and form 16,777,216 possible colours [22].

Grayscale Conversion: This is performed by taking a 2 dimensional colour fundus image and converting it into a grayscale. A grayscale image is denoted as a data matrix. Each element in the matrix has a value that is denoted by a shade of gray. This technique [37] includes histogram equalisation for contrast and provides an even brightness level distribution for the image. The order of application is firstly to the green channel which contains most of the information and then red and blue channels respectively in a colour fundus image in RBG colour space. The results are then combined to give a gray colour. It is performed to void any unique colour uniqueness that causes the model to pick additional characteristics resulting from its intensity information. It is implemented using the standard formula:

$$\mathbf{I}o = ((0.3\times\mathbf{R}) + (0.59\times\mathbf{G}) + (0.11\times\mathbf{B})) \tag{3.1}$$

Many variations extract only the green channel from a colour image and then apply grayscale.

Adaptive Histogram Equalization (AHE): Adaptive histogram equalization [11] is used to improve contrasts [85] in images and applied to grayscale converted fundus images. It allows poorly illuminated areas to become brighter while those which are brighter either to remain or reduce the brightness, thereby levelling off the differences. AHE divides the image into sections and computes their histograms which is then used to distribute the lightness value. This enhances the image details and local contrast.

Contrast Limiting AHE (CLAHE): Contrast Limited Adaptive Histogram Equalization (CLAHE)[149, 65] is an improvement over Histogram Equalization (HE) [11][122] and Adaptive Histogram Equalization (AHE) [92]. It computes several histograms for different sections of the image, and subsequently distributes the lightness values, but it caps the histogram to a predefined value to prevent over amplification that occurs with AHE. This improvement results in a low noise sharpened image which can assist in numerous medical diagnoses [10][110][45][32].

Wavelet Transformation: Wavelet transformation [93] is used to represent the signal by using wavelet functions that are easy to implement. The discrete wavelet is one of the available wavelet transformation. In this method, wavelets are discretely sampled and by using them as a filter-bank, they can be employed for signal compression [120]. The Haar wavelet [75], is a simple wavelet transform which is easy to implement and widely used. Its use allows significant reduction in image size without any information loss. This is useful as fundus images are getting larger in size and resolution with better imaging technology. The continuous wavelet transformation (CWT) [107] is another

branch of transformation similar to Fourier transform which provides a way of decomposing a signal into its constituent wavelets and analysing each wavelet at different scales. CWT based series of methods have been used across medical imaging to denoise diagnostic images [23]. Libraries with many types of wavelets are available that are useful for various purposes. Both compression and denoising are useful attributes which can be used for fundus images.

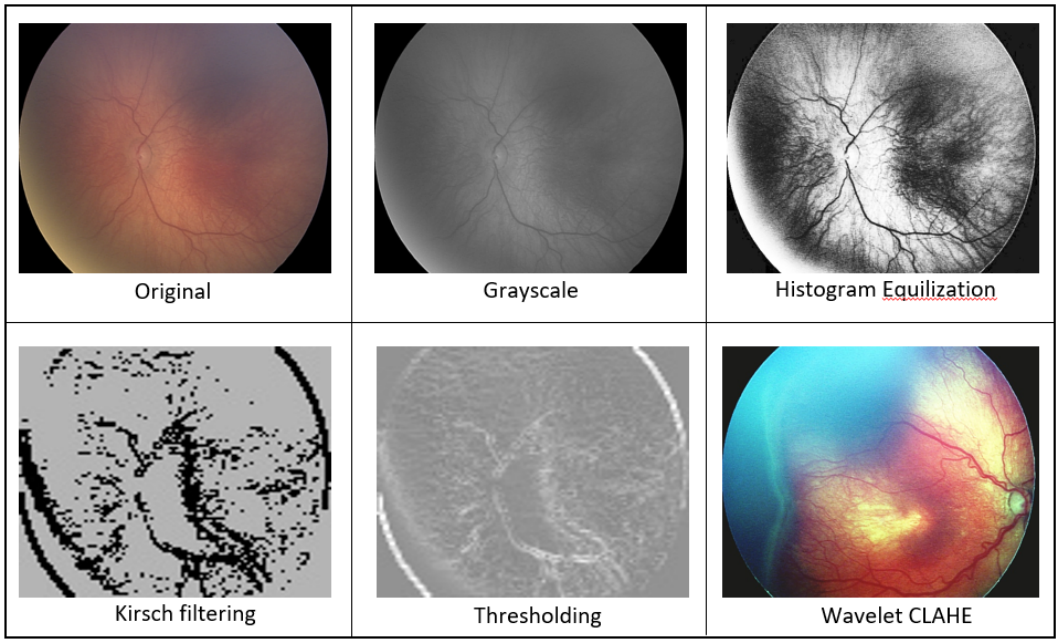
Filtering: This is another image enhancement method that improves the overall image quality or can assist in the recovery of image loss due to noise. It tries to maximise signal to noise in the presence of random background noise. There are three types of filtering, namely Median, Guassian and Kirsch. Gaussian is used widely and provides a good response to noise however in terms of segmentation. Kirch [51] method, which is an edge based segmentation method using a compass template with adjustable threshold values, provides excellent results for detection and identification of blood vessels and was the preferred filtering method for this thesis originally but was replaced by a better method using U-Net based segmentation.

Fuzzy C-Means Segmentation: This is a method of clustering which associates one portion of pixels with one or more sets pixel data portions in a retinal fundus image. The method aim is to detect blood vessels by segmenting the input fundus images. As noted, this was of interest for detecting ROP as well as DR but was not pursued in this thesis as U-Net CNN based segmentation methods provided a more accurate state of the art method.

Using combinations of the above primary methods, the resulting pre-processed images represent a standardised input to subsequent steps such as a CNN. An example of resulting outputs can be seen when applied to sample ROP RetCam images in Figure 3.1. We derived additional techniques for ROP Retcam, using

green channel extraction. In the literature survey, most prominent techniques in this class were CLAHE, grayscale using three channels only or grayscale via green channel only due to their simplicity, versatility and effectiveness. These have been used in many variations within ophthalmology settings for both, DR and ROP. Of note is CLAHE which is very versatile having implementations across different colour spaces as described by Ma et al., [65] and includes LAB color space. Overall, while image domain methods have demonstrated strength in simplicity with ROP RetCam images, the core challenge of restoration of ROP fundus images remains lack of clarity from haziness caused by internal reflection.

Figure 3.1: Image Pre-processing using ROP RetCam Image



3.1.2 Restoration methods

When light enters the retina, it undergoes refraction prior to hitting the retina. It, then, reflects from each layer it hits, including the sclera, lens, vitreous and retina itself. This reflection contributes a significant amount of noise. The formation of haze in the image is observed by the capturing fundus camera. The model for describing formation of haze is given in [148] as:

$$\mathbf{I}(x) = \mathbf{J}(x)t(x) + \mathbf{A}(1 - t(x)) \tag{3.2}$$

where **I** is observed intensity, **J** is scene radiance, **A** is global atmospheric light, t is the medium transmission describing the portion of light reaching the camera unscattered, and x is the pixel. The first term is known as direct attenuation and the second term is airlight [42]. We attempted, in our study, to leverage atmosphere based dehazing methods using Dark Prior Channel (DCP) as described first by He et al., [42] for each fundus image. DCP is based on the observation that in haze-free outdoor images, one of the channels will have a patch with the **A** variable having a low value in Equation (3.2). Haze removal is performed by estimating the transmission, refining the transmission by soft matting, final scene radiance recovery, and finally estimating the atmospheric light as the highest intensity in the input image. While good for terrestrial use, this did not yield much success for ROP RetCam images. Similarly, other powerful dehazing techniques as proposed by Zhu et al., [148], and Sami et al., [108] were also implemented without success in improving ROP RetCam images in terms of restoring details. The primary cause is these methods do not account for the fundamental challenge of reflection from within the eye.

For a hazy RGB image, neither channels will have non zero pixel values. Using amplification theory based on DCP theory, Gaudio et al., [34] proposed Pixel Color Amplification (PCA) which enhances a given fundus image. DCP theory permits inversions which can be used to derive additional Priors, namely Inverted DCP (Illumination Correction), and Bright Channel Prior (Exposure Correction). A fourth Prior based on the inter-relationship of three (3) priors is derived by Gaudio et al. [34]. Given Image **I**, transmission map **t** and atmosphere **A**, these four (4) Priors are unified with each revealing a weak and strong amplification including those of dark and bright pixel neighbourhoods. This yields 4 brightening/darkening methods which are referred to by letters A-D to brighten and W-Z to darken. These methods can be used individually or in combination to yield a merged image. Further a sharpening method is also available which can be activated by prefixing each letter with **s**. This allows sharpening of retinal features post prior computations which amplifies the difference between image and blurry computed version of itself. Using a combination of 4 Priors, PCA showed good retinal enhancement for EYEPACS/IRiD. It also held promise with our modification for ROP RetCam images that allows auto-balancing of illumination which is unpredictable. We noted that PCA suffered from being over-illuminated when used with adult fundus pictures as noted by Gaudio et al., [34]. The same problem was also noted when used with our ROP RetCam images dataset.

Zhang et al., [145] tackled the problem of reflection specific to a retina by building a multi-layer model of image formation that specifically dealt with reflective/illuminative imaging. The transmission term which He, Gaudio and others set to a constant value, is applied to illuminating light but also to reflected light. The first task is to estimate an enhanced restoration image

value using illumination, transmission of lens and a scatter matrix. The second component restores the image by focusing on the retinal area ignoring the black exterior box. Coarse illumination correction is performed across red, green, and blue channels followed by fine illumination boosting where the Grayscale DCP is used for dehazing. The last step is scatter suppression which results in an illumination corrected and dehazed image. The method is called Double Pass Fundus Reflection (DPFR). This method is very significant for ROP RetCam image pre-processing. It aligned with our own analysis of ROP RetCam images which notes that reflection from within the eye is the primary cause of quality degradation.

As the problem of ROP RetCam images quality is a multi-faceted challenge, we solved this by employing both image domain and restoration methods. First, we solved the problem of over/under amplification inherent in PCA while fine tuning DPFR specific for ROP RetCam images. These changes provide ROP RetCam pre-processed images with improved ROP clinical features. These pre-procesed images are more suitable for deep learning classifiers for ROP. We then combined our improved restoration methods and classic image domain method namely CLAHE to create a hybrid that further improved the overall ROP clinical features within the ROP RetCam image. Using each pre-processing method, a training and validation pre-processed dataset was created. Each pair was used with two different classifiers specific to Plus, Stages and Zones for ROP and their classification results were documented.

3.2 Image Segmentation

Image segmentation implies decomposition of an image into its constituent regions on the object such as optic disc, blood vessels, or lesions. A segmentation method ideally locates those sets which correspond to a distinct anatomical or pathological area of interest in the image. Blood vessels are a key structure within the retina and show up as a high contrast value on the fundus image. A fundus image has a wide range of blood vessels such as capillaries, various sized vessels, and leaking areas. Accurate segmentation is an essential part of detecting a pathology such as diabetes and ROP. Segmentation is the leading trend within computer assisted diagnosis.

There are multiple segmentation algorithms available that are combinations of image processing and classification methods that yield good retinal segmentation output with advantages and disadvantages. These retinal vessel segmentation methods are classified into 6 main categories. These are mathematical morphology, multi-scale approaches, pattern recognition, vessel tracking, matched filter. and model based approaches.

Pattern Recognition:

These computational methods use pattern recognition for automatically detecting and classifying retinal blood vessel, retinal features including artifacts and background. Pattern recognition methods for vessel segmentation are divided into two categories namely supervised and unsupervised methods. Supervised algorithms use prior labelling data for determining if a pixel is associated to a vessel or not. Unsupervised algorithms determine the vessel segmentation without any prior labelling knowledge.

In supervised algorithms, the rules to extract vessels are learned by the algorithms on the basis of a training. The training dataset is usually manually processed. The segmented reference images are termed as the ground truth or gold standard. The vascular structure in these ground truth images are precisely annotated by a specialist such as an ophthalmologist. For unsupervised classification, the algorithms attempt to find plausible patterns for blood vessels present in the retinal fundus images. These can be used to establish if a particular pixel is affiliated to a blood vessel or not. In this approach, hand labelled ground truths are not used in the overall design.

Matched Filtering: This technique convolves a two dimensional kernel with the retinal image [9]. The kernel models characteristics in any given position and arrangement. The presence of a feature is detected using matched filter response (MFR). 3 vascular properties are leveraged in order to compose the filter kernel. Firstly, blood vessels have a limited contours and are approximated using linear segments. Secondly the width of blood vessels decreases as they branch outwards away from the optic disk. Lastly, the vessel segments cross-sectional pixel intensity signature approximates a Gaussian curve. There is significant performance overhead as well as noise in an image that can result in a false positive but nevertheless, this is a very effective technique when used with other processing techniques [121].

Vessel Tracking: Vessel tracking algorithms [124][78] are able to segment a vessel between two points using localised data. This allows them to focus on a single vessel instead of the entire fundus vasculature. During tracking, properties such as average width, tortuosity, and gray level intensity are used to determine the center of the linear width of a blood vessel. The center line of the vessel aided by localised pixel data is used to track the vessel profile

that fits the profile of a vessel model. These methods are very effective in providing highly accurate individual vessel information such as vessel widths. This information is unavailable in other algorithms. This technique follows the entire tree which is more efficient rather than traversing the entire fundus image where the majority of space does not have vessels. These methods are generally paired with matched filtering for higher accuracy.

Mathematical morphology/Tracing: The word morphology is generally used in biology to define structural properties in a biological organism. The principle of mathematical morphology forms the basis for a set of tools that are used to extract features, outlines, and shapes that describe biological structures. The language of mathematical morphology [115] uses set theory that permits a unified approach towards many image processing problems. The unified approach offers comprehensive algorithms to solve these image processing problems. These include morphological operators which employ structuring elements to a given image. The images used can either be binary or grayscale. Morphological operators include dilation, erosion, and closing which can be used with each other [29]. Dilation performs an expansion operation using defining a structuring element, connecting sparse areas after filling holes [29]. Any gaps such as holes that are filled with disjointed regions are connected [29]. Erosion performs the opposite by using the structuring element to shrink objects [29]. Closing is defined as a dilation operation followed by an erosion operation [29]. Opening is an erosion operation following by a dilation operation [29]. Two other algorithms, namely top hat and watershed transformations are used in medical image segmentation [115]. These algorithms result in enhancements for blood vessels, if present in a retinal fundus image.

Multiscale Tracking/Tracing: The width of a vessel decreases in a gradual manner as it traverses away from the optic disk. Therefore, a vessel is defined *as a contrasted pattern with a Gaussian-like shape cross-section profile, piece-wise connected, and locally linear, with a gradually decreasing vessel width* [29]. The principal idea for this representation is for separating information related to the blood vessels that have differing widths across different scales as demonstrated by Farnel et al., [24].

Model Based: These approaches use explicit vessel models to extract the blood vessels from a retinal fundus image. Two categories are vessel profile and deformable models. The blood vessel cross sectional intensity profiles is used to approximate a Gaussian shape. The profiles which can also be used include second-order derivative Gaussian and Hermite polynomial profiles. The models include addition of the complex structures such as background, lesions, varying vessel widths, and other artifacts. These assist improving segmentation accuracy instead of assuming a simple flat background to segment against.

Active Learning: Active learning uses artificial intelligence (AI) either in the form of deep learning or machine learning. The basic learning block of AI lies with provided training data that is labelled in an orderly manner such that training is correctly performed. This phase is known as the learning or training phase. In case of unlabelled data, regression or clustering is performed to identify patterns. In this case the pattern is the retinal vessels for which a ground truth is drawn manually. It is used to train a model to identify blood vessels and predict a vessel map. This can be achieved either by a a Bayesian Convolution Neural Network or any other Convolution Neural Network (CNN) architecture. In our research, we focused on the Convolution Neural Network

namely U-Net [104] for our segmentation requirements, namely its variant in the state of the art LWNET [31]. These are discussed in subsequent sections.

3.3 Deep Learning

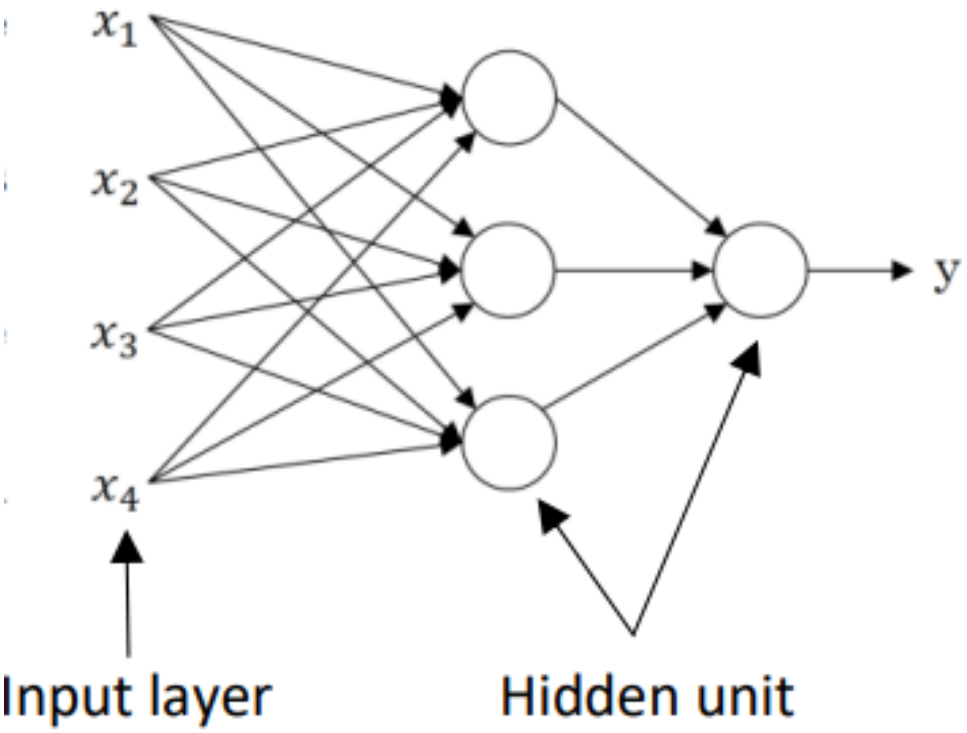
CNNs are being increasing employed across a wide range of applications such as computer vision, medical diagnosis, language recognition, voice recognition [4]. Recently CNN based image processing methods have demonstrated significant improvement in the examination of colour fundus retinal images for the presence of pathology or for the purposes of segmentation [4]. CNNs are able to recognise and analyse retinal vessels with a high degree of accuracy and performance that is close to a human expert [4]. In the case of fundus based CNNs, the success of these techniques are dependent on a critical component, namely the availability of manually labelled ground truth datasets. A high level of learning can be described in Figure 3.2 that shows how in each subsequent box learning is taking place.

Figure 3.2: Deep Learning

A basic neural network tries to mimic the structure of a working brain and learns progressively. There are input and output layers with more than one hidden layer where learning takes place as highlighted in Figure 3.3. A neural network is either classified as a shallow or deep neural network. A shallow

network has one output and one hidden layer while the later has many hidden layers [109].

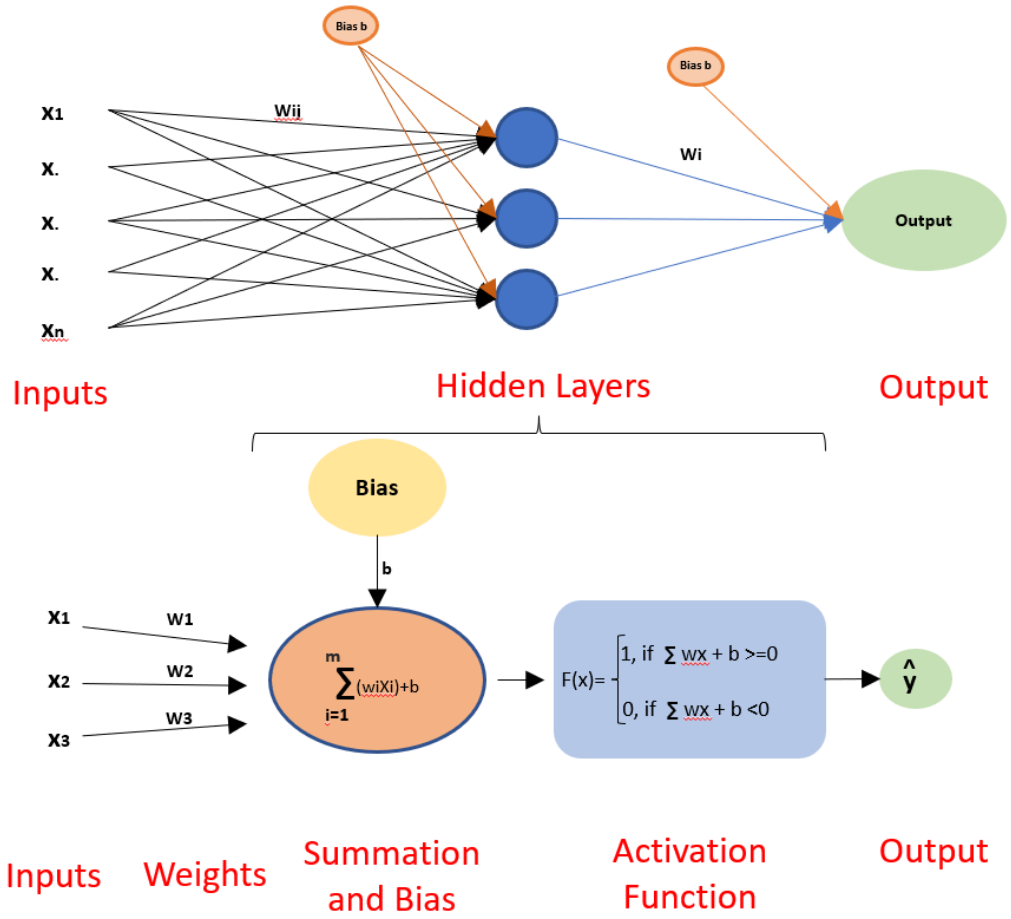
Figure 3.3: Basic Neural Network 1 [79]



The neural network consists of neurons. These two components attached to them in the form of weights (**w**) and biases (**b**). They take inputs from features or predecessor neurons and calculate a linear activation function (z) and an output function (a). The network component is described in Figure 3.4. It is comprised of interconnected neurons. The responsibility of each connection is to transfer the output of a neuron to the input of a connected neuron. There exists a propagation function which is responsible for computing the input to a neuron from the outputs of predecessor neurons using equation 3.3 [79] where

w is weight and b is bias. As the value of Z must be between between 0 and 1, an activation function [4] is employed for transforming the value of Z to a normalised scale as noted in Equation 3.4 [79]. Weight w's purpose is to reduce the output error that is defined as the difference between the desired output and the actual output. Activation functions such as Relu [4] are used. Lastly, there is an algorithm also known as the learning rule which modifies the parameters of the neural network, namely the weights (**w**) and biases (**b**) for an optimal output.

Figure 3.4: Components of a neural network [53, 140, 69]



$$Z = w.X + b. \tag{3.3}$$

$$h(x) = \begin{cases} 1, & \text{if } Z = w.X + b > 0 \\ 0, & \text{otherwise} \end{cases} \tag{3.4}$$

3.3.1 Convolutional Neural Network

Convolutional Neural Networks (Figure 3.5) are described as "*specialised neural networks designed with the aim of processing data with a gridlike structure*" [139]. As its architecture is being based on parameter sharing and sparse interactions, this allows for effective and efficient computation of spatial invariances in images [38]. Images are considered as matrices or sets of matrices as inputs. CNNs, like neural networks, are made up of neurons with learnable weights and biases. Each neuron receives several inputs, takes a weighted sum over them, passes it through an activation function and responds with an output. The whole network has a loss function leveraging various techniques used in standard neural networks being applied. A CNN is comprised of three layers which are the Convolutional, Pooling and Fully Connected layers [69, 38, 140, 4].

Convolutional layers consist of filters and feature maps [140, 38, 4]. Filters, also known as kernels are applied to an input and may have weights resulting in output values. Input size, that is also known as a receptive field is a fixed matrix. The convolutional layer may have an input layer which will be pixel values in case when the input is an image. The application of a filter on the input results in a feature map though activations. As one progresses

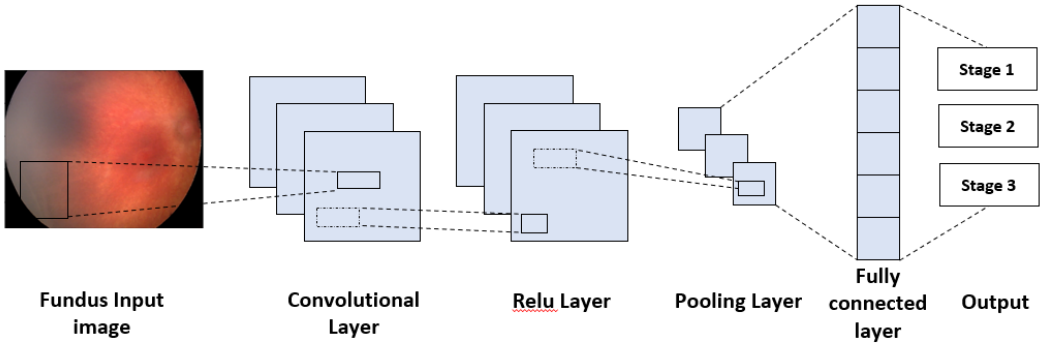
forward into the network, namely deeper, each subsequent convolutional layer will take input from a feature map from the predecessor convolutional layer. In other words, a feature map is defined as the output of one filter applied to the predecessor layer. Each layer performs feature extraction starting from the incoming matrix then to complex features in successive layers. It must be noted, image sizing must remain similar across training, and validation, otherwise the CNN will produce poor results.

Pooling layers summarise and expressed in the prior layers feature maps [4, 140]. The pooling layer introduces a technique for compressing or generalizing feature representations. Pooling layer is smaller than the prior convolutional layer by design as it is a downsizing operation. This results in reduction of learnable parameters [139]. They are often very simple and operate either by taking the average or the maximum of the input value in order to create its own feature map. These are referred to as average or max pooling. In the pooling layers, parameters are not used except for hyperparameters that control aspects such as filter size [139].

The output feature maps of the final convolution or pooling layer is transformed into a one dimensional array of numbers. These are then connected to either one of more fully connected layers which are referred to as dense layers. In the dense layers, each input is linked to every output with a learnable weight. With features extracted and downsampled completed, they are mapped using a subset of fully connected layers to final output of the CNN. These can be form of probabilities per class for a classification task. The final fully connected layers generally consists of output nodes which are same numerically as the number of classes. Each fully connected layer is proceeded with a non linear function such as ReLU as noted before.

Fully connected (FC) layer represent the output feature map which is represented as a flat single dimensional array [139, 79]. As the name suggests, each neuron in FC layer is connected to each neuron in prior layer. It is also referred to a dense layer. Typically it is the last layer in a given CNN which takes as inputs, the outputs resulting from downsampling from prior layers such as convolutional and pooling layers. An FC layer may have an activation function to produce final output class predictions [139, 79].

Figure 3.5: Convolutional Neural Network



Large neural networks have several challenges when we started this work. The higher the number of layers, the more abstractions that take place on top of prior layers. This causes backward propagation to be less effective in relaying back the information to lower layers. Consequently, the gradients start to become inconsequential relative to the weights of the networks. The second challenge was to avoid overfitting. Overfitting occurs when the model is unable to generalise and fits too close to training data [81, 4]. Underfitting can also occur due to high level of bias. Overfitting was a challenge we faced

due to limited data. Strategies to overcome this included augmentation, regularisation, and early stopping. These were used in our work. In case of larger dataset availability, resampling and ensemble would be the preferred way.

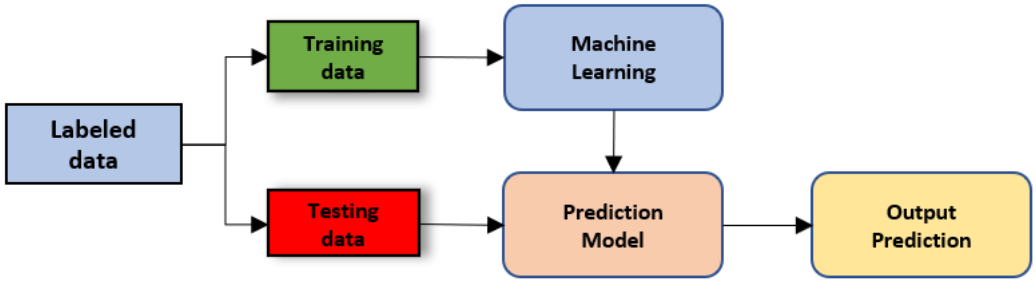
3.3.2 Machine Learning and CNNs

Machine learning is described as a set of powerful learning methods in neural networks [79]. These are the primary focus of research as they are used to solve challenging problems in the field of vision (image recognition), segmentation and various other fields. These algorithms treat each instance of a dataset as a collection of features[14]. Machine learning is categorised into supervised, semi-supervised and unsupervised learnings [79, 4, 69]. The choice of categorisation is based on data. If labelled data is present, this will be for supervised learning. Semi-supervised has a mix of unlabelled and labelled data. Unsupervised uses non labelled data and makes use of key techniques such as clustering, dimensionality reduction, and anomaly detection [14].

3.3.2.1 Supervised Learning

Supervised learning is very powerful and therefore widely used because it utilises labelled data for training [21, 79, 4, 69]. Using labelled data enables training a model such that it can accommodate newer data with a high degree of confidence. The downside of this method is obtaining labelled data. Preparing labelled data is also very laborious. Supervised learning architecture model can be shown in Figure 3.6.

Figure 3.6: Supervised Learning

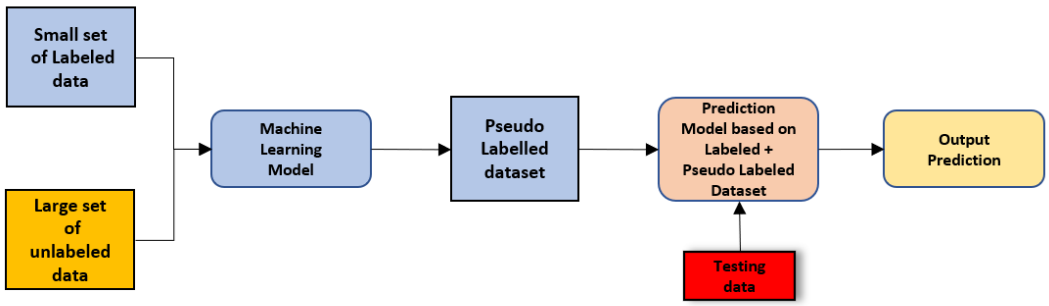


Supervised learning architecture starts with data collection which is then partitioned into training and testing datasets [21, 79, 4, 69]. These datasets can be pre-processed further for feature extraction. These feature extracted labelled datasets are used for training the model to recognise these features for relevant labels. Testing is lastly carried out with the model trying to predict vs the ground truth label [21]. It must be noted that there are 2 broad forms of supervised learning namely classification and regression. Classification is defined as predicting a set of output values based on a set of known values. During the training phase, the classification algorithm also referred to as a classifier, learns from training labelled dataset and assigned a data point to a particular class. At the end of classification, the model is considered complete and available for prediction. When predicting, this model is used to predict the class label for the test data. Regression method aims towards finding co-relations between a set of input variables and prediction of continuous output values. It is the opposite of classification which focuses on discrete values [21].

3.3.2.2 Semi-supervised Learning

Semi-supervised [87, 79, 4, 69] as the name states, is between supervised and unsupervised. Some labelled data is made available together with unlabelled data. In a majority of real world cases, this is the case. Labelled data is limited and expensive to collect while unlabelled data is readily available. As the aim is to have good matches, additional information is extracted from the unlabelled data [87]. First a model is trained in labelled data, then unlabelled data is self-labelled in an iteration and added back into the labelled training data as illustrated in Figure 3.7.

Figure 3.7: Semi-Supervised Learning



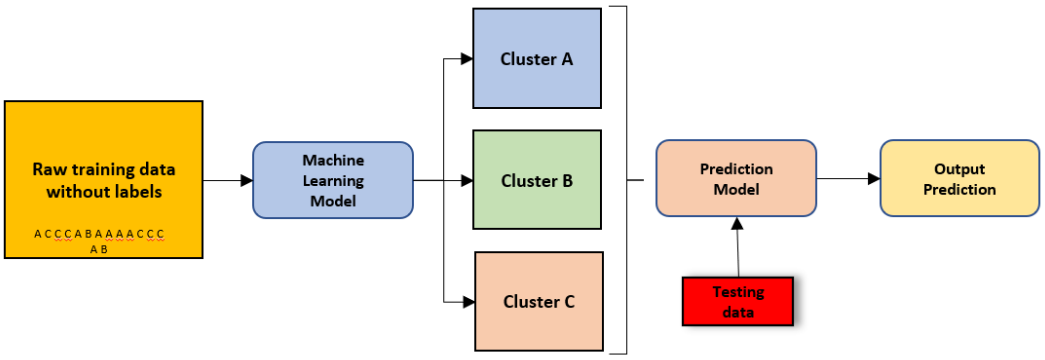
There are 4 broad method categories namely consistency training, proxy-label, generative models and graph based models [88]. Consistency training states a model is trained for consistent prediction such that if there are sufficient fluctuations applied to the unlabeled data points so that a given prediction does not change significantly [87]. For given similar inputs, Proxy-label uses a trained model to generate additional labelled data from unlabelled data based on rules. Generative models are able to transfer learned features from

one task to other tasks further downstream. Graph based models passes labels from labelled nodes to unlabelled nodes using strength of the edge between the nodes to signify similarity between them.

3.3.2.3 Unsupervised Learning

Unsupervised learning is a machine learning technique where no supervision is required for the model. The model is allowed to learn independent of labels so that it can identify cohorts of patterns and data which were scattered and disjointed. No labelled form of data is used [14]. It will attempt to group based on patterns by trying to discover hidden and novel patterns from the unlabeled data[71] as shown in Figure 3.8.

Figure 3.8: Unsupervised Learning



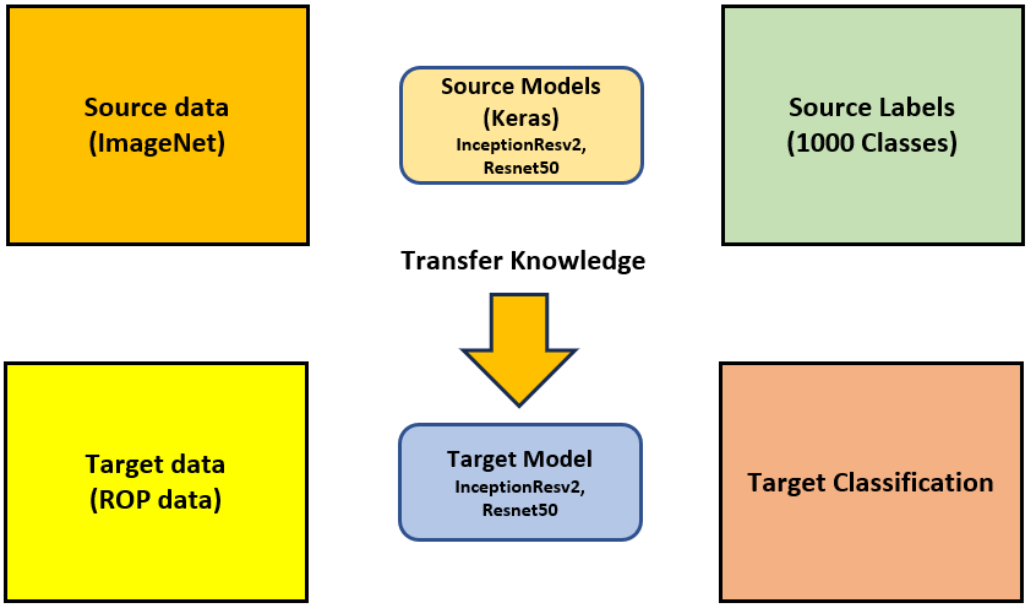
Unsupervised learning can be done in four ways namely by clustering, principal component analysis, anomaly detection and auto-encoders [71]. Clustering is most used in this problem area as it tries to find a cluster or groups where the properties of objects are similar in comparison to objects from other clusters [71, 14]. Principal component analysis uses association rule that describe the relationship in the form of rules. It is also used to identify new relationships

between non related objects in the dataset [71, 14]. An opposite approach to first two is anomaly detection which aims at the outliers of a dataset that looks for deviations in patterns. Lastly auto-encoders are feed forward networks where the input is the same as output. The input is decoded into a code that is of lower dimension and re-constructed into output which are typically found in compression methods [6].

3.3.3 Pre-trained Models and Transfer Learning

Transfer learning is a related problem to Semi-supervised classification where the concept of leveraging an existing pre-trained CNN model for one classification as a basis for a CNN model focused on a different classification task by using information from larger datasets using low computational cost [81]. In other words, prior training is leveraged instead of training from beginning. ImageNet dataset [16] has over 14 million images across 1000 classes. This dataset is considered as a gold standard and its use has led to complex CNNs being created. Most pre-trained models use ImageNet [16]. These models serve an extremely useful purpose especially when there is insufficient data to train a new model. This saves considerable time but allows for good generalisation that results in improved classifier results. Furthermore, the architecture allows additional layers to be added to the pre-trained CNN model which are allowed to be trained for Additionally, fine tuning can be employed by replacing or adding FC layers which then be retrained on the new training dataset describing new features [139]. Transfer learning process can be described in Figure 3.9.

Figure 3.9: Transfer Learning



Imagenet[16] has used as the benchmark dataset for several years for evaluating the performance of CNNs. It can also, easily be used to improve performance, especially training time on newer classifier data. This is done by adding layers to the network that can only improve its performance, not worsen it. Another example architecture is the DenseNet [54], which is a variation of the ResNet [58] model, that also utilises skipped connections. However, instead of connecting one layer to a previous layer, as in ResNet, the DenseNet connects all layers with all previous ones such that the input for a given layer is the feature maps of all earlier layers. The given layer's output is fed as input to all subsequent layers. Due to these dense connections, this model does not require as many parameters as the ResNet counterpart.

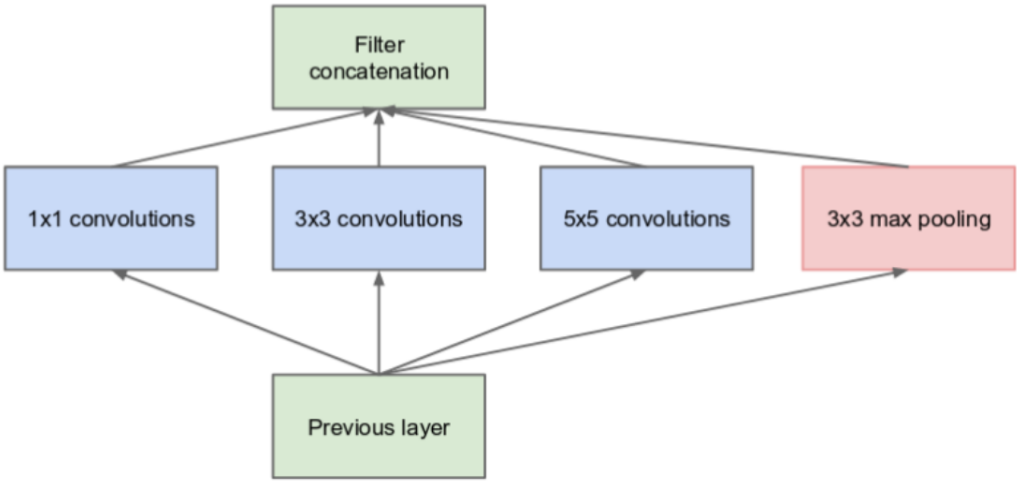
Numerous architectures of CNNs presently available, played a key role in building algorithms which are powering AI as a whole in the foreseeable future.

We identified these from annual ILSVRC [105] (ImageNet Large Scale Visual Recognition Competition) competition where best algorithms for computer vision problem areas are selected. Some of them include AlexNet, VGGNet (Oxford) [26] , GoogLeNet [118], ResNet [58] amongst many. These have been primarily used against the ImageNet deep learning challenge which serves a gold standard. This thesis considered these and eventually focused on VGG16 [26], ResNet50 [58], Inceptionv3 [56] and InceptionResv2 [56] as they are available with transfer learning. We briefly discuss GoogLeNet [118], ResNet [58] and Inceptionv2 [56], Visual Geometry Group (VGGNet)[26] and more newer state of the art Transformers and Foundation models.

3.3.4 Inception GoogLeNet

Conventional CNNs tend to stack convolution layers depth wise with the aim of getting better performance. Specific to imaging, there is a tremendous variation in the location of the information, and optimization of the kernel size becomes a challenge. A large kernel is necessary for widely distributed information and small kernel for localized information. These result in overfitting and performance degradation. Hence, Inception network use multiple sizes at the same level which appear to make it 'wider'.

Figure 3.10: Basic Inception model [118]

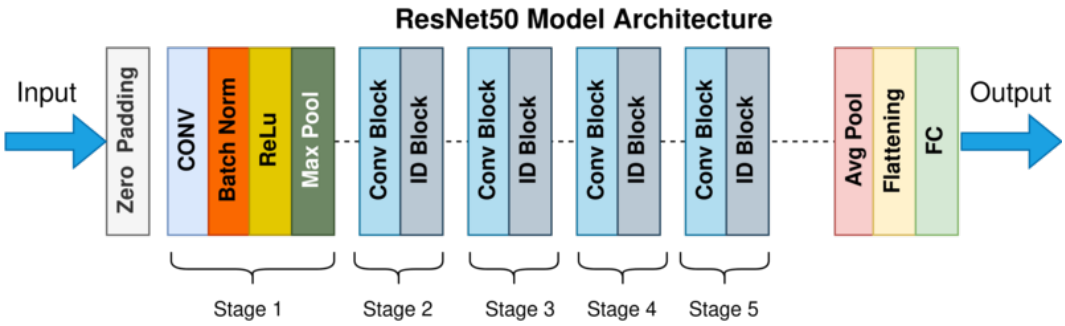


The above diagram in Figure 3.10, illustrates a basic inception model. Convolution is performed on an input with 3 distinctly different filters in addition to max pooling. The outputs are concatenated before being passed to the next inception module. There are various models of Inception [56] such as Inception v1/GoogLeNet, v2, v3 and most current being v4. Each iteration improved the previous model in terms of layering and performance.

3.3.5 The Residual Network CNN

The Residual Network also known as ResNet50 [58] was developed by Microsoft and is amongst the most widely recognised and utilised architectures in deep learning. It has been used in Kaggle competitions focusing on the numerous classification of pathologies related to DR. We used this as a primary method as it has been trained in the ImageNet validation set for transfer learning.

Figure 3.11: Resnet50 model [58]

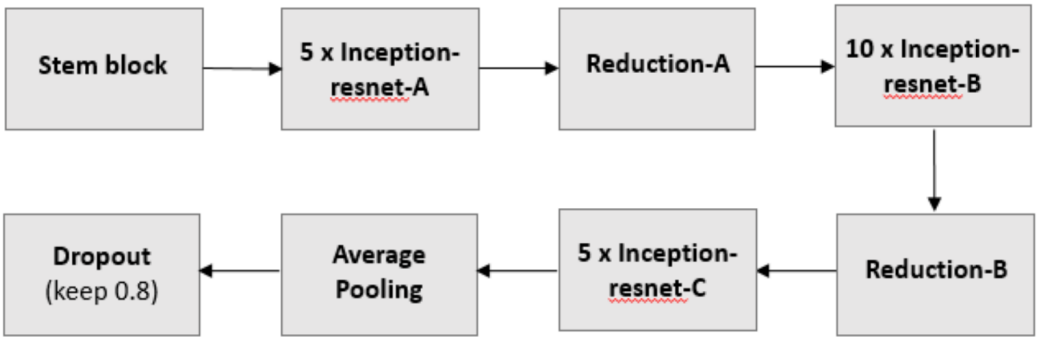


ResNet50, as the name implies, is comprised of 50 layer CNN that includes 48 convolutional, one MaxPool, and one average pool layer. These are composed of 5 stages as illustrated in Figure 3.11 with each stage having a convolutional and identity block. Its key strength lies in the use of a skip connection, which allows a block to be skipped if there is a problem in the layer.

3.3.6 InceptionResv2 CNN

Inception-ResNetv2 [56] is a CNN architecture based on the Inception series. This was the second best choice as it was easy to train and also had been trained with the ImageNet validation set for transfer learning. The model is illustrated in Figure 3.12.

Figure 3.12: InceptionResv2 model [56]

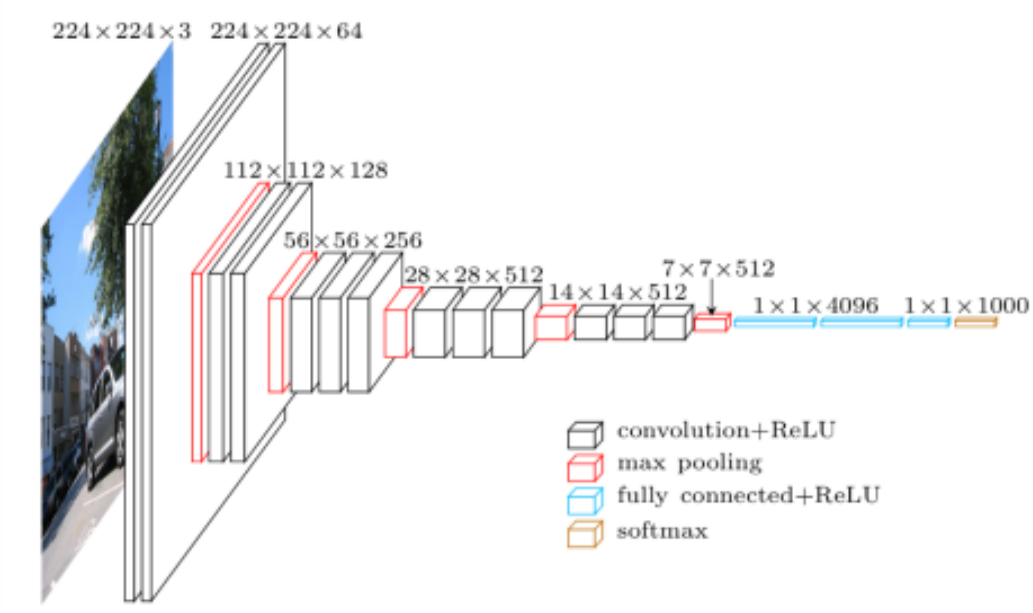


It combines Inception structure with residual connections by adding residual connections and by replacing filter concatenation stage of inception architecture. This greatly improves performance in training as degradation resulting from the deep CNN structure is eliminated.

3.3.7 Visual Geometry Group (VGG) CNN

The VGG network architecture was introduced in 2014 by the Variable Geometry Group [112]. It is elegant yet simple. It uses only 3x3 convolutional layers which are placed vertically on top of each other in an increasing depth. The reduction in volume size is the responsibility of max pooling. Two FC layers each having 4,096 nodes are then proceeded by a softmax classifier.

Figure 3.13: VGG16 Deep Neural Network [26]



The "16" and "19" represent the number of weight layers in the CNN network [112]. These weights are shown in Figure 3.14 as columns D and E. Hence VGG16 and VGG19 are two types of variants within VGG most commonly used. As with many CNN methods, this can be very complex and slow but provides very good results in imaging and was considered as a candidate classifier for this thesis.

Figure 3.14: VGG Configurations [26]

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

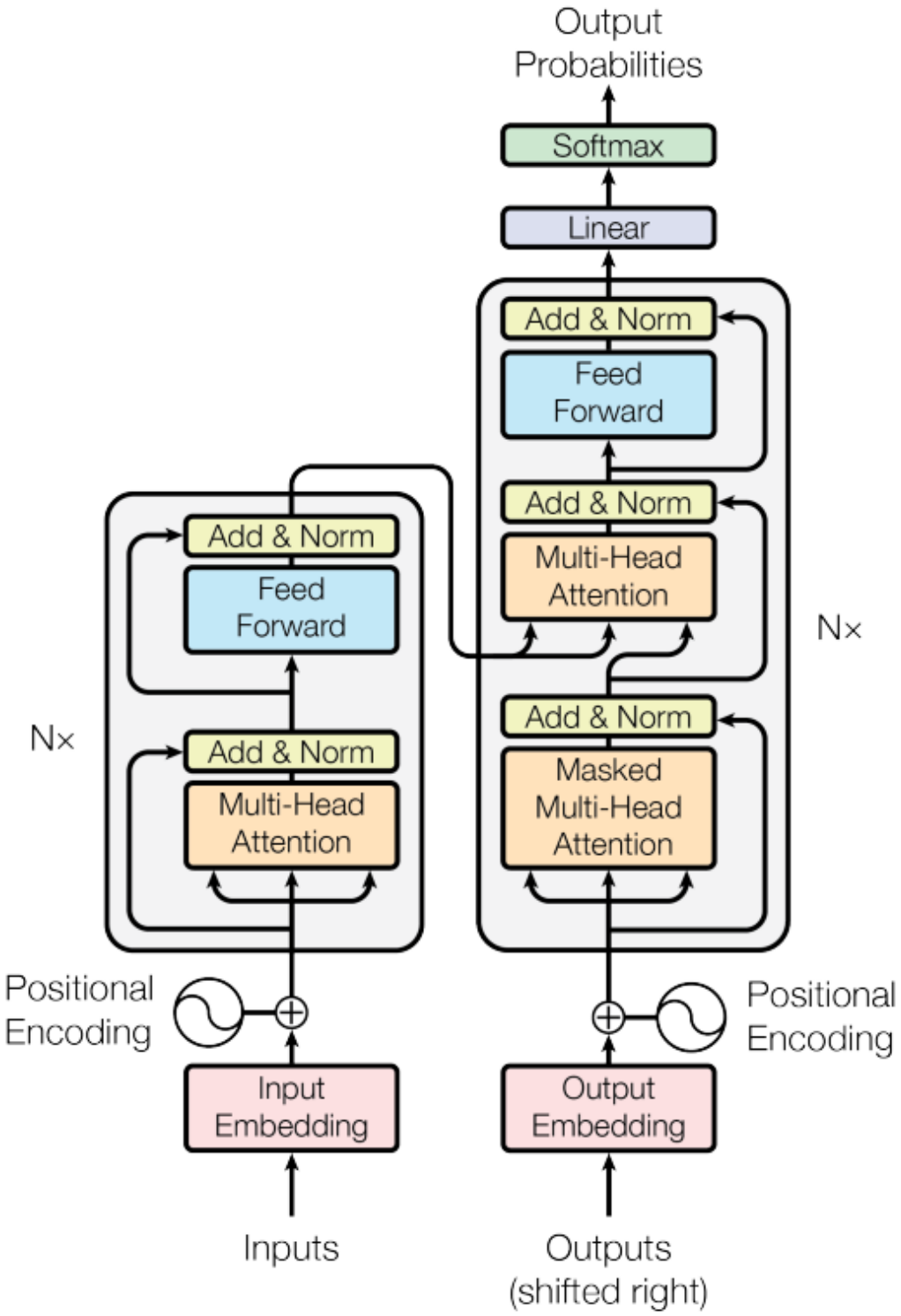
3.3.8 Vision Transformer (ViT)

Transformers are the latest state of the art innovative deep learning architecture which was first described in 2017 by Vaswani et al. [127]. Its key feature is self-attention capability which is absent in recurrent neural networks (RNNs) [62] which are unable to manage dependencies. It identifies long range

dependencies by analysing computed weighted sums of all tokens in a sequence as per their priority and position.

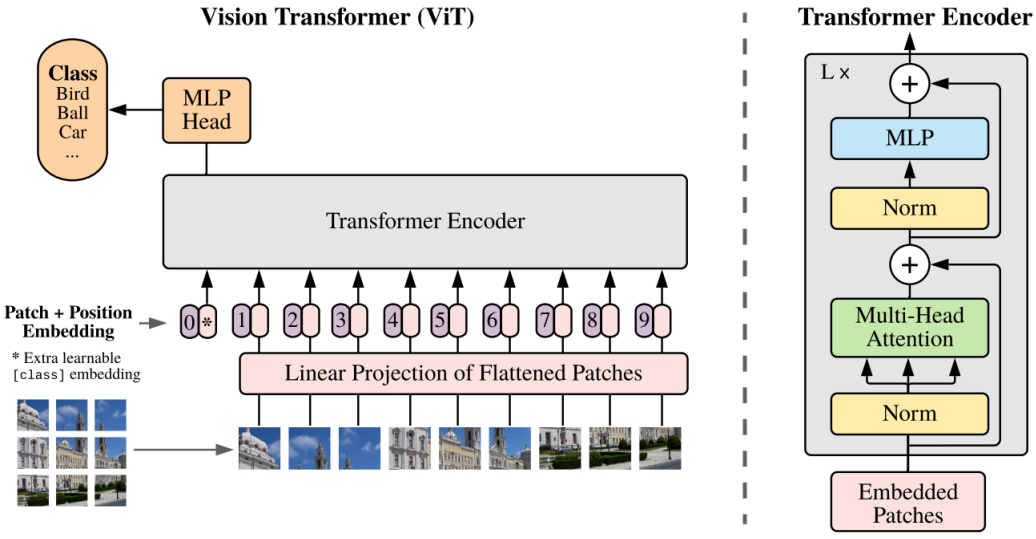
Transformers design includes encoder and decoder layers as described in Figure 3.15. These handle different input and output sequence lengths. Each encoder-decoder layer comprises of sub-layers which includes feed-forward networks and multi-head self-attention. Layer normalisation and residual connections are also used to improve training performance and provide stable models. These features have enabled transformers to improve data pattern identification for use in language processing and computer vision.

Figure 3.15: Transformer [127]



Dosovitskiy et al. [20] introduced Vision Transformers (ViTs). ViT architecture is illustrated on Fig 3.16. They modified Transformer architecture for use in image classification by using tokens representing image patches. Images are split into fixed-size patches. Each of these patches are linearly embedded with a positional embedding. A standard Transformer encoder is used to process the images. For precise classification of an image in this modified representation, a learnable 'classification token' is added to the sequence for classification tasks.

Figure 3.16: ViT Model [20]



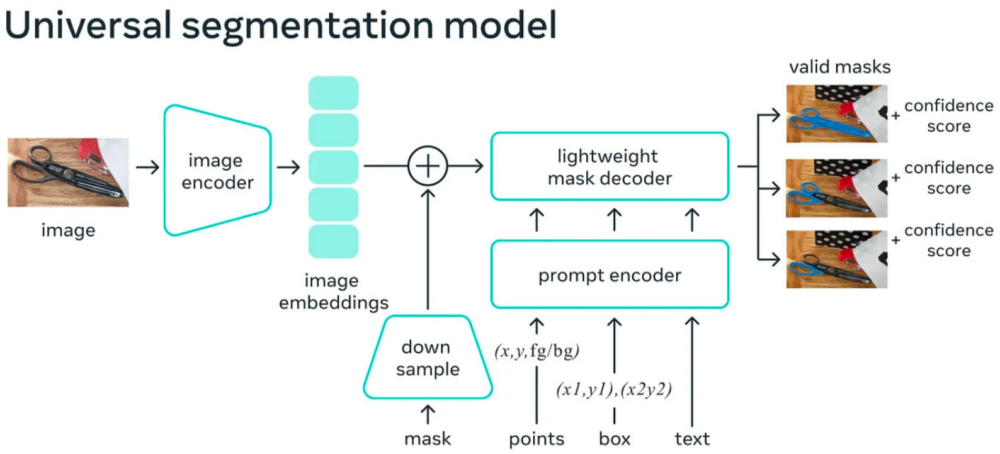
ViTs demonstrate key advantages over CNNs namely the ability to retain long range-dependencies and focus on key features while processing. However, unlike CNNs, ViT does not have the flexibility to add layers.

3.3.9 Foundation Models

Foundation models represent a new frontier in AI. The concept is based on having a universal generic model which can carry wide range of tasks. Such a model must be trained on very large data in order to successfully carry out a wide range of tasks. Foundational models in computer vision include Segment Anything Model (SAM) [66], CLIP [97] or DALL-E [100]. Ma et al, [66] note that in the medical images, use of SAM can be challenging as there is considerable difference between images used for training SAMs that are general in nature and medical images. Nevertheless, the ability to generalise is a significant step forward from deep learning models which are limited to their domain specific specified training and imaging requirements.

Meta's SAM [60] architecture, noted in Figure 3.17, comprises of 3 main components. The first component is an image encoder which uses MAE pre-trained ViT which is adapted to process high resolution inputs. Next component, is a prompt encoder which considers two types of prompts. The first type of prompts can be in the form of boxes or points or text. The other type of prompt is that of a mask and uses a fast off-shelf text encoder. The last component is a fast decoder that is based on a modified Transformer architecture which uses the forwarded image along with associated prompt embedding to generates masks.

Figure 3.17: Universal Segmentation Model [91, 60]



For any given image, SAM can generate masks for all objects present in the image. Its core strength is being able to interactively capture an object to segment by simply point/click with filtering or draw box or use a tool represents a significant jump in segmentation on demand. When dealing with uncertainty on the source object, it can generate multiple valid masks.

In terms of medical image tissue segmentation, SAM remain challenged as noted by Zhang et al. [144] survey paper. In particular, they note "*imbalance proportion of medical image data that exists among SAM's training dataset because the domain-specific dataset is extremely scarce in open-world object segmentation tasks*". Ma et al., [66] also note that "*the model exhibited substantial limitations in segmenting typical medical targets with weak boundaries or low contrast*". This is important as weak boundaries and low contrast are common problems within a given ROP RetCam image when identifying clinical ROP features.

Ma et al [66] noting limitations for medical images when using SAM proposed a fine tuned foundational SAM, referred to as MedSAM. The foundational SAM is fine tuned using 1.570 million medical image-mask pairs covering over 30 cancer types. The images were across different modes such as X-ray, CT, MRI, and endoscopy [66]. The segmentation results were then measured against SAM foundation model, specialised U-Net and DeepLabv3 models. MedSAM results were better with UNet coming second and SAM performing the worst.

3.4 Related Works

There have been numerous studies as well as papers published which used various techniques for the three compositional areas we identified earlier, namely image pre-processing, feature extraction and deep learning using various CNN models. Most studies have focused on DR with only a few contributions in these three segments for diagnosing ROP. The following reviews will cover specific papers where significant contributions have been made which have parallel relevance to ROP. Due to the rapid improvements in image pre-processing as well as various classifiers using CNNs, papers reviewed were restricted from 2010 until 2024.

3.4.1 Fundus Retina Image Pre-processing

Image pre-processing is a key component in highlighting or enhancing features normally not visible within a fundus retina image such as vascular features. In Section 3.1, we described various pre-processing techniques which are a must

for any computer assisted diagnosis. Below are some more prominent paper reviews.

Akram et al., [3] presented method that employs a Gabor wavelet which is combined with sharpening filter that which enhanced the presence of a vascular pattern. Using the sharpened retinal image, edge detection algorithm with additional morphological operation. When tested using, the DRIVE database, it gave 0.94 accuracy which was in line with results from Soares [113]. In subsequent paper, Akram et at., [3] refined the previous method further in a improved segmentation technique which removes noise by utilising hue and intensity channel values and identified dark background using local mean and variance. They referenced this [2] for the purpose of grading DR. With this technique, they are able to subsequently perform vessel segmentation as well as isolate exudates. Akram's techniques have shown to provide very good responses for microaneurysms, blood vessels, and exudates.

In the problem of diagnosing glaucoma, the problem is to focus on the optic head ignoring the rest of the vasculature in the retina. Several papers have described various pre-processing techniques with an example being Tjandrasa [123, 113]. In these papers, in order to compensate for the non uniform illumination, a homomorphic filtering method was utilised to reduce the problem of uneven illumination and intensity fluctuations in the image. Homomorphic filtering includes applying a Gaussian low pass filter to get an intermediate filtered image followed by application of dilated image which produces a uniformly illuminated grayscale image from which removal of vasculature and nerves can be performed.

Vernika et al., [134] note that image acquired by a fundus camera is in red and yellow shadings and is not sharp enough thereby requiring pre-processing

to remove unwanted artefacts and amplification of relevant data. They modified Otha's model which uses three orthogonal colour features: **I**1 (Equation 3.5), **I**2' (Equation 3.6), and **I**3' (Equation 3.7). The advantage of such transformation is its simplicity, compared to non-linear approaches and yields good blood vessel detection results.

$$\mathbf{I}1 = (\mathbf{R} + \mathbf{G} + \mathbf{B})/3, \mathbf{I}2' = (\mathbf{R} - \mathbf{B}) \tag{3.5}$$

$$\mathbf{I}2' = (\mathbf{R} - \mathbf{B}) \tag{3.6}$$

$$\mathbf{I}3' = (2\mathbf{G} - \mathbf{R} - \mathbf{B})/2 \tag{3.7}$$

Shilpa et al., [48] focused on the problem of identification of blood vessels. They used green channel from the colour fundus image to obtain the presence of blood vessels. A new algorithm method was designed that layered contrast enhancement, background exclusion and thresholding to eliminate background illumination variability allowing for easy foreground analysis. An additional morphological operation to smoothen the background allowed the vasculature to be very visible. This work highlights a simple and yet efficient algorithm which is applicable for retinal fundus vessel segmentation that worked well for both normal and abnormal images taken from the DRIVE database. Due to its efficiency, it is well suited for fast processing applications.

Wilfred et al., [28] were amongst the first to employ neural networks for retinal blood vessel segmentation. The first stage requires pre-processing by converting the colour fundus image into Hue Saturation Lightness (HSL) colour

space. Next, a mean filter is applied for the Luminance space background estimation. The image is split into background and foreground images are separated. Normalisation is carried out on the foreground split image. Then the image is converted back into RGB space. After the image enhancement, Gabor filter is applied at various orientations to extract smaller object features. These are merged resulting in improved accuracy of segmentation of varying vessel sizes. This is then passed along to the neural network which is discussed in the next subsection under segmentation.

Anila [5] used an interesting approach to exudate detection by using Local Binary Patterns (LBP). LBPs are a texture based operator which is able to cluster pixels within a given image and its surrounding area for localised variation. Different groups of patterns are labelled. Classification is then performed based on the label distributions. Pre-processing for LBP involves using Median filtering as a non linear operation which preserves the edges while excluding the blood vessels and the optic disc. This process is performed on the 3 channel components of a colour image (RGB). Their approach is able to identify a DR image from normal fundus image while avoiding any previous segmentation stage of retinal lesion.

With over 19 papers reviewed, the common practice for pre-processing is given preference with wide range of techniques depending on the pathology being sought after. This background information held a lot of insight into how to approach this issue for ROP fundus images.

3.4.2 Segmentation

Blood vessel segmentation is an essential process for diagnosing any fundus image. It offers insight into the vessel structure as well as early localization of pathology such as hemorrhages and exhudates as well as the demarcation line. It enables detailed vessel analysis such as vessel density measurements, and tortuosity which are relevant measures in ROP as well as in DR.

Many solutions have been described [29]. First attempts have been using standard feature based description of processed images. Soares [113] describes a dual step supervised vessel segmentation technique where the first step forms a feature description of the input fundus image using Gabor wavelets. This is followed by a Bayesian classification. Another approach leveraged supervised technique using Likelihood-ratio test which comprised of multi-scale matched filtering and measurements of vessel edges with confidence levels.

Second approach for segmentation is unsupervised learning. Nguyen et al., [77] make use of line detectors. These are applied at varying scales to a fundus image. Using a set of rotated lines, the devised method segments blood vessels at various angles. Osareh et al., [86] use multiscale tuned Gabor filters to identify vessel candidates. Principal component analysis is used for feature extraction. The feature vectors are used in conjuction with Gaussian mixture model (GMM) and Support Vector machines (SVM). This allows the fundus image pixels to be classified as blood vessels or none.

Huazhu [30] describes an innovative approach that is also a two stage solution. Segmentation can be thought of as a boundary detection problem. One approach to solving this is by applying a multi-scale and multi-level CNN with a side output layer to train and map an image's hierarchical representation.

A Conditional Random Field (CRF) is next used to model long range relationships between pixels. The results are finally combined as an integrated deep neural network which has been named DeepVessel. Results against the STARE database provided 0.9523 accuracy and 0.7603 specificity.

Debapriya et al., [68] focus on a similar challenge like Hauzhu [30] but achieve similar accuracy of 94.7 accuracy using a computational imaging framework that employed ensembled based CNNs to ensure high confidence detection of vasculature using colour retinal images. The ensemble based CNNs are used for segmentation training using unprocessed colour fundus images. When inference is performed, the results are averaged to form final segmentation by using responses from each CNN making up the ensemble. DRIVE database use utilised for testing.

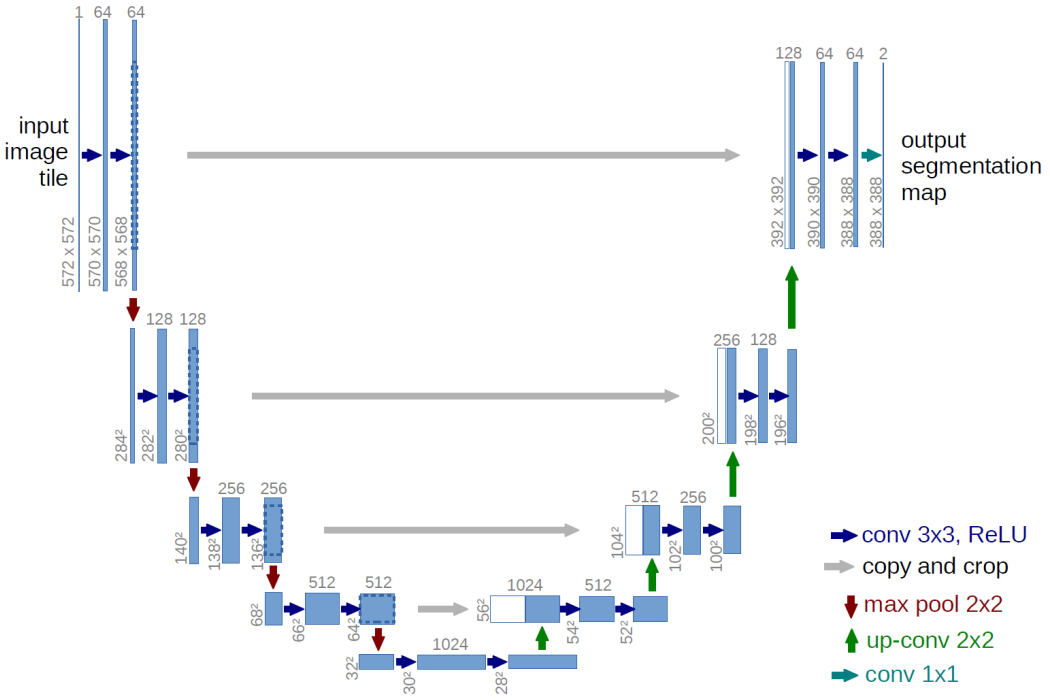
Wilfred et al., [28] after pre-processing the fundus images present enhanced features which were combined with moment invariant based features as input nodes into their CNN. The CNN is a five layered feed forward model with single neuron based output layer. The CNN has the ability to classify pixels either as vessel or non vessel classes. During the CNN training, pixel samples are obtained from blood vessel and non-blood vessel areas. Further processing is carried out in the hidden layers that includes converting the image into HSL and RGB spaces and image enhancement. Contrast enhancement is performed using distribution function. The overall process demonstrates the detection of abnormalities in the retina using image processing techniques with a low rate of false negatives.

Amna Waheed et al.,[131] proposed another blood vessel segmentation method. In this method, technique proposed by Akram et al., [2] is applied to extract vascular pattern. Connected component analysis is performed to

identify and extract candidate regions. A region is denoted as each connected segment in the vessel map. Features are defined using morphological and intensity values. Each candidate region is formulated by a seven dimensional feature vector. Identified features are fed into the Localized Fisher Discriminant Analysis (LFDA) for pixel classification. The method reported accuracy is 0.9581.

Most promising state of the art segmentation methods involves using CNNs. Olaf et al., [104] proposed a CNN architectural model constructed known as U-Net named after its symmetrical U-shape architecture as shown in Figure 3.18. U-Net has found a niche primarily in retinal image segmentation with its ability to learn complex representations. It's first part is an encoder which captures the contextual information from the input image during its downsampling steps that extracts hierarchical features. It then performs the upsampling steps using related features as detected in the downsampling preserving the spatial data which allows the formation of precise and accurate segmentation maps. Several variations on U-Net have been published on adult blood vessel segmentation [103, 49, 111, 132, 31]. It's strength lies in its ability to train using very few images.

Figure 3.18: U-Net Architecture [104]



Only few CNN based segmentation methods have been used for pediatrics opthalmology purposes primarily due to a lack of open source dataset. Goji et al., [36] used segmentation approach utilising 3 models earlier trained on open source fundus based segmentation datasets. They tested a limited number of ROP images with limited result due to challenges with choroidal vessels that are present in a premature infant retina, a fact highlighted by Reid et al.,[101].

Using U-Net, Yildiz et al. [142] performed retinal blood vessel segmentation with the aim of generating a vessel tree. The vessel tree was used to extract features to determine Plus disease.

Madhu et al. [67] used a modified U-Net with EfficientNet as its backbone. The pre-trained imagenet base EfficientNet [119] used together with its

weights. This model, referred to as ROP-seg, was then trained on new ROP images available to the authors. It showed good results visually.

With above literature review, U-Net was the preferred choice for ROP vessel segmentation. As a lot of work has already been done on this using adult fundus images. Therefore, we selected the latest state of the art LWNET [31] which is a minimalist model having a pre-trained model for usage with ROP RetCam image blood vessel segmentation as an intermediate step.

3.4.3 Classifiers for fundus pathology

a. CNN based Classifiers on Diabetic Retinopathy

María García et al., [33] pursued segmentation of hard exudates in a fundus image by using a multilayer perceptron (MLP) classifier. Alireza Osareh et al., [85] used similar classification extended it for for exudates and non exudates classes from segemented regions. They also compared the performance of various classifiers for this task. R. Priya and P. Aruna [95] utilised SVM to detect DR stages from colour fundus images.

Ghosh et. al., [35] describe a CNN to predict different disease grades for DR. Utilising the EYEPACS dataset, they were able to achieve 85% for the 5 class classification and 95% accuracy for the 2 class classification (DR or No DR). Of note was the use of flipped images for augmentation. Nguyen et al., [77] proposed a CNN using multilayer feed forward network to classify DR.

Parhan et al., [59] describe a CNN for analysing retinal fundus images. The aim was to detect and classify 3 pathologies in DR namely, exudates, haemorrhages and microaneurysms. It generated probability map for the 3 pathologies using softmax output of the layers. This probablity map was

then used to segment the fundus image and highlight the areas of interest by pathology. Test results show accuracy and selectivity of 0.98 each as well as sensitivity of 0.96 using the publicly available database DIARETDB1.

Aaron et al., [13] leverage CNN to classify whether the quality of fundus photos for ROP are good or not instead of the pathology itself. The CNN can sufficiently provide a high degree of discrimination between low and high quality fundus images. It has accumulated the feature knowledge that discerns the difference between a good and bad image quality. Such model can be trained for data quality analysis that serves as a quality metric for all imaging modalities ranging from fundus images, xrays and MRI.

Yang et al., [141] show an automatic DR analysis algorithm which is based on a two stage CNN. It is able to grade the severity of DR but also identify the location of the lesion in the fundus image which was a first. Both DR severity and lesions are completely separate classification, they integrate the networks providing for more complete and specific features for DR analysis. By introducing an imbalanced weighting map, priority is given to lesion regions for DR grading that results in significant performance. Results are comparable to trained human observers.

Gulshan et al., [41] created a Inceptionv3 based CNN algorithm for detecting DR and diabetic macular edema from a retinal fundus image. The network used a algorithm to combine local proximity pixels as local features. These were then aggregated those into global features. While the CNN did not explicitly detect lesions, it had learnt to recognise them using the local features. Using EyePACS-1 and Messidor-2 datasets, very high confidence results were obtained against ground truth. Further in 2019, Gulshan et al., [41] leveraged an optimised deep neural network which they used in 2016 against real data

of 3049 patients at Aravind and Sankara Eye Hospitals in India. Optimization was performed using larger training sets, better tuning of hyper-parameter exploration, larger input image resolution and improved Inception-v4 architecture. Adjudicators Results were validated against both and validated results against a retina specialist as well as trained grader. Deep learning showed k-Quadratic Weighted Scores for 5 Point DR Grading at 0.85 vs 0.0.82 for a retina specialist.

b. Transformers based classifiers for Diabetic Retinopathy

ViT [20] been investigated for DR classifications and has outperformed CNN. Key works are noted below.

Mohan et al., [72] proposed using state of the art ViT [20] for DR classification. Using Kaggle and IDRid DR datasets, the images were separated into non-overlapping patches for location retention. Sequences were generated from these patches prior to linear/positional embedding. Final representation was performed with token sequences used for classification. Comparative transfer learning CNNs experiments were also performed using the same datasets. ViT results were better than peer transfer learning CNNs.

Nazih et al. [76] used most recent open source dataset created specifically for ViT and referred to as fine-grained annotated diabetic retinopathy (FGADR) [147]. It is noted that this dataset is imbalanced and data augmentation techniques were used in conjunction with identify correct metric, weights, focal loss, and label smoothing. They used a novel ViT based deep learning pipeline to grade DR stages in comparison to peer transfer learning CNNs. This architecture achieved better results than comparative CNN and baseline ViT.

c. Foundational model based classifiers for Diabetic Retinopathy

Li et al. [64] proposed a prompt based strategy for guiding automatic inference. This was to overcome poor results for segmenting DR lesions especially trying to distinguishing fine grain DR lesion categories. They used Vision Language Model (VLM) that uses exact lesion description instead of implicit class name. This allowed for differentiating within a given class of categories. By using an injector which accepts the explicit textual directive allows them to locate DR lesions correctly resulting in better segmentation.

d. CNN based classifiers on Retinopathy of Prematurity

Recent approaches with promising results for automated ROP pathology detection are based on convolutional neural networks (CNN). These use ROP RetCam images as input and do not require manual annotation. These methods for ROP detection are capable of coarse-grained classification, such as discriminating severe from mild ROP. They do not specifically assess disease Stages or Zones. While the literature suggests that severe disease rarely develops without changes in posterior pole vasculature, providing additional outputs of the Zone and Stages could improve the interpretability of the system's assessment.

Worral et al., [137] were the first to use CNN for ROP. Using approximately 1500 images from 35 patients and 347 exams (2-8 images per eye), they utilised a Bayesian CNN approach for posterior over disease presence, and thereafter used a second CNN trained to return novel feature map visualisations of pathologies. They achieved an accuracy of 91.80% in terms of

identifying eyes with ROP versus healthy eyes, with sensitivity of 82.5% and specificity of 98.30% per eye.

Wang et al., [133] architected the largest automated ROP detection system named DeepROP using DNNs. The largest dataset at the time used 20,795 images, 3722 cases, and 1273 patients, with clinical labels added by clinical ophthalmologists. ROP detection was divided into ROP identification and grading tasks using two DNN models, namely Id-Net and Gr-Net. They first determined if ROP was present, and if present, they graded ROP either as Minor (Zone II/III and Stage 1/Stage 2) or Severe (threshold disease, Type 1, Type 2, AP-ROP, Stage 4/5). No finer details are determined such as Plus disease, which Stage of ROP and Zones individually. Standard imaging technology was used, with private datasets and there is no published way to reproduce these findings. Images used were of a single population cohort and pre-processing datasets results were not clearly identified prior to training. Their results were validated against 472 screenings from 404 infants, and resulted in identification of ROP with sensitivity of 84.9%, specificity of 94.9% and accuracy of 95.6%. Minor/Severe classification was achieved with sensitivity of 93%, specificity of 73.6% and accuracy of 76.4%. It outperformed 1 out of 3 experts.

Brown et al., [8] devised a two stage CNN i-ROP-DL which combines prediction probabilities via a linear formula to compute an ROP severity score, which can serve as an objective quantification of disease. A similar idea could provide finer grading of Plus disease. The i-ROP-DL deep learning system was the first to detect specific ROP classifications. 6000 images were used and professionally labelled as normal, pre-Plus, Plus ROP stages. Partial/full detachment images were excluded. Two stage CNNs were used. One was to

segment and the other was to classify. The classifier classified only Plus disease, and their solution was limited to handling one type of imaging source, as indicated by the authors . They achieved 91% accuracy with sensitivity and specificity of 93%, 94%, respectively. For Plus disease, they achieved 100% accuracy, and a maximum of 94% for Pre-Plus. This outperformed 6 out of 8 ROP experts.

Mulay et al., [74] focused on detection of the demarcation line using a CNN based model, Mask R-CNN which used generated mask to predict ROP Stage 2 classification based on ridge line presence. Image pre-processing was used to overcome poor image quality. Detection accuracy was 88% with positive predictive value of 90%/75% and with negative predictive value of 97%/42% in terms of sensitivity/specificity. This demonstrated that image pre-processing assists deep learning classification for ROP Stages.

Vinekar et al., [130] used deep learning with 42,000 images from a tele-ROP screening in India, with the goal of detecting the presence of Plus disease. Using two test sets that excluded pre-Plus disease, they were able to achieve 95.7%/99.6% and 97.8%/68.3% in terms of sensitivity/specificity.

Tong et al., [125] architected another novel approach to identifying ROP and its stages using 2-layer CNN DL and predict lesion location in fundus images. Datasets of over 36,000 RetCam images were reviewed and labelled by 13 ophthalmologists into training and validation dataset using 90:10 split. Each image was labelled with a classification (ROP severity) and identification label (ROP clinical stage). Normal when no abnormalities, mild for Stage 1/Stage 2 without Plus disease, semi-urgent for Stage 1/Stage 2 with Plus disease, urgent for Stage 3, Stage 4, and Stage 5 with or without Plus disease. Very basic pre-processing which included adding noise and brightness to

images was performed, and the dataset was then augmented for volume and images were compressed. Resnet 101 CNN using transfer learning was used for classification and Region-Based Convolutional Neural Network (R-CNN) was used to identify ROP stage, presence of Plus disease and predict objective boundary of lesions. Very good accuracy was achieved by using transfer learning of 88.30%, 90%, 95.70% and 87% for Normal, mild, semi-urgent and urgent results with additional accuracy of Stage and Plus disease at 95.70% and 89.60%. A repeat without transfer showed poor sensitivity and accuracy. They do show breakdown for Stages as well Plus disease with which we can compare our results (see 6.3.6). As they were unable to differentiate Stage 0 and Stage 1 due to very subtle demarcation line presence, Stage 0 detection was dropped.

Wu et al., [138] developed 2 deep learning models to demonstrate that a CNN can be used to predict the occurrence and severity of ROP. Two models were specifically designed to look for presence and severity of ROP using ResNet50. Severity was defined as *mild* if the patient had Type II ROP, Zone II Stage 1, Stage 2 ROP without Plus disease and Zone III Stage 1,2, or 3 ROP. Severity was *severe* if the patient had Stage 4 or 5 ROP, Type I or aggressive posterior ROP. Results for presence of ROP was 100%/38% and severity of ROP was 100%/47% in terms of sensitivity/specificity. Here too, image pre-processing was not highlighted as having been used.

Ding et al., [18] proposed a hybrid architecture focused on localization of the demarcation line and in parallel feeding a black and white reciprocal image as 2 channel input. The former provides the input of an area of interest. It is fed into traditional CNN. The paper conducted experiments using 2759 images from a local hospital, that were classified by Stages 1-3 only. Images

were pre-processed using contrast enhancing features. The pre-processed image was used for object segmentation by applying Mask R-CNN to highlight the demarcation line at pixel level and generate a bounding box around the area of interest, creating a binary mask for the demarcation line. The 299x299 resized image was then used for training with Transference learning to prevent over-fitting. Inception v3 pre-trained on ImageNet was used for classification. Stage 1, 2 and 3 classifications achieved 77%/78%, 62%/61%, 62%/62% in terms of sensitivity/specificity. Of significance is their ability to detect between Stage 1 and 2. However, it was not clear if Stage 0 images were completely removed and we assumed this was the case given that Stage 0 versus Stage 1 is a challenging area due to the faint nature of the demarcation line.

Zhao et al., [146] turned to deep learning for identifying Zone I using RetCam images. Their approach is based on first identifying the position of the optic disc and macula. Thereafter, the center point of the optic disc and macula were calculated, from which Zone I was then calculated. This is the only study we know of that identifies Zone I. They achieve accuracy of 91% with an IOU threshold of 0.8. They acknowledge that disc and macula positioning is required without which this study is incomplete. An automated Zone Quality Filter is a good contribution.

Agarwal et al [61] used OD detection using circle Hough transforms in conjuction with IUnet based blood vessel segmentation. Using this they determine Zone classification. They were able to detect classify ROP in Zone I with 88.23% accuracy.

The above works showed promising results but had similar gaps. First challenge was with non-independently verified private ROP RetCam datasets. Next, dataset used were lacking geographical diversity to address this as a

worldwide challenge. At time of writing, there was no public labelled ROP population diverse dataset which could be used to compare and measure results with, due to the sensitive nature of paediatric data. A standardised open source ROP labelled dataset is critically needed by ROP AI researchers to create diagnostic solutions. It will allow for reproducibility of published work and aid future works. In terms of purist approach to Stages, and Zones, only Ding et al., [18] who focused on Stage only. Others such as Tong et al., [125] combined various Stages and Plus disease for broader outcomes. Similarly Wu et al., [138] combined Stage and Zones minus Plus disease. Fundus pre-processing in all papers to date is very general except for Ding et al., [18] and Mulay et al., [74] who used R-CNN to amplify the demarcation line. This motivated our study on researching improvements in the image pre-processing to assist in ROP classification.

As part of our belief in open and collaborative research for the automation of ROP diagnosis using CNNs, Dr Sabri and team created a labelled ROP repository which is available to use for fellow researchers with prior authorisation. This repository continue to will add more ROP labelled images over time for future use. This repository was used to support our focus on ROP RetCam pre-processing techniques that enhance overall features pertinent in providing high quality classifier outcomes in terms of Plus, Stages and Zones individually instead of grouping different ROP aspects. Data augmentation is used to supplement the lack of a larger ROP dataset with transfer learning CNNs. We demonstrated that improving ROP features for ROP RetCam images using our improved pre-processing methods does result in improved deep learning classifier results. These methods can also aid other researchers who

can re-try their experiments and provide their input on improvements using these methods.

As the problem of ROP RetCam images is a multifaceted challenge, we solve this by employing both image domain and restoration methods. In the latter, we solved the problem of over/under amplification inherent in Pixel Colour Amplification while fine tuning Double Pass Fundus Reflection specifically for ROP RetCam images. These changes allow the pre-processed images to be more suitable for deep learning classifiers for ROP. We then combined our improved restoration methods and classic image domain method as hybrids which further improved overall features. Using each pre-processing method, a training and validation pre-processed dataset was created. Each pair was used with two different classifiers specific to Plus, Stages and Zones for ROP and results noted.

3.5 Chapter Summary

This chapter summarised the background to several areas on image pre-processing including related works. This includes image pre-processing with respect to fundus images, different segmentation approaches that range from classical to newer CNN based models. CNNs and their relationship to machine learning are described. Various classifier models that use transfer learning are reviewed. Literature review on the use of CNNs in ROP classification and DR. The review highlights challenges in ROP especially lack of public data and private datasets which cannot be corroborated in terms of results replication for detailed peer review. This review to date, did not identify any new novel method usage prior to using the classifiers specifically using restoration domain methods.

Hence, this summarisation allows us to restate the RetCam image quality problem as a multi-faceted challenge. Both image domain and restoration methods are necessary for a hybrid approach leveraging their unique strengths. Using the restoration method as best focus area, Pixel Colour Amplification (PCA) and Double Pass Fundus Reflection (DPFR) were identified as good research areas with potential use in ROP provided their key deficiencies were addressed. With this in mind, this research focused on finding effective fundus pre-processing techniques that can enhance overall ROP clinical features that provide best classifier outcomes in terms of Plus disease, Stages and Zones. Their efficacy in CNNs can then be shown by creating a CNN transfer learning based model. This can be regarded as one of the major achievements in this research and is now discussed in detail in subsequent chapters.

Chapter 4

Research Methodology and Dataset Collection

This chapter presents the design of this research in Section 4.1. Data collection process is noted in Section 4.2 which includes detailed criteria for grading Plus disease, Stages and Zones. The data augmentation is a necessary component for limited data. The methods used are described in Section 4.3 leading to unique images ideal for training purposes. Metrics for data analysis are described in Section 4.4. Summarised dataset results from augmentation are collated for Plus disease, Stages and Zones and reviewed in Section 4.5 . Methodology summary and data collection is noted in Section 4.7. A proper dataset which is graded and labelled correctly for Plus disease, Stages and Zones is a key requirement for the respective CNNs such that they can grade the ROP conditions correctly.

4.1 Research Design

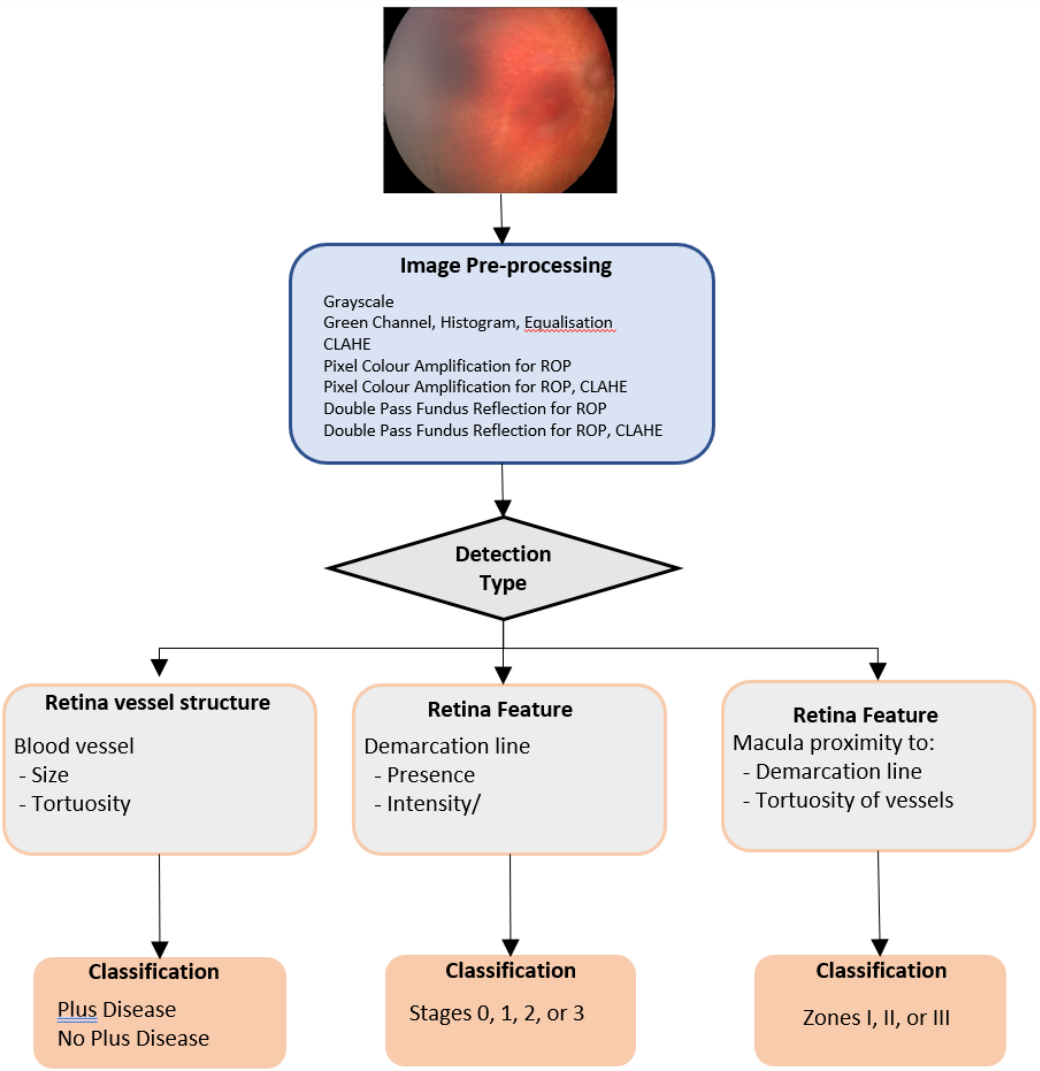
The topics performed during this research included the following:

1. Understood the ROP RetCam quality problem
2. Reviewed the related CNN based ROP classification literature as well as expanding on similar aspects in an adult pathology namely DR
3. Reviewed the image domain and restoration methods available to date and their application to DR and ROP
4. Proposed the purpose of the study to improve ROP RetCam quality
5. Collection of data which includes images and labels and creating correct sets for Plus Disease, Stages and Zones
6. Solved the challenge of ROP RetCam image quality by pre-processing the data using improved novel restoration based methods and combined with traditional methods.
7. Staged the data for the classifiers using augmentation for training and non-augmented for validation
8. Developed classifiers for Plus disease, Stages, Zones
9. Trained the classifiers for Plus, Stages, Zones using respective data repeated per pre-processing method
10. Collected classifier results and analysing them

Figure 4.1 represents the high level design for the overall development for detection and classification of ROP RetCam images using the outputs the

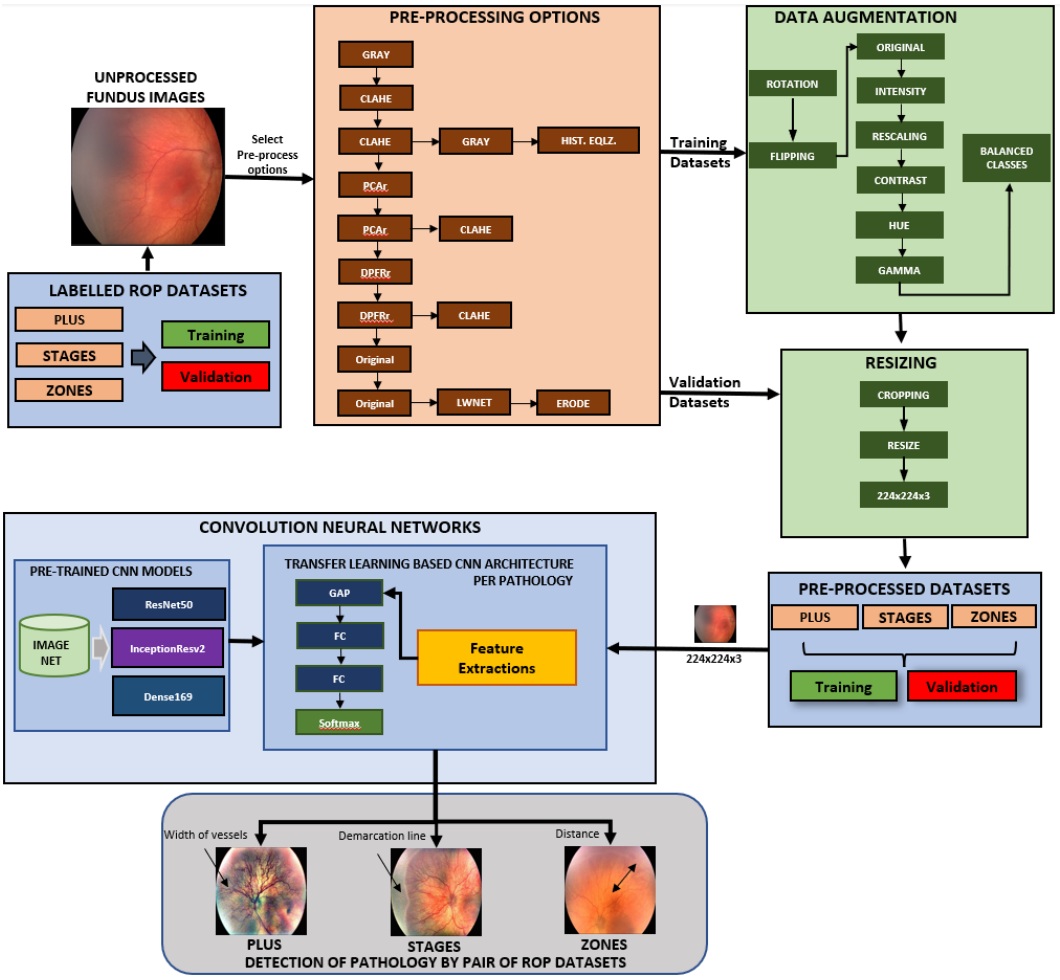
various image pre-processing methods that includes the improved novel pre-processing contributions.

Figure 4.1: ROP Research design flow



From the high level design, we developed with the following architecture we refer to as McROP. Described in Figure 4.2, it facilitates the steps required by the research process. These steps are as follows:

Figure 4.2: McROP Overview



1. **Data Collection**: Creation of Plus disease, Stages and Zones datasets. These were split into Training and Validation pairs

2. **Image Pre-processing:** One of the seven available option (excluding LWNET) chosen

3. **Resizing:** Cropping to maximise the image and image resizing was carried out

4. **Data Augmentation**: This optional step was needed when there was insufficient training data to create unique images from each original single image. It was also used to balance the training dataset

5. **Pre-processed dataset**: Resizing of data was done for both Training and Validation pair of either Plus disease, Stages and Zones dataset.

6. **Convolution Neural Networks:** These transfer learning based CNNs were created for Plus disease, Stages and Zones. They were trained using the received datasets. Validation and testing produced the results for these separately by respective CNNs.

The output of these steps was the development and validation sets of (pre-processed) images, CNN model and CNN validation results for that set. These sets were repeated for each of the pre-processing techniques.

4.2 Data Collection

We now review the data collection method. We first describe each ROP RetCam images and provide their labels. We define how each image was qualified for Plus, Stages, and Zones in Section 4.2. As we had limited set of training data available, effort was required to expand the number of unique images. Section 4.3 provides oversight on how data augmentation was used to create more unique ROP images using the original image. Section 4.5 concludes this overall chapter.

Dataset collection for ROP remained the biggest challenge as there are no open source ROP datasets. A small ROP RetCam dataset was obtained as part of a collaboration agreement with University of Calgary with Dr Anna

Ells's support. We refer to it as the Calgary ROP Dataset. A dataset of 1778 images was provided in 191 folders. Each folder represented patient number and examination number. In each folder, there were two folders. One folder had left and other with right eye images for the patient. Each folder had between 4-8 images per eye. This represented a total of 29 patients with up to 9 visits. Each visit captured bilateral images (left/right) of up to 6 images.

These original images were captured using RetCam camera in 640 × 480 pixels. Dr Ells provided us with a ground truth for these images. These were rechecked for a second time by Dr Kourosh Sabri as second grading specialist. Where there was a difference, the new ground truth was noted. Data was anonymous when received. Age, gender and race were not identified as per health data privacy requirements. Data quality guidelines included firstly identifying and excluding images which were badly focused, over illuminated, missing key retina features, having laser treatment marks or too many external reflections. Image of the exterior of the eyes were also excluded. This manual process highlighted a crucial gap in overall ROP data collection process, namely an automated CNN based Data Quality Filter tool. We will consider this as part of our future work. The remaining images selected for the further classification were still challenged by poor illumination, and at times are slightly unfocused.

The grading on qualified images, was completed based on the guidelines set out by the American Academy of Pediatrics [25]. The general criteria for determining the RetCam fundus image qualification for Plus disease, Stages and Zones are as follows:

1. **Plus disease**

For a given image, determination of the blood vessels' tortuosity is key. Plus disease looks for the abnormal width and shape of the retinal blood vessels. If larger and showing tortuosity, then it is categorised as Plus Disease. There does however exist a Pre-Plus disease which we do not consider in this research. Any images with Pre-plus disease were removed and kept aside for future work. If neither was present, it was denoted as No Plus.

2. **Stages 0-3**

   For Stages, the criteria is to look for the demarcation line. As per the criteria laid out, the demarcation line can appear as not being present within the peripheral region. This is categorised as Stage 0. The observability of the demarcation line and its magnitude of size and shape determines if it is Stage 1, 2, or 3. In Stage 3, the ridge of the line is very pronounced. Stage 4 and Stage 5 are when the retina detachment has commenced leading to full retina detachment. In our dataset, we did not have Stage 4 and 5 and therefore did not categorise these 2 conditions. Only Stages 0, 1, 2, and 3 were categorised.

3. **Zones I-III**

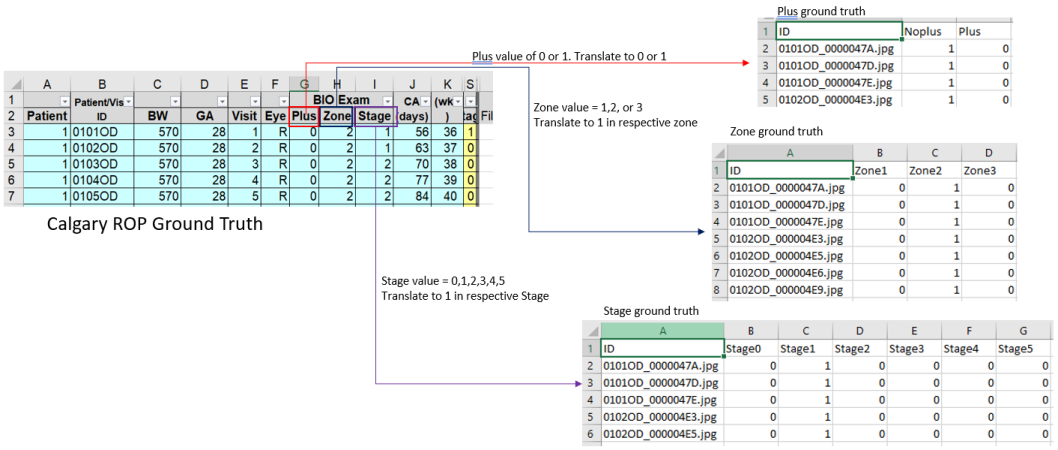
   Zone is a multi-dimension challenge. The determination is made from within the 360 rotation around the macula, noting exactly where the demarcation line and / or tortuosity lies. There are 3 Zones as previously noted in 2.2. Assessment of Stage/Plus then categorises proximity from the macula to determine the zone. In the Calgary ROP Dataset, there were few images which qualified for the criteria by having the centre of the vision and peripheral area present. This geometric aspect was the

most difficult to assess with the small number of ROP RetCam images available.

Following the above guidelines, 3 separate datasets were created. One each for Plus diseases (pre-Plus was excluded), Stages 0-3, and Zones I-III classification. Each image in the respective dataset was labelled with a classification label and filename identification label. Classification labels of Plus/No Plus were used for Plus, 0-3 were for Stages, and I-III for Zones. Figure 4.3 proves an illustration of labelling results for Plus, Stages and Zones.

Next, we needed to split these datasets into two subsets termed as Training and Validations datasets. The splits utilised were 80:20 for Stage/Zone and 90:10 split for Plus. The split also ensured that data from the same patient did not appear in the validation for the same eye. The alternate eye was used if required as a replacement. Selection for Plus disease was made by the specialists which ensured the optic disc and vascular features were visible. For Stage, the image may have the optic disc visible and/or the periphery where the demarcation line may or may not be present. For Zone, the criteria were stricter, the peripheral area must be present as well as the optic disc and macula; both must be present in order to be able to provide an assessment. This is also in line with Zhao et al. [146].

Figure 4.3: Creating, and Labelling of Plus disease, Stages, Zones datasets using Calgary ROP Dataset



There was imbalance in these three training datasets namely Plus/No Plus, Stages 0-3, Zones I-III. To make up for the lack of training data, data augmentation was performed in order to have sufficient volume that was unique individually to allow training to occur. The finalised number of images are noted in Figure 4.3.

4.3 Data Augmentation

After preparing our datasets from the prior section for Plus disease, Stages and Zones, the number of images per dataset was very limited. Data augmentation was performed to expand the image diversity but preserving the actual image prediction characteristics. This is also required to have a balanced training dataset. Several methods can be used including the geometric orientation such as rotation, lateral and vertical flip with each change in orientation being treated with scaling intensity, adjusting image contrasts. Each image is a

unique variation of the original image. For our research, the number of images required were a few thousands, so the following techniques were performed.

Algorithm 1 Augmentation Steps on an original ROP Image into multiple unique images

1: For each original image, Perform following orientation then call variations of image in 2.

   1. Original
   2. Upside Down
   3. Flip Upside Down image from 1.2 - Left to Right
   4. Rotate Original by 90 degrees
   5. Flip Original rotated by 90, Left to Right
   6. Flip Original Left to Right
   7. Flip Original Left to Right, Upside Down.

2: For each of orientation above perform:

   1. Original oriented image
   2. Change scale intensity by 0.999
   3. Rescale the intensity
   4. Contrast of Original oriented image
   5. Saturation of Original oriented image
   6. Gamma of Original oriented image
   7. Introduce hue into Original oriented image
   8. Gamma contrast of Original oriented image
   9. Sigmoid contrast of Original oriented image
   10. Linear contrast of Original oriented image

3: Save each generated output using filename, original ground truth label

The above algorithm can be explained as such. First image orientation is performed described in actions 1.1-1.7. For each image orientation, an image change was applied by each method described 2.1-2.10. Together, this mechanism gave a maximum of 70 augmented images from a single original image. The augmentation number is controllable at class level. By using this flexibility, it allows for X number of augmentation per image category. Each output file name was suffixed with Xth number being generated and outputted in a flat file. Augmentation is only required for the training dataset to provide a large balanced dataset inclusive of labelling ground truth so that overfitting during the training stage of the classifier. Examples for Steps 1.1-1.7 are shown in Figure 4.4, with corresponding image changes for a given flip in Figure 4.5. This step is not required should there be a larger training dataset.

Figure 4.4: Flipping of an ROP RetCam Image

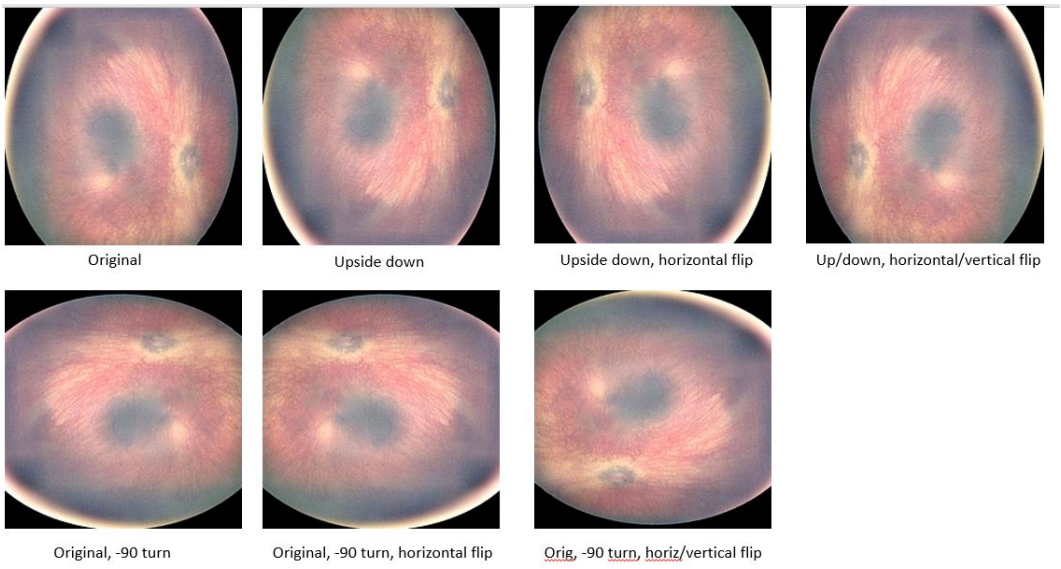
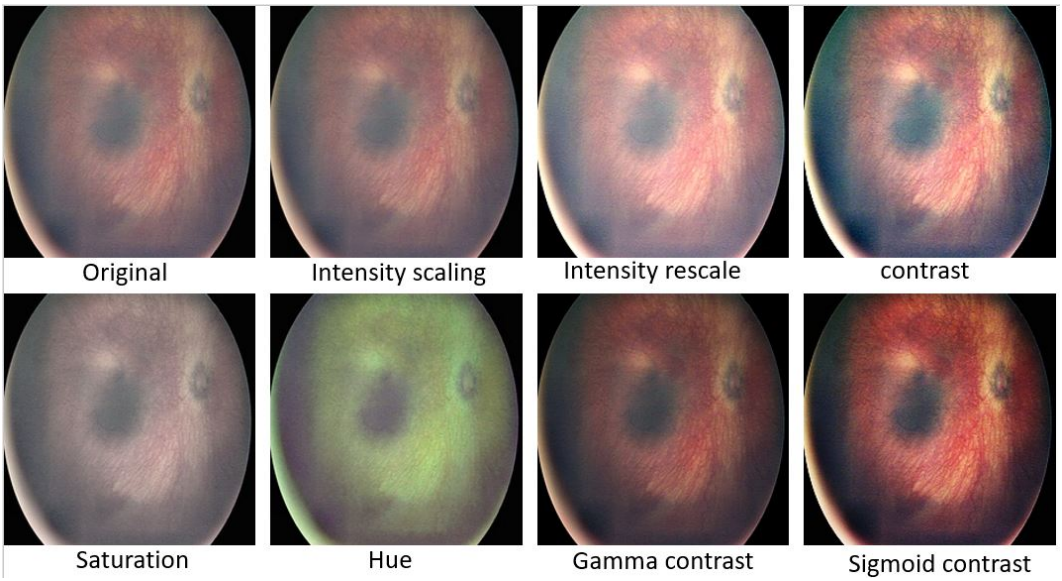
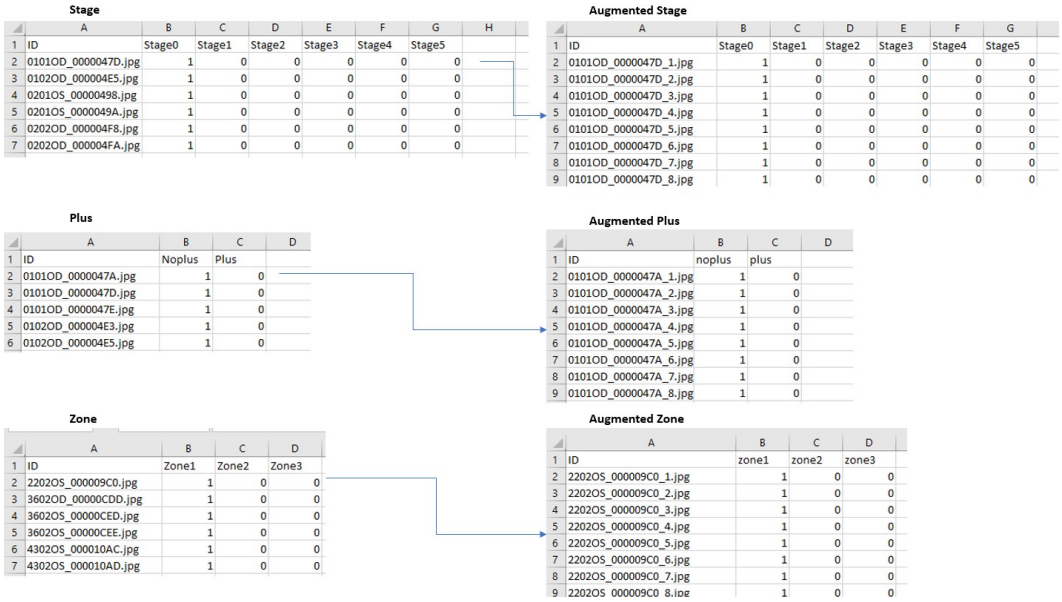


Figure 4.5: Varying image within a flip cycle



An augmented training image makes use of the original pre-processed labelled image name suffixed with a sequential key. The original ground truth preserves the diagnosis label with this new image. As noted previously, this was written out as a new sequential file name and entry for the classifier list where corresponding name and ground truth was recorded. This all done as part of the augmentation process automatically without any manual effort. The resulting augmented dataset label file including filename and labels for Plus disease, Stages, Zones can be shown in Figure 4.6.

Figure 4.6: Augmented ground truth files



Stage

	ID	Stage0	Stage1	Stage2	Stage3	Stage4	Stage5
2	0101OD_0000047D.jpg	1	0	0	0	0	0
3	0102OD_000004E5.jpg	1	0	0	0	0	0
4	0201OS_00000498.jpg	1	0	0	0	0	0
5	0201OS_0000049A.jpg	1	0	0	0	0	0
6	0202OD_000004F8.jpg	1	0	0	0	0	0
7	0202OD_000004FA.jpg	1	0	0	0	0	0

Augmented Stage

	ID	Stage0	Stage1	Stage2	Stage3	Stage4	Stage5
2	0101OD_0000047D_1.jpg	1	0	0	0	0	0
3	0101OD_0000047D_2.jpg	1	0	0	0	0	0
4	0101OD_0000047D_3.jpg	1	0	0	0	0	0
5	0101OD_0000047D_4.jpg	1	0	0	0	0	0
6	0101OD_0000047D_5.jpg	1	0	0	0	0	0
7	0101OD_0000047D_6.jpg	1	0	0	0	0	0
8	0101OD_0000047D_7.jpg	1	0	0	0	0	0
9	0101OD_0000047D_8.jpg	1	0	0	0	0	0

Plus

	ID	Noplus	Plus
2	0101OD_0000047A.jpg	1	0
3	0101OD_0000047D.jpg	1	0
4	0101OD_0000047E.jpg	1	0
5	0102OD_000004E3.jpg	1	0
6	0102OD_000004E5.jpg	1	0

Augmented Plus

	ID	noplus	plus
2	0101OD_0000047A_1.jpg	1	0
3	0101OD_0000047A_2.jpg	1	0
4	0101OD_0000047A_3.jpg	1	0
5	0101OD_0000047A_4.jpg	1	0
6	0101OD_0000047A_5.jpg	1	0
7	0101OD_0000047A_6.jpg	1	0
8	0101OD_0000047A_7.jpg	1	0
9	0101OD_0000047A_8.jpg	1	0

Zone

	ID	Zone1	Zone2	Zone3
2	2202OS_000009C0.jpg	1	0	0
3	3602OD_00000CDD.jpg	1	0	0
4	3602OS_00000CED.jpg	1	0	0
5	3602OS_00000CEE.jpg	1	0	0
6	4302OS_000010AC.jpg	1	0	0
7	4302OS_000010AD.jpg	1	0	0

Augmented Zone

	ID	zone1	zone2	zone3
2	2202OS_000009C0_1.jpg	1	0	0
3	2202OS_000009C0_2.jpg	1	0	0
4	2202OS_000009C0_3.jpg	1	0	0
5	2202OS_000009C0_4.jpg	1	0	0
6	2202OS_000009C0_5.jpg	1	0	0
7	2202OS_000009C0_6.jpg	1	0	0
8	2202OS_000009C0_7.jpg	1	0	0
9	2202OS_000009C0_8.jpg	1	0	0

4.4 Data Analysis

To summarise this process, the ROP dataset as provided by University of Calgary was original and new. This dataset was first checked for anonymity. Independent specialist grading was performed and respective ground truth labels provided. Images of very poor quality for this research were removed and kept aside for future use. The data was prepared for classifier training as validation and training datasets using splits as described previously. Augmentation was performed on the training dataset. Both training (augmented) and validation datasets keep their respective assigned ground truth labels. Training and validation Data which has been graded has 4 possible outcomes in all 3 ROP sub-areas (Plus Disease/Stages/Zones) by the classifiers:

1. **True Negative (TN)**: image was normal and reported as normal

2. **True Positive (TP)**: Image showed Plus disease or having a Stage, or Zone and correctly reported by the screener as having such corresponding outcomes. This would require specialist intervention.

3. **False Negative (FN)**: image had a pathology but was reported with incorrect outcome.

4. **False Positive (FP)**: Image was normal but disclosed by respective classifier as having any of the ROP conditions.

We used the following four metric indicators to evaluate the classifiers. These are Sensitivity (Equation 4.1), Specificity (Equation 4.2), Precision (Equation 4.3), F1-Score (Equation 4.4), and Accuracy (Equation 4.5) is the percentage of accurately classified images. These are described [15] below.

$$\textbf{Sensitivity/Recall} = TP/(TP + FN) \tag{4.1}$$

$$\textbf{Specificity} = TN/(FP + TN) \tag{4.2}$$

$$\textbf{Precision} = TP/(TP + FP) \tag{4.3}$$

$$\textbf{F1} = 2 * (\textbf{Precision} * \textbf{Recall})/(\textbf{Precision} + \textbf{Recall}) \tag{4.4}$$

$$\textbf{Accuracy} = (TP + TN)/(TP + TN + FP + FN) \tag{4.5}$$

Sensitivity (SN)/recall (REC) describes the true positive percentage which is the total correct positive predictions in relation to total number of positives. It is also referred to as True Positive Rate (TPR). The best sensitivity is 1.0 with 0 being the worst. Specificity describes percentage of identified as correct negatives using the total number of correctly identified negatives in relation to total number of negatives. It is also referred to as true negative

rate (TNR). The best specificity is 1.0 with 0 being the worst. Precision also known as positive predictive is computed using the number of correct positive predictions with the total number of positive predictions. Accuracy is determined by using the all correct predictions divided by the total number in the dataset. The choice of correct metric is essential when there are imbalanced training datasets. In these cases, F1 score is better guide as a mean of precision and recall as it improves upon classifier correctly classifying the correct classes.

To aid in this classification categorisation, we used confusion matrix that is used in classification methods. It is in a matrix based on the number of classes. For example, in Plus diease, there are two classes namely an image has no Plus disease or has Plus Disease. Similarly Stages has 6 classes namely Normal/Stage 0, Stages 1-5. However in our case, we only categorise Stages 0-3. Zones will have 3 classes, Zones I, II and III. Using binary classification as an example, we show this confusion matrix in Table 4.1.

Table 4.1: Binary Confusion Matrix

	Positive	**Negative**
Positive	**TP**	**FP**
Negative	**FN**	**TN**

In binary classification, either the prediction is true or false. Correctly classified run diagonal from top left most to bottom right most. For a multi-class classification such as for Zones, the confusion matrix will run as shown in 4.2. The top horizontal row designated the predicted classes while the left vertical row designates actual classes. Correctly predicted zones/actual zones are diagonally from top left to bottom right.

Table 4.2: 3 classes Confusion Matrix for Zones

		Prediction		
		Zone 1	**Zone 2**	**Zone 3**
Actual	**Zone 1**	**Predicted Zone 1 as Actual**	Predicted Zone 1 as Zone 2	Predicted Zone 1 as Zone 3
	Zone 2	Predicted Zone 2 as Zone 1	**Predicted Zone 2 as Actual**	Predicted Zone 2 as Zone 3
	Zone 3	Predicted Zone 3 as Zone 1	Predicted Zone 3 as Zone 2	**Predicted Zone 3 as Actual**

With these calculation basis identified, our experiments will focus on obtaining highest level of sensitivity and specificity possible between comparative pre-processing method based CNN classification. Sensitivity will provide us correctly identified positives if this number is relatively high e.g. greater than 60% as our chosen threshold. Similarly, we are also seeking for a high specificity rate which ascertains correctly identified negatives that are genuinely negatives e.g. greater than 60% as our chosen threshold. In conjunction with these 2 metrics, precision is sought to be of a higher value indicating that these values can be trusted by virtue of its formula that uses FP in its denominator. Lastly F1 score will provide us a good ratio of precision and sensitivity. All papers we have reviewed focused on these first three key metrics which provides a comparative basis between our work and theirs.

4.5 Developed ROP Datasets

As discussed previously, there were significant challenges in getting ROP data as there is no public graded dataset. Upon obtaining the Calgary ROP dataset, we were able to review, grade and filter corrected labelled datasets. Using the information as presented in Section 4.2, we obtained and then created training and validation datasets for Plus, Stages, and Zones categorization.

As the volume of images was challenged, augmentation had to be used that is described in Section 4.3. Each image is a unique variation of the original image. Depending on the class imbalance, the multiple was altered to ensure we get a close enough balanced set. We also used weights in the classifier to mitigate overfitting. Extensive use of the OpenCV library of image processing methods [84] was used. Very slight imbalance was present nevertheless.

For Plus disease, a total of 1001 images were isolated of which the training data set had 916 split between 855 with No Plus disease and 61 as Plus disease. The training data set had 85 images which were split between 72 as No Plus Disease and 13 with Plus disease. With augmentation on the training dataset, this number was increased to 8550 No Plus and 6039 as Plus disease. There was a slight imbalance between No Plus and Plus but acceptable slight overfit may occur mitigated by use of weights in the respective CNNs.

For Stages, we had 369 total images for Stage 0 - 3. 300 images were segregated for Training while 60 were separated for validation. The training data set had 71, 47, 101, 85 representing Stage 0,1,2,and 3. These were augmented to 3905, 2585, 4040, 3400 for the respective stages noted earlier. The validation data set of 60 images was split between 15, 9, 20, and 14 for each of the stages respectively.

Zones had the least amount of data given the multiple criteria that need to be fulfilled. In total 214 images were available; split between 173 and 41 for training and validation. The training dataset had 11, 143, 19 images for Zone I, II, and III respectively; with augmentation, this was increased to 539, 5005 and 931 respectively. The validation dataset of 41 had 3, 33, and 5 images for Zone I, II, and III respectively.

This research was limited by the amount of ROP data, and even with augmentation the balancing was not perfect. The augmented datasets were created post the pre-processing using the resized images. Each training augmented dataset had a folder with its series of images and a labelled file that had the filename, augmented series number plus the grading as shown in Figure 4.6.

Summarised Datasets for Plus, Stages and Zones are illustrated in the following three Tables 4.3, 4.4, 4.5 below:

Plus Disease Datasets			
	Training	**Augmented Training**	**Validation**
No Plus Disease	855	8550	72
Plus Disease	61	6039	13
	916	**14589**	**85**

Table 4.3: Plus Disease Dataset - Training, Augmented, Validation

Stage Datasets			
	Training	**Augmented Training**	**Validation**
Stage 0	71	3905	15
Stage 1	47	2585	9
Stage 2	101	4040	20
Stage 3	85	3400	14
	304	**13930**	**58**

Table 4.4: Stages Dataset - Training, Augmented, Validation

Zone Datasets			
	Training	**Augmented Training**	**Validation**
Zone I	11	539	3
Zone II	143	5005	33
Zone III	19	931	5
	173	**6475**	**41**

Table 4.5: Zones Dataset - Training, Augmented, Validation

4.6 ROP Website

The datasets created are a result of work outlined in this chapter and are used by this research project. Dr Kourosh Sabri created a website that stores the graded and labelled ROP RetCam images so that they are accessible by qualified researchers using prior authorisation. The dataset is accessible at https://screenropimages.ca. The website contains the original dataset and graded dataset with specialist labelled diagnosis. The availability of this graded dataset for ROP for authorised research use is of real benefit to the research community, given the unavailability of graded and labelled ROP RetCam data.

4.7 Chapter Summary

This chapter discussed overall data collection and grading approach ROP RetCam fundus images. Data Collection which categorized images into 3 separate classes namely Plus Disease, Stages and Zones. The limited dataset meant we had to use augmentation to expand and balance the number of unique images

for overcoming this deficit. Augmentation carried forward the ground truth into the augmented ground truth. This culminated in a developed dataset for Plus Disease, Stages, and Zones. This was split into training and validation datasets, based on the ratio for Plus Disease, Stages and Zones. This completed the most critical component of dataset development. It ensured data was graded within sufficient quality. For training datasets only, augmentation had to be used employing a wide variety of methods to create a larger variety of data automatically while preserving the initial ground truth. The datasets prepared as part of this project are available on a secure website for use with correct authorisation. These graded datasets can be beneficial for further research using ROP RetCam image based diagnosis and evaluation of peer CNN based systems.

Chapter 5

Image Pre-processing for improving ROP RetCam CNN classification

Previous chapters have described key aspects required to address the challenges of ROP classification for Plus disease, Stages and Zones. The first step in our frame work is to use image pre-processing capable of resolving ROP RetCam images that highlighted the salient clinical features of Plus disease, Stages and Zones. Thereafter, transfer learning based CNNs are created and trained using these datasets. The validation data is then used to measure the accuracy of the CNNs. Section 5.1 introduces the initial background for this work specifically using dehazing theory. Section 5 focuses on image pre-processing and transfer learning CNNs. Section 5.2 discusses Pixel Colour Amplification extension which resolves the illumination problem. Section 5.3 examines Double Pass Fundus Reflection and its gap in terms of ROP Retcam. Also, the solution to this gap is outlined. Further noise reduction by the erosion of blood

vessels in the retina is described Section 5.4 using LWNET to generate vessel maps using them for erosion. Section 5.5 notes the challenge of resizing and recommends the correct algorithm for ROP RetCam use for CNN. Lastly Section 5.6.2 outlines the different transfer learning CNN setups for ResNet50 and InceptionResv2 for classifying Plus disease, Stages, and Zones. Section 5.7 summarises our contribution to image pre-processing for ROP classification.

5.1 Background

Image pre-processing should improve the results of deep learning classifiers. As noted in Chapter 3.1.1 and 3.1.2, there are two main groups of image processing, namely image domain and restoration methods. Image domain methods have been successfully used in various fundus processing. CLAHE is the most widespread method in fundus based pathology research including non ophthalmology medical research areas. While CLAHE is good, it does not get over the primary problem of internal reflection. Our research focused primarily on two methods recently introduced by Gaudio et al., [34] Pixel Colour Amplification (PCA) and by Zhang et al., [145] Double Pass Fundus Reflection (DPFR). Both are based on Dark Channel Prior (DCP) work which was first developed by He et al.,[42]. These and many related works are described in Chapter 3.1.2.

Starting with the image domain class models, we introduce three methods for our use. These are Grayscale, CLAHE and our derived third method using a set of image domain class methods. Extensive use is made of the OpenCV library [83, 84] for these methods. As noted before, image domain pre-processing does not completely remediate the underlying problem but simply reduces it in different ways. We refer to these so that we have a baseline comparison

against our newer methods we will introduce in terms of ROP clinical feature resolution and the subsequent CNN classification results.

1. Grayscale (**Gray**) [37] converts the colour to grayscale thereby voiding the image of any unique colour uniqueness. This prevents the CNN from identifying any colour based characteristics.

2. Contrast Limited Adaptive Histogram Equalization (**CLAHE**) [149] is used to enhance the overall contrast of the image and smoothen out the pixels. A modified approach to CLAHE introduces a cliplimit of 2.0 across all three separated channels (Red, Green, and Blue) and merged into a single output. This gives a good contrast of ROP features.

3. We derived another method by using CLAHE output and extracting the green channel only. Green channel always carries the highest level signal which aids in identifying the demarcation line visually while keeping other noise balanced. To this, Histogram Equalization is then applied for better contrasting. We refer to it as the CLAHE-Green-Histogram (**CGH**) method.

The image domain methods do not address the underlying problem of ROP RetCam image quality, which emerges when the incidental light entering the retina is reflected back from not only the retina but the additional layers starting from the surface of the eye, lens, vitreous and retinal layer itself. After reviewing the haziness of the pictures, we deduced that this reflection problem is due to the close proximity of the capturing device laying on the surface of the retina and light induced into the eye. This reflection in nature is familiar to us, as it can be seen in domestic cats or dogs eyes at night. A

human eye too has the same reflection property albeit at a reduced level. This realisation led to our improved image pre-processing based on dehazing theory [42].

A dehazing image is described by He et al., [42] and noted earlier in Chapter 3 and repeated here in Equation 5.1:

$$\mathbf{I}(x) = \mathbf{J}(x)\mathbf{t}(x) + \mathbf{A}(1 - \mathbf{t}(x)) \tag{5.1}$$

Where for a given pixel x, image $\mathbf{I}$ is observed intensity calculated from $\mathbf{J}$, the distortion free scene radiance, $\mathbf{A}$, being the global atmospheric light which approximates the colour of scattering light, and a gray scale transmission map image $\mathbf{t}$, which describes the portion of light reaching the camera unscattered. $\mathbf{A}$ can be represented as a RGB vector. The value of $\mathbf{t}(x)$ is either 0 or 1 for pixel value of x. Distortion is represented by the non-negative airlight term $\mathbf{A}$(1- $\mathbf{t}(x)$).

In order to obtain $\mathbf{J}$ following steps are needed:

Algorithm 2 Steps to get Distortion Free Image $\mathbf{J}$

1: Define atmosphere term $\mathbf{A}$

2: Derive the Transmission map $\mathbf{t}$ for the distorted image $\mathbf{I}$

3: Solve for Distortion Free Image $\mathbf{J}$

This is the foundation for DCP. Additional Priors which were noted in Section 3.1.2 using Equation 3.2 are extrapolated leading to PCA and used by DPFR. This research reviewed these two novel methods and improved them for use in resolving ROP RetCam quality for use with CNNs.

5.2 Pixel Colour Amplification Extension

5.2.1 Preliminaries

The DCP method for dehazing starts with the principle of colour in which an image is composed of 3 colours shown in Equation 5.2 that is visible as a distorted image, Equation 5.3. DCP assumption for a given non-hazy RGB image is that there will be at least one zero intensity pixel in either one of red, green or blue channel represented as $((0, g, b), (r, 0, b), (r, g, 0))$.

i. Dehazing - Dark Channel Prior Method

The mathematics of DCP as noted by Gaudio et al. [34] and rephrased below as follows. Solve Equation 5.1 for transmission map $\mathbf{t}$ by inverting the image as noted in Equation 5.4, for a given atmosphere represented in Equation 5.2. Equation 5.3 denotes the dark channel where c represents the 3 colour channels and Ω being a set of pixels in $\mathbf{I}$ neighbouring pixel x. Equation 5.5 is used to restore fine details that the prior equation loses. The undistorted final image $\mathbf{J}$ is available from Equation 5.6.

$$\mathbf{A} = (r, g, b) \tag{5.2}$$

$$\tilde{\mathbf{I}}_{dark} = \min_{c} \min_{\gamma \in \Omega R_I(x)} \frac{\mathbf{I}c(\mathbf{y})}{\mathbf{A}c} \tag{5.3}$$

$$\tilde{\mathbf{t}}(x) = 1 - \tilde{\mathbf{I}}_{dark} \tag{5.4}$$

$$\mathbf{t}(x) = guidedFilter(\mathbf{I}, \tilde{\mathbf{t}}(x)) \tag{5.5}$$

$$\mathbf{J(x)} = \frac{\mathbf{I} - \mathbf{A}}{\max(t(x), \epsilon)} => \mathbf{J} = \mathbf{f}dcp(\mathbf{I}, \mathbf{A}) \tag{5.6}$$

DCP becomes the basis for two additional Priors by using inversions. The first additional Prior is known as Inverted DCP (IDCP) and also referred to as Colour Illumination Correction. The second additional Prior is known as Bright Channel Prior (BCP) or also known as Exposure Correction. IDCP functions by taking the original image and inverting it as first step. The assumption is that the image is white balanced and reflected as such in Equation 5.7. Next, it is dehazed and re-inverted back resulting from Equation 5.8 from DCP.

ii. Inverted Dark Channel Prior - Illumination Correction

We use the assumption is that the image is white balanced and this is reflected in Equation 5.7. In IDCP, the original image is first inverted, then dehazed before being re-inverted back again shown in Equation 5.8.

$$\mathbf{A} = (1, 1, 1) \tag{5.7}$$

$$\mathbf{J} = 1 - \mathbf{f}_{\text{dcp}}(\mathbf{1 - I}, \mathbf{A}) \tag{5.8}$$

iii.Bright Channel Prior - Exposure correction

In BCP, the minimum is replaced by maximum in Equation 5.3 in DCP resulting in Equation 5.9 below. Similar to DCP, the final image is available from Equation 5.10.

$$\tilde{\mathbf{t}}(x) = 1 - \max_{c} max_{\gamma \in \Omega R_I(x)} \frac{\mathbf{I}c(\mathbf{y})}{\mathbf{A}c} \tag{5.9}$$

$$\mathbf{t}(x) = guidedFilter(\mathbf{I}, \tilde{\mathbf{t}}(x)) \Rightarrow \mathbf{J} = \mathbf{f}_{\text{bcp}}(\mathbf{I}, \mathbf{A})) \tag{5.10}$$

Gaudio et al. [34] then proceeded to create a fourth Prior combining inversions of BCP and IDCP. With these 4 priors, 4 transmission maps $\mathbf{t}$ are

obtained. These maps are derived using maximum and minimum operations as shown below in Equations 5.11, 5.12.

$$\mathbf{t}(x) = 1 - \min_{c} \min_{\gamma \in \Omega R_I(x)} \frac{\mathbf{I}c(\mathbf{y})}{\mathbf{A}c} \equiv \max_{c} \max_{\gamma \in \Omega R_I(x)} \frac{1 - \mathbf{I}c(\mathbf{y})}{\mathbf{A}c} \tag{5.11}$$

$$\mathbf{t}(x) = 1 - \max_{c} \max_{\gamma \in \Omega R_I(x)} \frac{\mathbf{I}c(\mathbf{y})}{\mathbf{A}c} \equiv \min_{c} \min_{\gamma \in \Omega R_I(x)} \frac{1 - \mathbf{I}c(\mathbf{y})}{\mathbf{A}c} \tag{5.12}$$

Based on these identities, Gaudio derived functions which can be used to reveal weak and strong amplifications including those of dark and bright pixel neighbourhoods. The resulting identity in Equation 5.13, for a given Image $\mathbf{I}$, having transmission map $\mathbf{t}$ and atmosphere $\mathbf{A}$, leads to 4 image brightening methods and 4 image darkening methods.

$$\mathbf{J}(\mathbf{I}, \mathbf{t}, \mathbf{A}) = 1 - \mathbf{J}(\mathbf{I}, \mathbf{t}, 1 - \mathbf{A}) \tag{5.13}$$

These brightening and darkening methods are derived as functions based on solving for $\mathbf{t}$ using minimum or maximum values from identities earlier in 5.11, 5.12 using $\mathbf{I}$ or $\mathbf{I}$ - 1 and $\mathbf{A} \in$ (0,1). These combinations results in 2 groups amplification that is shown in Table 5.1. When $\mathbf{A}$=0, brightening methods are derived and when $\mathbf{A}$=1, darkening methods are derived from their transmission map $\mathbf{t}$. A,B,C, and D methods in PCA are referred to as brightening while W, X , Y, and Z are referred to as darkening. Note, $\mathbf{A}$ always is 1 when solving for $\mathbf{t}$.

Table 5.1: Derived Transmission Maps from t(x)=guidedFilter(**I**, $\tilde{t}$(x) [34]

Amplification	**Amplify Dark Areas**	**Amplify Bright Areas**
Weak	min t(1-**I**, **A=1**)	min t(**I**, **A=1**)
	1- max t(**I**, **A=1**)	1- max t(1 - **I**, **A=1**)
	Colour Illumination Prior	**Standard DCP**
Strong	1- min t(**I**, **A=1**)	1-min t(1 - **I**, **A=1**)
	max t(1 - **I**, **A=1**)	max t(**I**, **A=1**)
	PCA DCP	**BCP**

With these now understood, we focus on the PCA problem relating to ROP Retcam.

5.2.2 PCA - Illumination and over-saturation problem

PCA by Gaudio et al., [34] provide a toolkit (IEKT) which uses 4 letters A-D to brighten/and 4 letters W-Z to darken. Presently, PCA is called using an image and parameters which denote either brightening or darkening letters. The illumination of the input or that of the resulting output is not known. This defines the first problem of PCA as the outcome of each result is not predictable because the source image illumination is not considered in order to get the correct outcome. In most cases, the image is over illuminated or extremely dark. With a large dataset, there are numerous variations in the original image where some images are very dark with others are very bright. Predictability to select moderately amplified image is required for handling varying illumination of input ROP RetCam images. Therefore, we solved this problem of identifying the best illuminated result. The second

problem affecting PCA output is from over-saturation. For given ROP RetCam images, PCA tends to highlight the demarcation line as well as the vasculature. However, resulting PCA output images are challenged with over-saturation of reds in the center vision region as described by Dissopa et al.[19] which affects the use in Deep Learning. Our research also confirmed this to be a significant challenge with ROP RetCam images when using PCA methods. In order for PCA to be effective in ROP Retcam, it needs to firstly automatically select the correct method for illumination and then resolve the resulting over-saturation. We review the solutions to these problems next.

5.2.3 PCAr - Auto-Illumination

The research solved the first PCA illumination problem by identifying its best output image in terms of illumination. First, the transmission maps **t** are obtained for each of the 4 methods. These are named as A, B, C, and D for brightness, and W, X, Y, and Z for darkness. Brightness and darkness is selected by altering the value of atmosphere **A**. Atmosphere value of 0 is for brightness and 1 for darkness. Transmission **t** is solved using the identities in Equations 5.11 and 5.12. Final non-distorted image **J** is derived using Equation 5.13, using each of the 8 distinct transmission maps t. This results in 8 output images produced for each input image. When deriving transmission map **t**, the depthmap is normalised such that by using a median of 0.5. Then, calculated score value relative to 0.5 for each method is returned. In order not to over or under amplify, the image is centered and cropped to isolate region of interest map. This centre cropping is provided in IETK library. YCrCb [114] colour space was used for computing the depth map. This was chosen as

its Y component captures the luminance/brightness with Cr and Cb that note the colour differences for red and blue. Once all the solutions and their scores are obtained, record the scoring result with closest value to the median. The closest value to the median is the final selected balanced illuminated output as undistorted image **J**. The PCA Auto-Illumination algorithm we refer to as PCAr for ROP RetCam is described in Algorithm 3.

Algorithm 3 PCA Auto-Illumination (PCAr)

Input option for singleimage or compositeimage

Load image (imageX)

Crop and center the image for area of interest

Normalise depthmap for solving **t** for median of 0.5

Evaluate all brightening/darkening functions A-D, W-Z

Set Atmosphere **A**=0 for brightness

Generate solution of J for darkening methods and obtain median pixel offset score:

A= **J**(imageX,**A**,**t**(1-I))

B= **J**(imageX,**A**,**t**(I))

C= **J**(imageX,**A**,1-**t**(I))

D= **J**(imageX,**A**,1-**t**(1-I))

Set Atmosphere **A**=1 for darkness

Generate solution of J for darkening methods and obtain median pixel offset score:

W= **J**(imageX,**A**,**t**(1-I))

X= **J**(imageX,**A**,**t**(I))

Y= **J**(imageX,**A**,1-**t**(I))

Z= **J**(imageX,**A**,1-**t**(1-I))

USING scored images A-D, W-Z

IF singleimage **THEN**

pcarimage = image with lowest score

ELSIF compositeimage **THEN**

img1 = PCAr with lowest score in brightening group

img2 = PCAr with second highest score in brightening group

img3 = PCAr with second highest score in darkening group

pcarimage = img1 + img2, img3

ENDIF

Output pcarimage

This auto-illumination method avoids manually identifying the correct outcome which is unpredictable. To illustrate this problem, we demonstrate the intermediate results for 2 sample images. In Figures 5.1, 5.2 we show dark and bright RetCam image with output results using original PCA algorithm with our additional scoring added. Each output image also has a median offset score which is used to select the best outcome.

Figure 5.1: Dark Image McROP RetCam across PCA A-D, W-Z and their scores

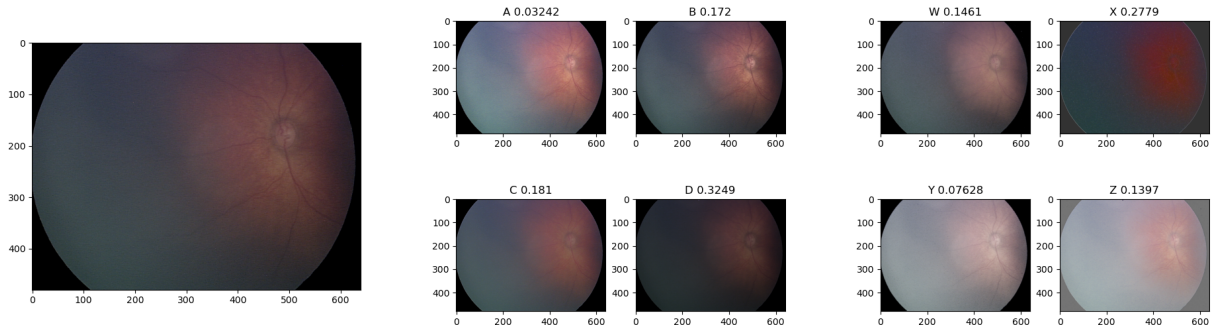
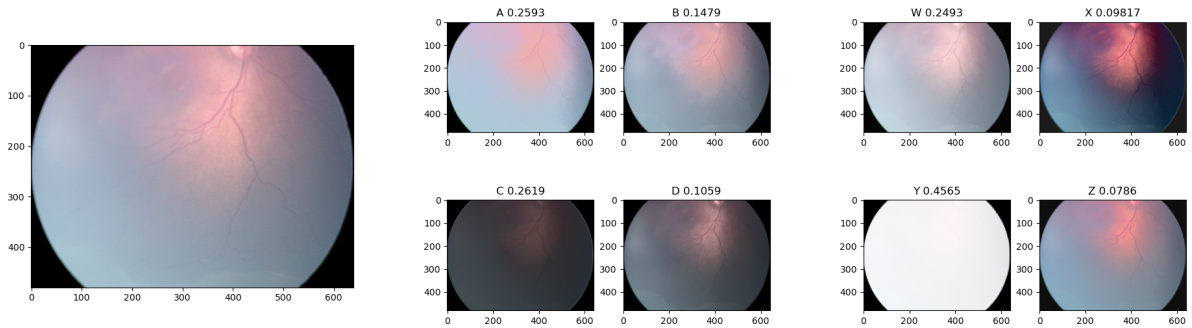
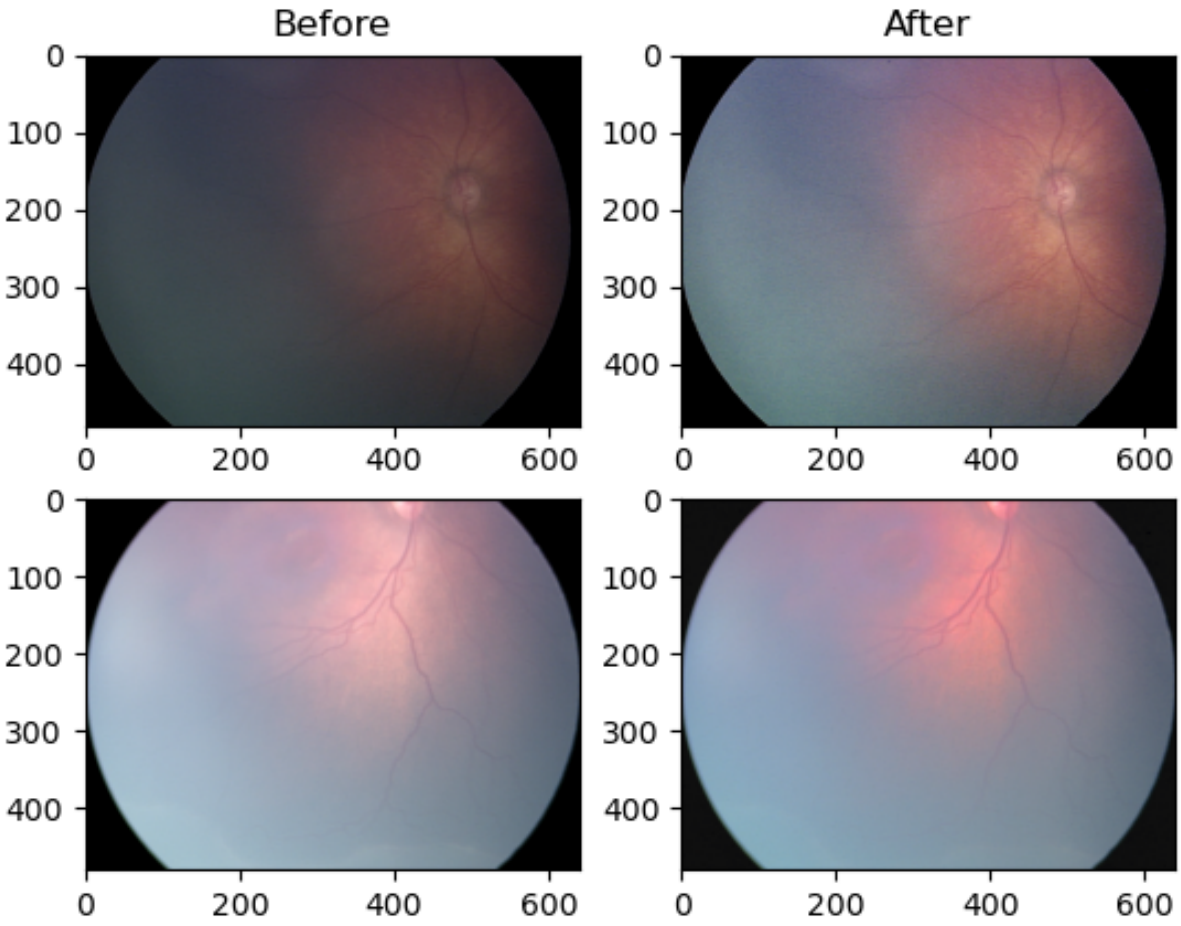


Figure 5.2: Bright Image McROP RetCam across PCA A-D, W-Z and their scores



The above examples show wide range of varying outputs for a dark and bright input image, This demonstrates PCA's challenge of manually identify a balanced output image relative to the input illumination variation. Using these examples, PCAr projects all 8 outcomes and uses its calculated scores to determine the image with closest score to median. This produces the most balanced illumination image. The final output example is shown in Figure 5.3. In the example, the original images where one is dark while other is bright, PCAr was able identify the correct output using the calculated offset from the median. It must be noted that PCA allows for multiple combinations of these to coexist. We further enhance the PCAr selection to cater for deriving a composite of 3 outputs only. As we have already generated 8 single outputs for scoring in PCAr, three of these are selected to create a composite image. This is performed by taking 2 outputs from brightening group that has lowest score closest to the median and third lowest from the median. The third output is taken from darkening method which is the highest value furthest from the medium. The merging of selected 3 images allowed us to get more ROP features into the image than by employing a single output.

Figure 5.3: PCAr final result for dark and bright RetCam image

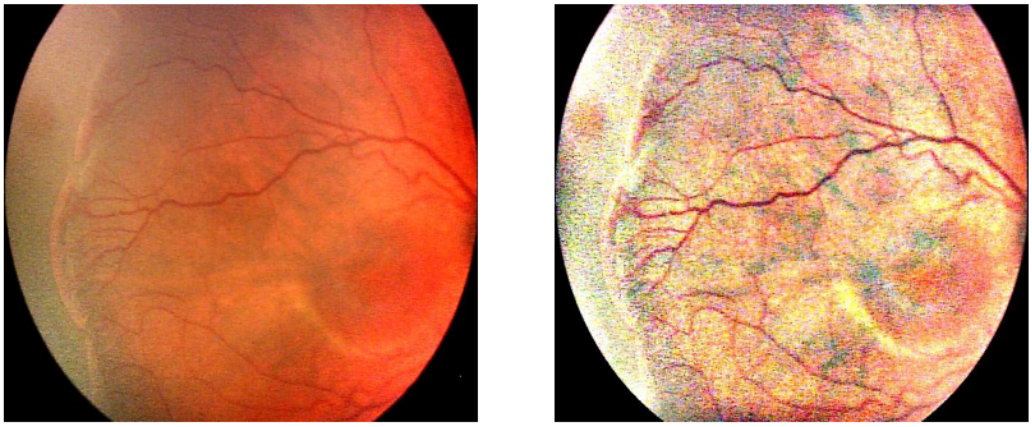


5.2.4 PCAr-CLAHE

With PCAr solving the auto-illumination problem and further extending to merge multiple best case outputs, we can use the output as an input into the CLAHE method which was described earlier. PCA output suffers from over saturation which can suppress key features necessary for ROP. This includes

the demarcation lines and blood vessels. By applying CLAHE to the PCAr merged image output, the resulting output showed significantly improved the overall contrast and colouration balance which negated the over-saturation problem associated with PCA. We refer to this as PCAr-CLAHE. The sample results of this hybrid method are shown in Figure 5.4, where PCAr-CLAHE ROP image is shown on the right and PCAr on the left. The resulting PCAr-CLAHE image shows clearly visible blood vessels and demarcation line as well as the macula while suppressing the high amount of colour that can degrade an CNN classifier. PCAr-CLAHE produces useful enhanced ROP RetCam image with clearly visible ROP features that can be used for classification.

Figure 5.4: PCAr merged image vs PCAr-CLAHE



5.2.5 Summary

To summarise, we examined DCP and 3 additionally derived Priors with deeper understanding of PCA and how our new method (PCAr) allows us to select an outcome which is more balanced irrespective of how dark or bright the original ROP RetCam image was. The resulting image can be further enhanced by

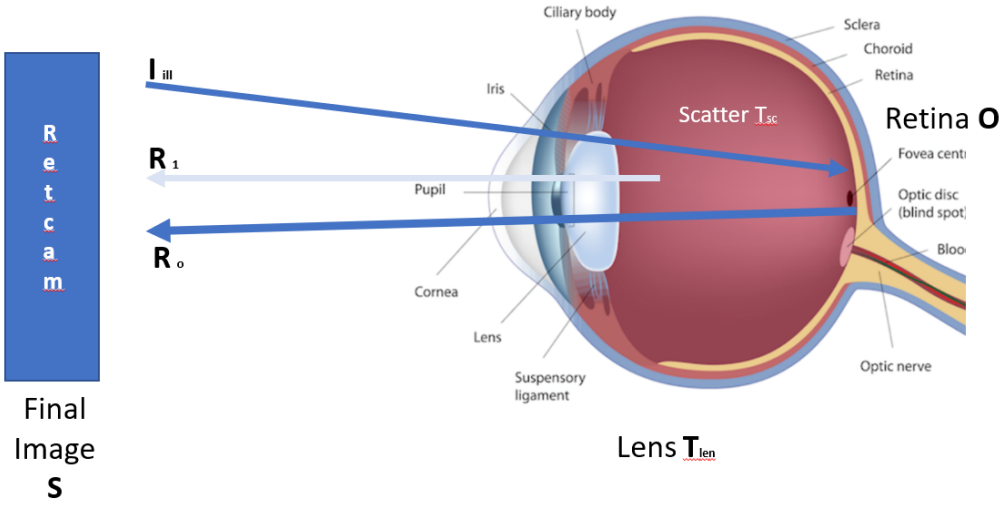
passing it to 3 channel based CLAHE (PCAr-CLAHE). PCAr-CLAHE reduces the over colour saturation as earlier reported by Dissopa et al, [19] as well as in our findings for ROP Retcam. This makes it useful for CNN in DR usage as well as in our case for ROP RetCam images.

5.3 DPFR

5.3.1 Preliminaries

Zhang et al., [145] introduced the DPFR method which answered a key question in this research's hypotheses. ROP RetCam image captured is subjected to a high degree of reflection from different objects as light enters the eye including the retina itself. DPFR took into account the image capturing mechanism which includes illumination, its forward journey into the eye and reflection. The transmission index was also estimated and not fixed as in the case of PCA. The final image, **S**, was computed using the sum of two back scattering components $\mathbf{R}o$, Equation 5.14, from the retina and $\mathbf{R}1$, Equation 5.15, from intra-ocular. $\mathbf{I}ill$ (r) represents the uneven illumination from the exterior of the eye, $\mathbf{T}lens$ represents filtering by the eye, and $\mathbf{T}sc$ interoccular scatter. It is noted that square terms in both equations denote the double interaction with scatter and lens. DPFR leveraged DCP to solve this. Rephrasing Zhang et al., [145] below, describes summarises the model.

Figure 5.5: Fundus Illumination and Reflection [145]



$$\mathbf{R}_o = \mathbf{I}_{ill}(r).[\mathbf{T}_{lens}(r).\mathbf{T}_{sc}(r)]^2.\mathbf{O}(r) \tag{5.14}$$

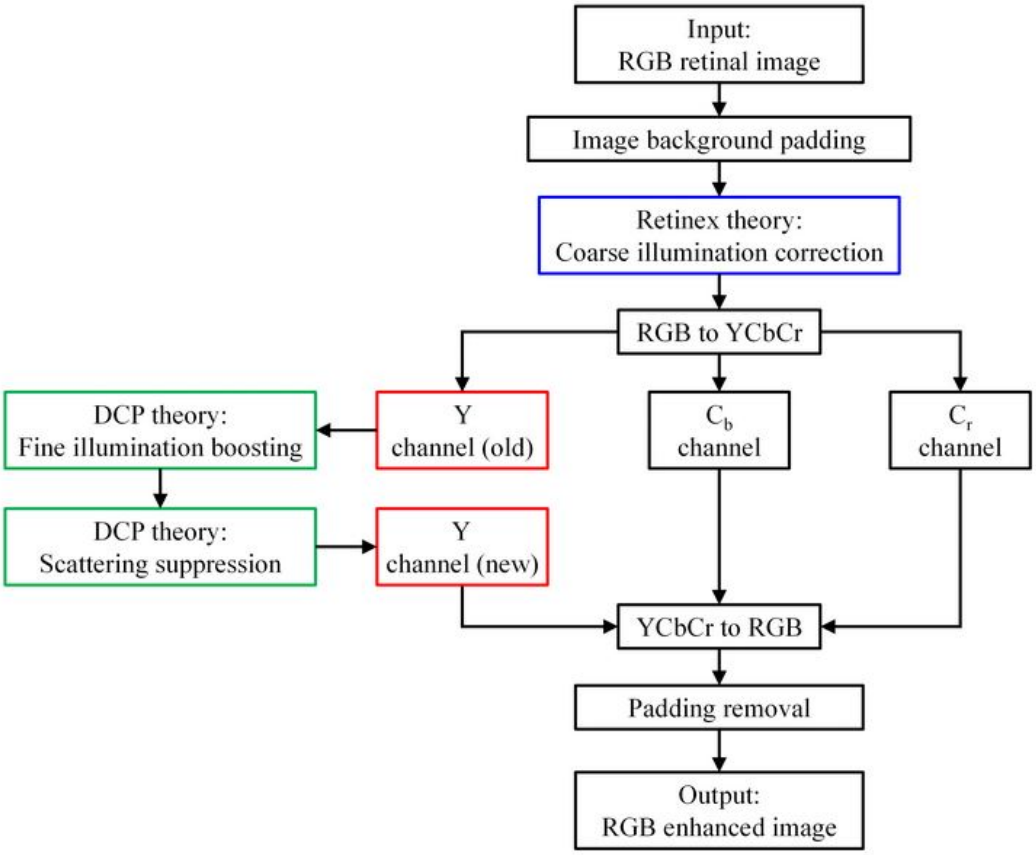
$$\mathbf{R}_1 = \mathbf{I}_{ill}(r).\mathbf{T}_{lens}{}^2(r)[1 - \mathbf{T}_{sc}(r)] \tag{5.15}$$

$$\mathbf{S}(r) = \mathbf{R}_o + \mathbf{R}_1 = \mathbf{I}_{ill}(r).\mathbf{T}^2{}_{lens}(r).[\,\mathbf{T}^2{}_{sc}(r).\mathbf{O}(r) + 1 - \mathbf{T}_{sc}(r)] \tag{5.16}$$

Restoration of image $\mathbf{O}$, namely the retina, is implied if illumination matrix $\mathbf{I}ill$ and transmission matrix $\mathbf{T}lens$ and $\mathbf{T}sc$ can be measured using DCP which is pre-known information. By estimating $\mathbf{I}ill$, $\mathbf{T}lens$ and $\mathbf{T}sc$ and solving Equation 5.16, an enhanced version of $\mathbf{O}$ is obtained.

The inverse problem is solved in four (4) steps: 1. Retinal image background padding, 2. Coarse illumination correction, 3. Fine Illumination boosting and, 4. Scatter suppression.

Figure 5.6: Double Pass Fundus Reflection [145]



In **Step 1**, fundus image circular exterior composed of a black region. This is padded to reduce over enhancement on the edges of the fundus image. **Step 2** involves the 3 channels (red, green, blue) preliminary illumination correction that removes fluctuations in the environment prior to hitting the retina. Equation 5.16, is now rewritten as follows:

$$\mathbf{S}(\mathbf{r}) = \mathbf{I}_{coarse}(\mathbf{r}).\mathbf{S}_o(\mathbf{r}) \tag{5.17}$$

with

$$\mathbf{S}_o(\mathbf{r}) = \mathbf{I}_{fine}(\mathbf{r}).\mathbf{S}_{ic}(\mathbf{r}) \tag{5.18}$$

and reflectance object[90] as:

$$\mathbf{S}ic(\mathbf{r}) = \mathbf{T}^2{}_{sc}.\mathbf{O}(\mathbf{r}) + 1 - \mathbf{T}_{sc}(\mathbf{r}) \tag{5.19}$$

The challenge of uneven illumination is resolved by using Retinex theory on illumination correction [47] by using logs on both sides of Equation 5.17 where log $(\mathbf{S})$ = log$(\mathbf{I}coarse)$ + log $(\mathbf{S}o)$. $\mathbf{I}coarse$ is generated by filtering $\mathbf{S}$ with a low pass filter noted in Equation 5.20. This yields Equation 5.19 which is applied to each channel followed by a colour correction. The ϵ value is of interest to us in Equation 5.21 that needs to be tuned for ROP RetCam image.

$$[\frac{1}{2\pi w^2}\exp(-\frac{\mathbf{r}^2}{2w^2})] \tag{5.20}$$

$$\mathbf{S}_0(\mathbf{r}) = \exp\big(\log(\mathbf{S}+\epsilon) - \log\mathbf{S} * [\frac{1}{2\pi w^2}\exp(-\frac{\mathbf{r}^2}{2w^2})] + \epsilon\big) \tag{5.21}$$

Step 3 and **Step 4** are combined in a unified model using DCP which was earlier discussed in PCA in Section 5.2.1. Fine illumination boosting converts the colour image to YCbCR maximising the Y channel. Y = (6.481 R + 128.553xG + 24.966*B + 16)/255) is used to provide maximum weight to the green channel with least to blue channel. This is used for grayscale DCP for dehazing. Here the estimation of reflection object $\mathbf{Y}_{\text{ic}}$ Equation 5.18, and $\mathbf{Y}_{\text{o}}$ is $\mathbf{Y}$ channel of $\mathbf{S}_{\text{o}}$ in YCbCR colour space.

$$\mathbf{Y}_{ic}(\mathbf{r}) = \mathbf{1} - \frac{\mathbf{1} - \mathbf{Y}_o(\mathbf{r}) + \mathbf{I}_{fine}(\mathbf{r}) - \mathbf{1}}{\mathbf{I}_{fine}(\mathbf{r})} \tag{5.22}$$

where $\mathbf{I}_{fine} \approx \mathbf{1} - \alpha.(\mathbf{1} - \mathbf{Y}_o)_{dark}$

Dark channel estimation is done using a local min filter. A value of 0.001 is passed for the filter. Lastly, scatter suppression is done to remove the haze effect of the images.

This method showed great promise when using adult fundus photos taken from a slit scope. A camera is visible from a distance and light is shown in. However, when the fundus image is that of a premature infant taken with a Retcam, there are significant physical/optical challenges if it is to be used for CNN based classifications. We now highlight the modifications to make it usable for ROP RetCam images.

5.3.2 Double Pass Fundus Reflection - grainy ROP RetCam problem

DPFR by Zhang et al., [145] provided an effective restoration which followed the optical truth to correct the elimination which PCA does not. This was accomplished by eliminating the reflective component by using a three steps process where a coarse illumination correction was done first, followed by fine illumination, and scatter suppression. The resulting image was very grainy when viewed closely for ROP RetCam use. Unlike adults, a pre-term and an infant have more choroid vessels visible due to age [101]. The existing DPFR filters caused a lot of deeper seated choroid vessels to show up. These vessels are a distraction when diagnosing Stages or Zones, and may be challenging for CNN diagnosis of Plus disease.

The biggest challenge was the identification of the Stage number in ROP RetCam images. The demarcation line is the critical factor for determining the Stage number. The ability to detect Stage 1 compared Stage 0 was significantly

impacted as the demarcation line was very faint. A standard DPFR processed image was very grainy, and this caused this line to vanish at times which prevented a diagnosis. Secondly, the number of choroid vessels showing up on DPFR were an additional burden, since they present a colour factor that poses a challenge with any CNN. An improvement to resolve both these aspects was key for enabling DPFR to work well with ROP RetCam images such that the resulting image could be used with CNNs for Stage or Zone detection but also can be helpful with Plus disease as well.

5.3.3 Double Pass Fundus Reflection - for ROP RetCam Contribution

In DPFR by Zhang et al., [145] two challenges were noted earlier. To remediate these, we identified 3 tuning improvements to reduce the graininess namely in the guided filter and dehazing estimation in D. We refer to the tuned method as DPFRr, with r referencing ROP Retcam. The changes reduced the level of choroid vessels which became less visible thereby making this method suitable for ROP Retcam. Once these improvements were done, the colouration was further reduced by passing the DPRFr resulting image into the 3 channel CLAHE process described earlier. The latter, further reduced the red colour saturation factor.

The first changes made were in the parameters of choice in the following areas, suggested by studying the above methods:

1. **Coarse Illumination Correction**

- Low pass filter value of ϵ was changed from 0.01 to 0.001 in Equation 20. Other values were also tried but this appeared to be most effective.

2. **Dehazing estimate target**

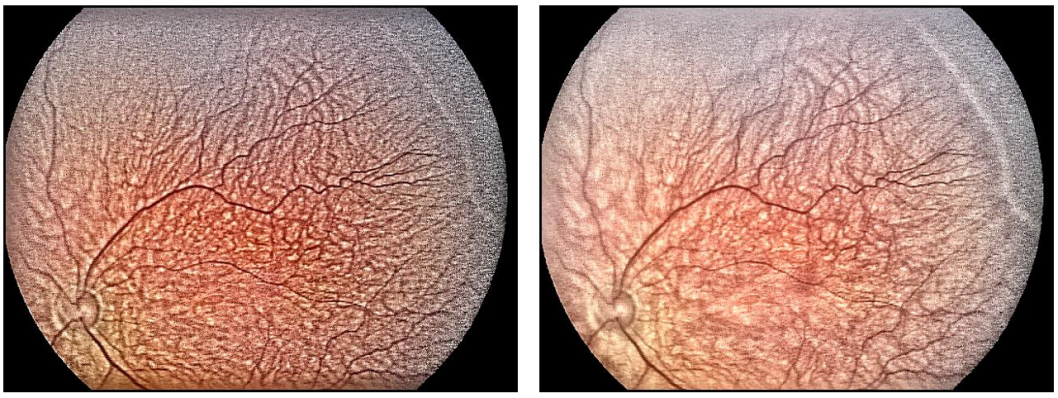
- In supporting Eq. 20, the estimate for dehazing was altered from 0.1 to 0.2 to reduce the level of correction for choroidal vessels. Other finer values were also tried but this value appeared to provide most balanced output when combined with other changes.

3. **Fine illumination boosting**

- In Equation 20, for the Dark channel estimation local min filter, the value is altered from 0.001 to 0.000000001 to allow a finer selection. This value was derived after repeating the changes from 0.001 to 0.0001, 0.00001 until we reached 0.000000001. This significantly reduced the grainy pixel nature of the DPFR by making the pixel cluster area smaller. The smaller cluster enables better identification of the demarcation line.

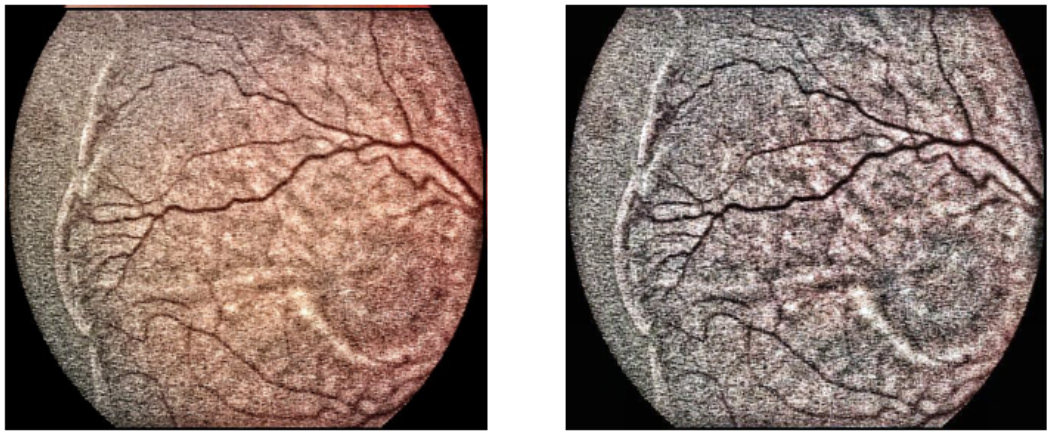
The resulting corrections from these three changes noted above, resulted in significant improvements allowed the demarcation line to be better visible, reduced choroid vessels noise, and achieved a very low grain pixelisation effect when compared with the original original DPFR. Figure 5.7 shows the improvements between standard DPFR (left) and ROP tuned DPFRr (right), which was then used for the rest of this project's experiments for Plus disease, Stages and Zones.

Figure 5.7: DPFR vs DPFRr for ROP RetCam Images



With this, we defined a tuned DPRFr which was ideally suited for ROP RetCam fundus images. The changes reduced the level of visible choroid vessels and reduced the granular pixels to a smaller amount. The output image still had very high illumination and colour factor. By using DPFRr output as input to a 3 channel CLAHE approach, the DPFRr output image's contrast was reduced considerably yielding the presence and features of the demarcation line even better. This removed any additional colouration which can influence the classifier while making the demarcation line, main blood vessels and optic disc clearer. We refer to this hybrid as DPFRr-CLAHE as our eighth method. Figure 5.8 shows further improvements between DPFr and DPFRr-CLAHE with less colour and improved demarcation line, blood vessels and optic disc.

Figure 5.8: DPFRr to DPFRr-CLAHE for ROP RetCam Images



5.3.4 Summary

To summarise, we examined DPFR and its core derivation logic and fine-tuned key parameters in three core functional areas to allow it to be more suited for ROP RetCam images. DPFRr further was fed into CLAHE and now called DPFRr-CLAHE. It provided for a better demarcation line and vessels outline. Using this we have a comparative analysis available for DPFRr and DPFRr-CLAHE as two distinct pre-processing techniques for ROP RetCam images for CNN classifiers primarily for Stages and Zones but can also be used for Plus disease.

In summary, standard image domain methods and extended image restoration for ROP RetCam images, each provides independent salient features on their own. These can be more powerful when combined, as in the case of PCAr-CLAHE and DPFRr-CLAHE. The use of these restoration methods namely PCAr, DPFRr and combinations with CLAHE are the first known attempts aimed at improving ROP RetCam images specifically.

5.4 LWNET Segmentation for Vessel erosion

After resolving the image pre-processing using the newly introduced methods namely PCAr, PCAr-CLAHE, DPFRr and DPFRr-CLAHE, we further focused on further reducing the blood vessels in the ROP RetCam image such that only demarcation line is available on the image. We considered two additional approaches. The first approach was to denoise the entire image resulting in lower intensity of the vessels in the ROP RetCam image. The alternate approach was the erosion of blood vessels in such a way that the background is blended into the surrounding without any degradation to the intensity of the demarcation line. We discuss and highlight these novel approaches next.

5.4.1 Preliminaries

In preparing the RetCam fundus image for Stages detection, we now pay particular attention to whether there is a demarcation line or not. If the line is present, its magnitude can be closest to the fovea (center vision) - Zone I, or further out towards the periphery in Zone III. The idea conceptualizes the elimination of retinal vascularity as a potential variable from the classifier model. The purpose is to help the respective classifier consider only the rest of the image where the demarcation area may be present, thereby classifying the Stage. In other words, reduce the noise from other aspects of the fundus image such as blood vessels.

5.4.2 Erosion using Wavelets

This research first focused on how to accentuate the demarcation line while reducing all other vasculature. Consideration was first given to use wavelets [40] as an attempt to denoise and then follow it with a guided filter. Wavelets provide a good method of approximation at differing orders that preserve critical information. This is a viable alternative to thresholding noted earlier in Section 3.1. Yulong et al., [143] note the wavelets role in signal processing to remove noise in a captured image to determine damage to thin plates in engineering. Similarly, we considered the retinal image as a flat 2 dimensional surface and applied wavelets to denoise ROP RetCam images. The algorithm created is described as:

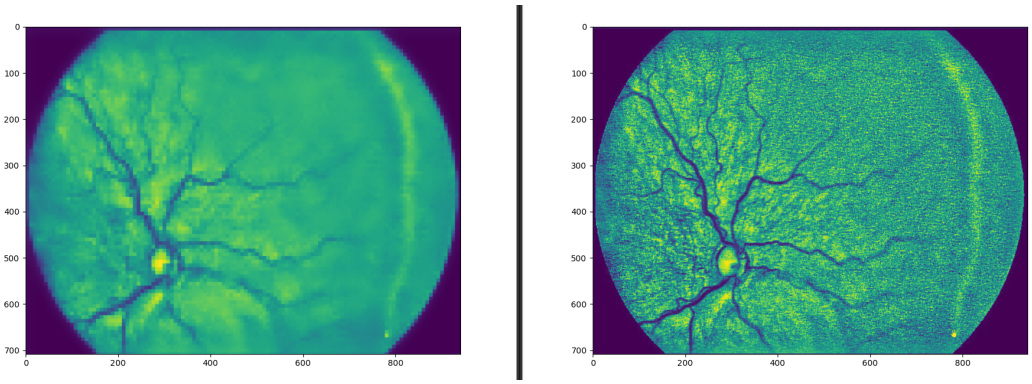
Algorithm 4 Wavelet denoising and smoothening

1: **function** WAVELETDENOISE($imageX$):
2: grayimg = convert image($imageX$) to Grayscale.
3: waveimg = wavedec2(grayimg)
4: denoiseimg = denoise(waveimg) using a threshold value
5: fineimg = guidedfilter(denoiseimg)
6: **return** combinedchannelimg(r,g,b)

We used pywt.wavedec2 [17] method and CV2 library's guidedFilter [84] for this proof of concept. Wavedec2 allowed the derivation of a 2D multi-level decomposition of the image. Denoising was then performed by using 1st, 2nd and 3rd coefficients of the decomposition. The result of each was used to re-constitute the image. Then a guided filter was applied. This acts as an edge preserving smoothing filter that filtered out the noise and retained the edges. The results using one of McROP RetCam Stage ROP images is shown

in Figure 5.9, with the left image as denoised versus pre-denoising on the right:

Figure 5.9: Using wavelets to reduce vasculature noise



The resulting data showed promise by removing most of the retinal background such as fovea and smaller vessels. However the blood vessels still remained prominent. Other research showed that wavelets have been used to train CNNs for other diagnoses such as echocardiogram [89]. We propose future work where wavelets can be used to detect the presence of the demarcation line presence, by using CNNs, and possibly isolate and measure the intensity as a measure for Stages.

5.4.3 Vessel erosion using LWNET Segmentation

As previously noted, retinal blood vessel presence was highlighted as being a contributor of noise/additional factor during the training for Stages. Reduction of the vessels using wavelets also negatively impacts the intensity of the demarcation line which can be erased, if in an early stage. Therefore, consideration was given to completely removing them from the image as if they

were absent. This may assist in the overall detection and concentration on the demarcation line which determines the ROP Stage.

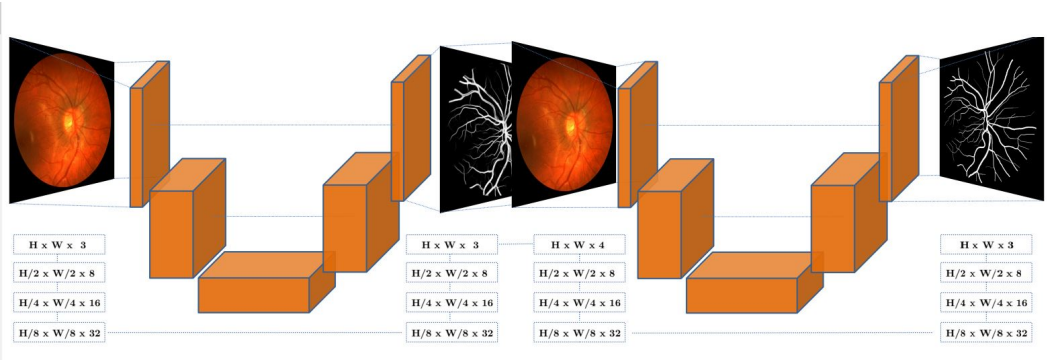
5.4.3.1 Generating a vessel map using LWNET for ROP RetCam images

We first demonstrated a way of doing this by using a vessel segmentation generated map. To perform this, CNN based vessel segmentation was a critical computational requirement for an ROP RetCam image. The generated segment vessel map for this project served the task of a vessel map. For this, the choice was to use a minimal weight established CNN based U-Net[104] segmentation. Using this, there was an existing precedence and allowed the project to shortcut the research and focus on the problem area. We chose LWNET [31] which is a very minimal model segmentation design. U-Net and its adaptations are better performers than those achieved via mathematical morphology where the idea of performing a type of mathematical transformation to accentuate vessel profiles and thereafter threshold the output resulting in a segmentation.

LWNET uses the U-Net architecture twice (Figure 5.10); referred to as W-Net ϕ with input image x. The first U-Net result ψ is concatenated to image x. The concatenated image is then passed through a second U-Net which can be denoted in Equation 5.23. In other words, the first U-Net generates prediction of vessel localization. This is used for focusing into the areas of interest by the second U-Net.

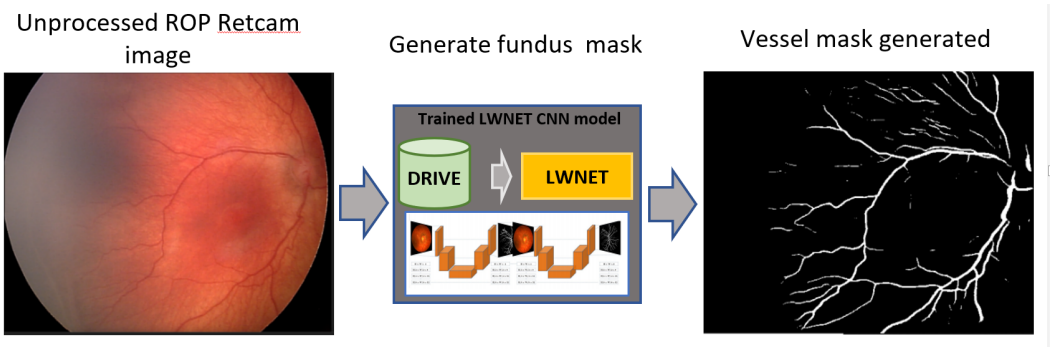
$$\phi(x) = (\phi^2(x), (\phi^1(x))). \tag{5.23}$$

Figure 5.10: LWNET Double U-Net [31]



As the intention was to use it for ROP Retcam, we did not have ground truth manually annotated vessel maps where the segmentation map was generated. Therefore, the assumption made was to use the adult based DRIVE public dataset to train LWNET for vessel generation. The DRIVE dataset has fundus images of healthy adult population with hand segmentation maps for validation. DRIVE fundus images have excellent resolution quality and contrast. Once the LWNET was trained with DRIVE, the model was ready to generate predictions for vessels. The McROP ROP RetCam Stages datasets (training/validation) were passed through LWNET. This generated a vessel map for each individual image using the same filename in a separate folder. This overall flow is described in Figure 5.11.

Figure 5.11: ROP RetCam vessel map generation using LWNET



Using this approach, we successfully used DRIVE to train the LWNET CNN model, and then used the model to generate vessel maps from ROP RetCam images. In other words, we used an open source retinal images with open ground truth mapped vessels for training and then successfully using it with ROP RetCam pediatric images. This was used to generated segmentation maps for the training and validation datasets for Stages. This is first known use of any minimalist CNN based segmentation for ROP RetCam images.

5.4.3.2 Erosion Average of Blood vessels using vessel map

As noted before, the area of interest remains on removing blood vessels so that noise could be reduced further to better enable classification for Stages. As shown before, large blood vessels remain prominent after attempting to denoise using wavelets. Hence the next step was to understand how to remove these. All work to date focused on extracting these blood vessels but here the challenge is how to erode the vessels but preserve the background so that removal of vessels was not visible. We have now demonstrated that segemen-

tation can be used to remove distracting elements while retaining background information.

This was accomplished by feeding LWNET [31] with Stage ROP Datasets as noted in the prior section 5.4.3.1. We then developed a new method demonstrated that took the segmented vessel image mask generated by LWNET and its original ROP RetCam image and removed vessels based on the generated mask.

The principle of erosion of blood vessels can be described as taking the image x and using the segmented vessel map ϕ (x) as a mask. We used this erosion method applied to each of the three channels, namely red, green, blue channels of image x. Since an ROP RetCam image is in colour (RGB) and it is therefore important that all 3 channels are addressed. Convolution of the image is performed using a wider 32 x 32 array that provides a blur box. This is then flattened together with a channel and mask. In our method, we identify a vein's presence in mask containing vessel segmented map by pixel value greater than 0.1. These are the selected pixels which are assigned either an average value or Gaussian value of the surrounding area; the others are left with prior existing values. The resulting reshaped shape is then convolved further in the 2d space using 16 x 16 array with the same subsequent steps; repeating for 8 x 8 and 4 x 4 respectively. The following algorithm describes the process:

Algorithm 5 Erosion using Segment masking either by Average or Gaussian method

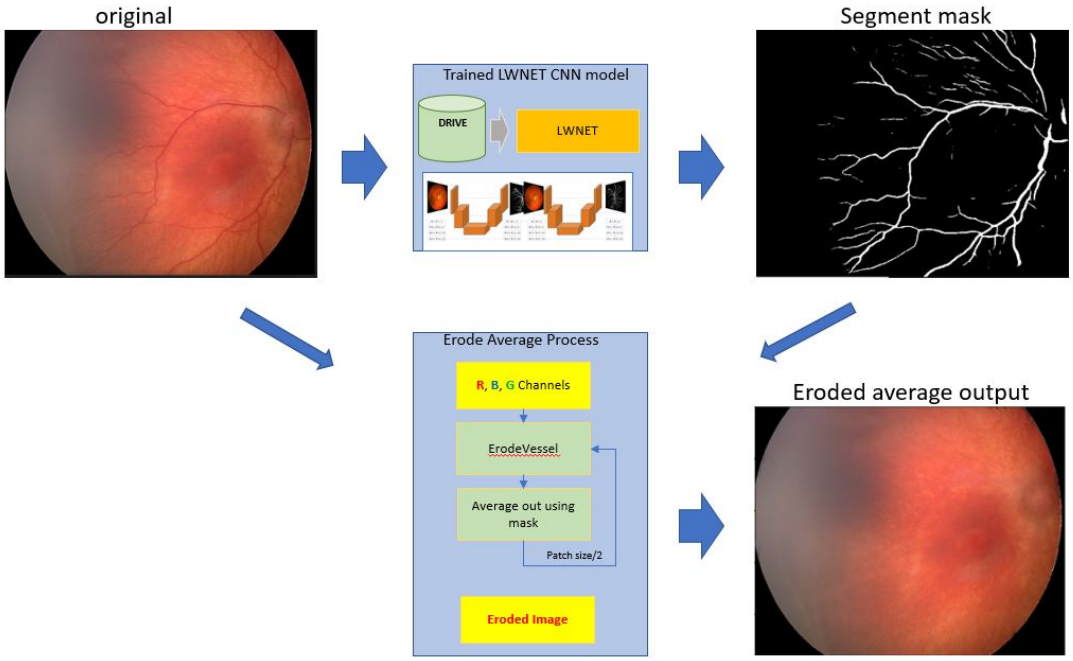
```
function CLEANIMAGE(imageX, maskY)
    split into 3 channels and employ gaussian to the vessel using mask
    r = blendvessel(imageX(r channel), maskY)
    g = blendvessel(imageX(g channel), maskY)
    b = blendvessel(imageX(b channel), maskY)
    return stackedimage(r,g,b)
function CHOOSESIDE(boxareaV, channelV, maskV)
    boxarea = boxareaV.flatten()
    channel = channelV.flatten()
    mask = maskV.flatten()
    veins = (mask >0.1)
    retval = np.zeros(boxarea.shape)
    assign gaussian to selected pixels
    retval[veins] = boxarea[veins]
     keep remaining ones same as original
    retval[∼veins] = channel[∼veins]
    return retval.reshape(channelv.shape)
function BLENDVESSEL(channel, mask)
    – removes vessel from each channel using gaussian
    patchsize = 32
    loop while patchsize ≥ 2
     blursqr = convolve2d(channel, np.one((patchsize,patchsize)) using wrap)
     if guassianopt then blursqr = gaussian(blursqr)
     else blursqr = average(blursqr)
     end if
     curedchannel = chooseside(blursqr, channel, mask)
     channel = curedchannel
     patchsize = patchsize/2
    end loop
    return curedchannel
```

Using this model, the eroded fundus image left a blended replacement which started from a rough map and repeatedly narrows the patch size thereby blending in the averaging colouration. Additionally by doing this across all the channels, this allowed any other image processing to take advantage of the newly formed image. Although not perfect, results showed promise, but the method needs to be further improved if it is to have more pronounced effect.

We believe, this is the first known attempt of erasure of blood vessels in the fundus image using this approach.

Below Figure 5.12, illustrates the overall method. Using the ROP RetCam image, this is fed into the LWNET [31] resulting in a segmentation vessel map. Both original and segmentation map are fed into the curing process which results in the eroded average ROP RetCam fundus output image.

Figure 5.12: Fundus segmentation and Erasure



5.4.4 Summary

To summarise this section, this problem was identified as an approach to erase retinal vessels from the original image while preserving the remaining ROP clinical features. The intent was that the removal of blood vessels can aid a CNN focus on other aspects of ROP RetCam image e.g. the demarcation

line if present. Additionally, an alternate approach using wavelets was also studied to attenuate the blood vessels signal. The segmentation approach to erosion demonstrated a unique way to generate a segmentation map using an existing technique which was trained on adult eyes but then re-used in paediatric ROP RetCam images with success even though there are significant differences between the biological structures. While the CNN success using this method was not very conclusive, the segmentation based erosion does have potential use in a clinical setting where some features may need to be eroded. This can also be applied to other challenges where other segmented areas need to be merged into the background.

5.5 Resizing evaluation for ROP

5.5.1 Preliminaries

Resizing implies image transformation [83] and OpenCV[84] which is widely used provides a resize function with several choice of filters such as nearest, bilinear, bicubic and Lanczos. This implies transforming a larger image into a smaller sized image. Resizing is a critical standard component in Deep Learning as larger size images require longer time to train on as opposed to a smaller sized image. For CNN use, a larger sized image needs to be downsized or resized. The assumption for resizing is that there must be no loss of critical information. In our case, for ROP RetCam based Stages/Zones classification, it is imperative that there is no significant loss of relevant information in all related datasets e.g. training, validation and/or testing datasets.

5.5.2 Resizing for ROP RetCam images

In this work, we resized ROP RetCam images from 640x480 to 224x224 as optimal size. This also implies that the aspect ratio changes from 16:9 to a square. This optimal size is recommended by Keras as it will be used in our keras based CNNs [57]. It is essential to understand which resizing methods will work for ROP RetCam based on the challenges reported by Venturelli M. [128]. To date, there was no mention of this challenge in any ROP or Diabetic Retinopathy papers employing CNNs.

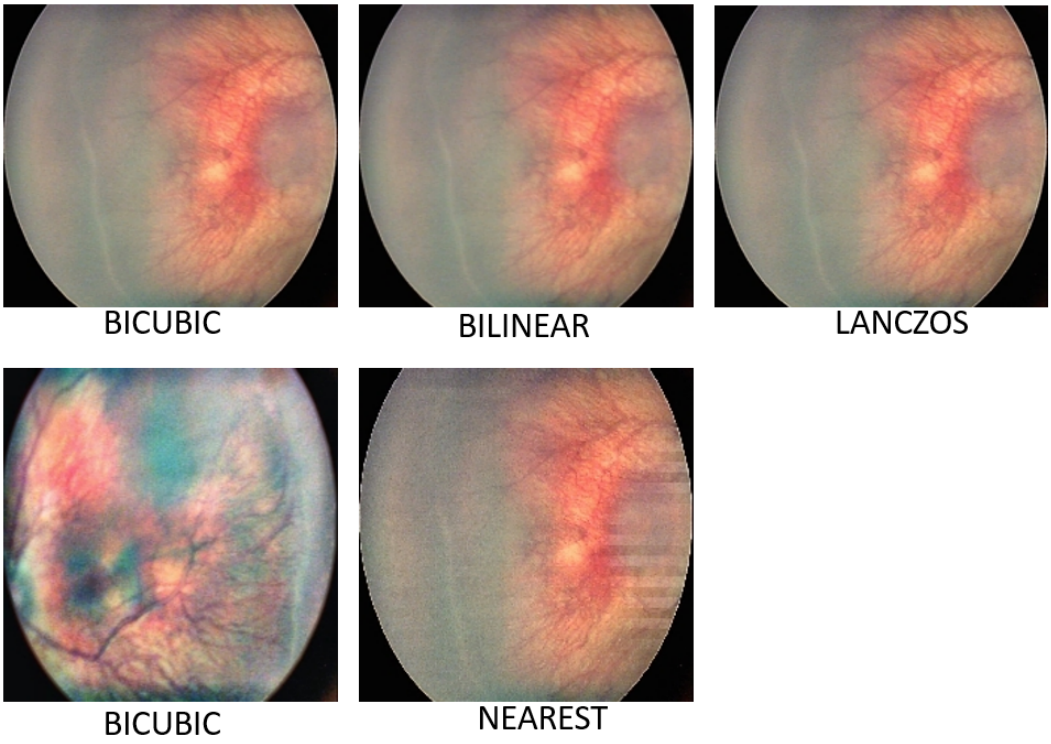
We reviewed each resizing method's output for loss of data and effectiveness in retaining information such as boundaries, texture, edges as well as lack of artifacts from original information such as vessels, Optic Disc, macula, demarcation line (if present). The best choice observed anecdotally was Lanczos filter based on these criteria. It uses a truncated sinc filter applied on an 8x8 neighbourhood. The other second best method was the bicubic method. Other methods such as bilinear, and nearest did not do well as waves / ripple lines were noted on the other methods.

We resized the pipelined images from each of pre-processed methods as noted in section to 224x224 pixels RGB images using OpenCV resize with lanczos method. Figure 5.13 shows the differences between these in a few samples used for Stage resizing to illustrate the differences. In the cases of bicubic method, it can give false colours from time to time, nearest method induces wave patterns at times where bilinear showed some details in lower intensity than Lanczos. With Lanczos resizing, we observed an additional benefit of smoothing effect that also further reduced the pixelisation effect for DPFRr.

The most rigorous testing approach will be to test each method and run it against each classifier. This is left as a future publication to test it in conjunction with Transfer Learning models.

In our preliminary research, prior literature reviews [99] noted the use of Haar wavelet for resizing/compression. This was our initial method for resizing in order to maintain image details while compressing the image. Our findings with Haar wavelet compression did not yield any improvement in excess of 1%. As Lanczos method resized the image better, therefore, Haar was no longer pursued as an option for resizing.

Figure 5.13: Resizing Examples -Lanzos, Bicubic, Bilinear, Nearest



5.5.3 Summary

To summarise this section, resizing is a critical component bridge between the image pre-processor and the CNN. Resizing enables faster CNN processing as opposed to a larger image and therefore is required pre-CNN usage. Therefore, it is essential for the smaller image to preserve key features in relation to the original image. This ensures there is no loss of data which can adversely affect the results in a CNN. Amongst the various resizing methods, we identified one method, namely Lanczos. This identified method worked well for ROP RetCam pre-processed images. With this complete, we have completed all primary pre-processing and resizing steps necessary in order to now focus on CNNs to classify the pathologies.

5.6 CNN Classifiers for Plus disease, Stages, Zones using - Transfer learning

As discussed earlier, lack of data and prior learning leads this research into leveraging transfer learning models. We described the transfer learning concept earlier in Section 3.3.3 and reviewed leading deep convolutional network for object recognition, including Oxford's Visual Geometry Group (VGG)[26], Inceptionv3, ResNet50 [58], InceptionResv2 [56] and Dense169 [54], as these had been pretrained using ImageNet and provided by Keras [55]. These networks are ideally suited as they are pre-optimised for identity mapping during the training process. We chose to use ResNet50 [58], InceptionResv2 [56], and Dense169 [54] which had been pre-trained using ImageNet and provided by Keras [55]. It should be noted that VGG16, VGG19 and Inceptionv3 transfer

learning models were developed along similar lines and rigorously analysed. They did not provide adequate classification results in the most recent upgraded libraries versus earlier versions.

5.6.1 CNN Transfer learning model

Using the standard methods as prescribed for transfer learning CNNs in all cases, we do not want the last fully connected layer. Algorithm 6 describes the approach used to leverage an existing model.

Algorithm 6 Using Transfer Learning for hybrid model

1: Freeze weights of each model to false - no updates
2: Add custom block of CNN layers on top of each model
3: Commence learning using local data

The algorithm's steps summarise the steps where the last layer was replaced by the task at hand e.g., Plus disease, Stage and Zone. We froze the weights of each model to false to stop updates to pretrained weights as we wanted to preserve knowledge retained earlier using the ImageNet dataset. We added a custom block of CNN layers on top of each one of these models to build a hybrid model.

Transfer Learning [43] has been successfully used to improve classifiers [110, 45, 32] and is critical as the number of ROP images was limited. It is helpful when there is not sufficient data to train a full model as it prevents overfitting and aids in the training process. The shape of each pretrained model's input layer and final dense layer is modified. Weights are additionally calculated based on the number of training images by class to prevent overfitting. This included L1 regularizer and Dropout which were also employed to prevent

overfitting. L1 and L2 regularizers are effective in reducing the variance when the training does not generalise. This is helpful in cases of imbalanced datasets. GradCam[39] was also used to check for features being considered.

5.6.2 Transfer Learning CNNs for Plus disease, Stages, Zones

For Plus disease, Stages, and Zones, in ResNet50, the following layers were frozen and additional ones added with softmax given below:

Algorithm 7 ResNet50 Transfer Learning model - Plus disease

1: Freeze all layers except: ‘res5c_branch2b‘, ‘res5c_branch2c‘, ‘activation_97‘

2: Add fully connected layers:

GlobalMaxPooling2D

Dropout(0.6)

Flatten

Dense(1024,‘relu‘,l1)

Dropout(0.6)

Dense(1024,‘relu‘,l1)

Dropout(0.6)

Batchnormalization

3: Final classifier prediction using softmax activation function

Algorithm 8 ResNet50 Transfer Learning model - Stage, Zones

1: Freeze all layers except: 'res5c_branch2b', 'res5c_branch2c', 'activation_97'

2: Add fully connected layers:

GlobalMaxPooling2D

Dense(1024,'relu',l1)

Dropout(0.2)

Flatten

Dense(1024,'relu',l1)

Dropout(0.2)

Batchnormalization

3: Final classifier prediction using softmax activation function

Following similar practice, we extended InceptionResv2 for Plus disease and Stages in the following summary which was slightly different and Zones in subsequent definition.

Algorithm 9 InceptionResv2 Transfer Learning model - Plus disease, Stages

1: Freeze all layers except: 'block8_10_mixed' that is trainable

2: Add fully connected layers:

Flatten

Dense(1024, 'relu')

Dropout(x) where x=0.6 for Plus, x=0.2 for Stage

Dense(1024,'relu')

Dropout(x) where x=0.6 for Plus, x=0.2 for Stage

Batchnormalization

3: Final classifier prediction using softmax activation function

Algorithm 10 InceptionResv2 Transfer Learning model - Zones

1: Freeze all layers except: ‘block8_10_mixed‘ that is trainable

2: Add fully connected layers:

Flatten

BatchNormalisation

Dropout(x) where x= 0.2 for Zones

Dense(1024,‘relu‘,l1)

Dropout

Dense(1024, ‘relu‘,l1)

Dropout(x) where x=0.7 for Stage and 0.2 for Zones

Batchnormalization

3: Final classifier prediction using softmax activation function

Lastly, Dense169 was used for Stages to cross validate results between ResNet50 and InceptionResv2 for Stages.

Algorithm 11 Dense169 Transfer Learning model - Stages

1: Freeze all layers except: ’block8_10_mixed’ that is trainable

2: Add fully connected layers:

GlobalMaxPooling2D()

Dense(1024, ’relu’, ’l1’)

Dropout(x) where x=0.2

Dense(1024,’relu’, ’l1’)

Dropout(x) where x=0.2

Batchnormalization

3: Final classifier prediction using softmax activation function

A standardized set of models were prepared for Plus disease, Stage and Zone. To miminize loss, a self-adjusting training rate was added with start rate of 1e-5 which is called by LearningRateScheduler, batch size of 32 for minibatch optimisation, epochs of 10, Adam optimizer, patience of 5 early stop Inference rate was set at standard 0.5. Cross-entropy loss was also used in the training process.

To summarise, the CNNs described were able to be trained with ease using augmented datasets for Plus, Stages, and Zones. These formed the final portion of the McROP architecture.

5.7 Chapter Summary

Retinopathy of Prematurity (ROP) poses a significant challenge within prenatal care system and potentially long term implications in terms of sight threatening conditions to the infants if not addressed early. ROP is a potentially life altering disease which requires early detection and intervention. An accurate diagnosis which is able to grade the severity various aspects of ROP can assist in its early detection. The research contribution in this chapter gives a complete set of processes which contribute towards the creation of a base ROP diagnosis system. This uses ROP RetCam images and can be used by clinicians to either trial diagnose ROP at an early stage with a high degree of confidence in terms of accuracy and safety or even use the image pre-processing in a clinical mode in Telehealth.

This research studied and contributed by solving a critical problem with restoration based image pre-processing method namely PCA by resolving its illumination problem and then creating a composite using a prescribed set

of scores. Next, the second restoration based image pre-processing method, DPFR was fine tuned to make it suitable specifically for Stages and Zones when using ROP RetCam images. The results of both pre-processing methods were combined with a 3 channel based CLAHE method. The hybrid provided significantly improved ROP specific features. These features are suitable for later use in transfer learning based CNNs for Plus disease, Stages and Zones. Two main comparative deep learning processes namely CNN InceptionResv2 and ResNet50 were extended with ROP specific layers for Plus, Stages and Zones. A new method to erase retinal vessels from the original image while preserving the remaining features was described using a minimalist U-Net based segmentation. An efficient re-sizing method was identified suitable for these pre-processed images which kept intact retinal data such as vessels, demarcation lines.

All components together combined completed the creation of McROP diagnosis system and we next examine the experiments and results.

Chapter 6

Experiments Results Analysis and Discussion

The focus of this research was to improve image pre-processing methods to improve the clinical features visibility in an ROP RetCam image. These pre-processed images are to be used in CNNs resulting in improved classification of Plus disease, Stages, and Zones. In doing so, this would create an overall process that contributes to improved classification. Our work focused on the quality of ROP RetCam images that are key inputs for the processes which are outlined in this section.

This chapter presents the result analysis of the experiments using pre-processed images resulting from McROP datasets and CNNs that classify the various conditions. Section 6.1 describes the high level flow of the experiments. Section 6.2 describes the overall configuration of the system used and datasets used to train CNN classifiers. Section 6.3 reviews the results of the CNNs for Plus disease, Stages and Zones. It details the confusion matrix, and other

measurements for further discussion and comparison with relevant reviewed papers. Section 6.4 summarises this chapter.

6.1 Experiments

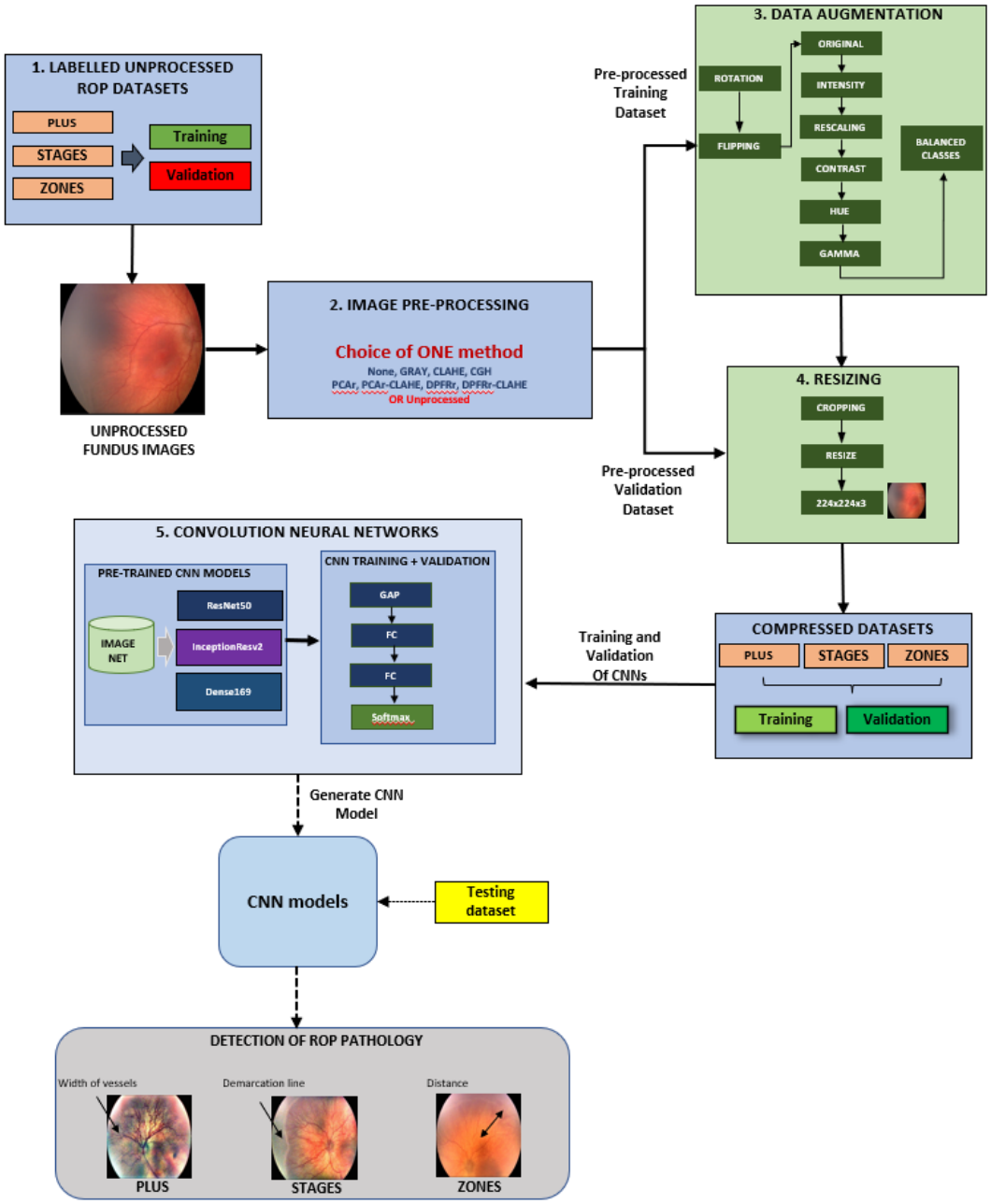
We describe in this section the overall process in terms of configuration and experiments which were executed as part of McROP architecture. High level process flow steps include using the labelled pair of training/validation datasets and a choice of image pre-processing method. For the training dataset, augmentation was applied as described earlier in Section 4.3. Training/validation datasets were then resized prior to training/validation in the selected CNNs. The algorithm is iterative and the steps in each iteration are described in algorithm 12. Each iteration applies to a specific pre-processing method including one iteration that does not include any pre-processed dataset pair.

Algorithm 12 Steps for each iteration

1: Define experiment for Plus disease, Stages, or Zones.
2: Choose correct pair of labelled unprocessed dataset for experiment
3: **LOOP** For M = pre-processing methods incl. unprocessed **LOOP**
4: **IF** m=pre-processing method **THEN**
5: Generate pre-processed dataset pair using identified method M
6: **ELSE**
7: Use unprocessed dataset pair.
8: **END**
9: Run Cropping, Resizing to produce 224x224x3 images for paired dataset
10: Run Augmentation for Training dataset
11: Compress training and validation datasets from Steps 5,6
12: Run ResNet50 CNN for select Experiment
13: Run InceptionResv2 CNN for select Experiment
14: Capture validation results from CNNs
15: flush cache
16: **END LOOP**

These steps for a single iteration (pre-processing method) can be visualised in the processing flow diagram provided in Figure 6.1.

Figure 6.1: Processing Flow



6.2 Configuration

These experiments were conducted on a Dell T5810 workstation with Xeon 10 core CPU, 96gb memory and RTX3060 12Gg GPU card with python 3.9, Keras 2.9, Tensorflow 2.9 with GPU support and CUDA 11.3.

6.2.1 Datasets

Datasets were prepared and labelled into Plus disease, Stages and Zones as described in Section 4.2. The training datasets were further augmented for unique additional images as highlighted in Section 4.3. The augmented datasets carried over the labelling from the original dataset images. This allowed standardized developed datasets as described in Section 4.5 was then be used to train and validate respective CNNs for Plus disease, Stages and Zones. In the CNN classifiers, self-adjusting training rate was added with start rate of 1e-5 which is called by LearningRateScheduler, batch size of 32, epochs of 10, Adam optimizer, patience of 5 early stop Inference rate was set at standard 0.5. Early stopping was employed by monitoring the validation loss.

Experiments were arranged in the order of Plus disease, Stage, and Zone classifiers. For each, a series of 8 datasets (training/validation) were created following pre-processing of each original image using the methods as described in Sections 5.1, 5.2, 5.3, 5.4 with sample output referenced from Figures 6.2, 6.3 and 6.4.

- Unprocessed Original (baseline) (Figure 6.2.original)
- Grayscale (Figure 6.2.grayscale)
- CLAHE-GRAY-Histogram (CGH) (Figure 6.2.CGH)

- CLAHE (Figure 6.2.CLAHE)
- Pixel Colour Amplification (PCAr) composite (Figure 6.2.PCAr)
- PCAr-CLAHE (Figure 6.2.PCAr-CLAHE)
- DPFRr (Figure 6.2.DPFRr)
- DPFRr-CLAHE (Figure 6.2.DPFRr-CLAHE)
- Vessel eroded images from Original (Figure 6.4.Original eroded)
- Eroded DPFRr (Figure 6.4.DPFRr-CLAHE Eroded) is shown but not used in the classifiers.

The last two datasets denoted above as 'Vessel eroded images from Original', and 'Eroded DPFRr' are shown as results of pre-processing. They were not used in the Stages classification with any of the transfer learning based classifiers we used. They will be used for future work namely, Stage segmentation based demarcation detection.

For each image pre-processing method, a pair of datasets were generated from original training and validation dataset. This uses a pre-labelled originally segregated dataset created by data collection in Section 4.2. Resizing was performed for the images/labels prior to training and validation using Lanczos method as recommended in Section 5.5. For each method's pre-processed pair of datasets (training, validation), both training and validation is performed using two classifiers namely ResNet50 and InceptionResv2. If required Dense169 was run to verify any outliers for Stages. This was to ensure that results were replicable, and no anomaly entered the results. Also, it paved the way for the ensemble framework.

The above classifiers were based on transfer learning using imagenet with new layers added for Plus disease, Stages and Zones. The results were then recorded.

6.3 Results and Discussion

The results from each classifier and method of pre-processing are discussed next separately by ROP pathology namely Plus disease, Stages and Zones.

6.3.1 Results

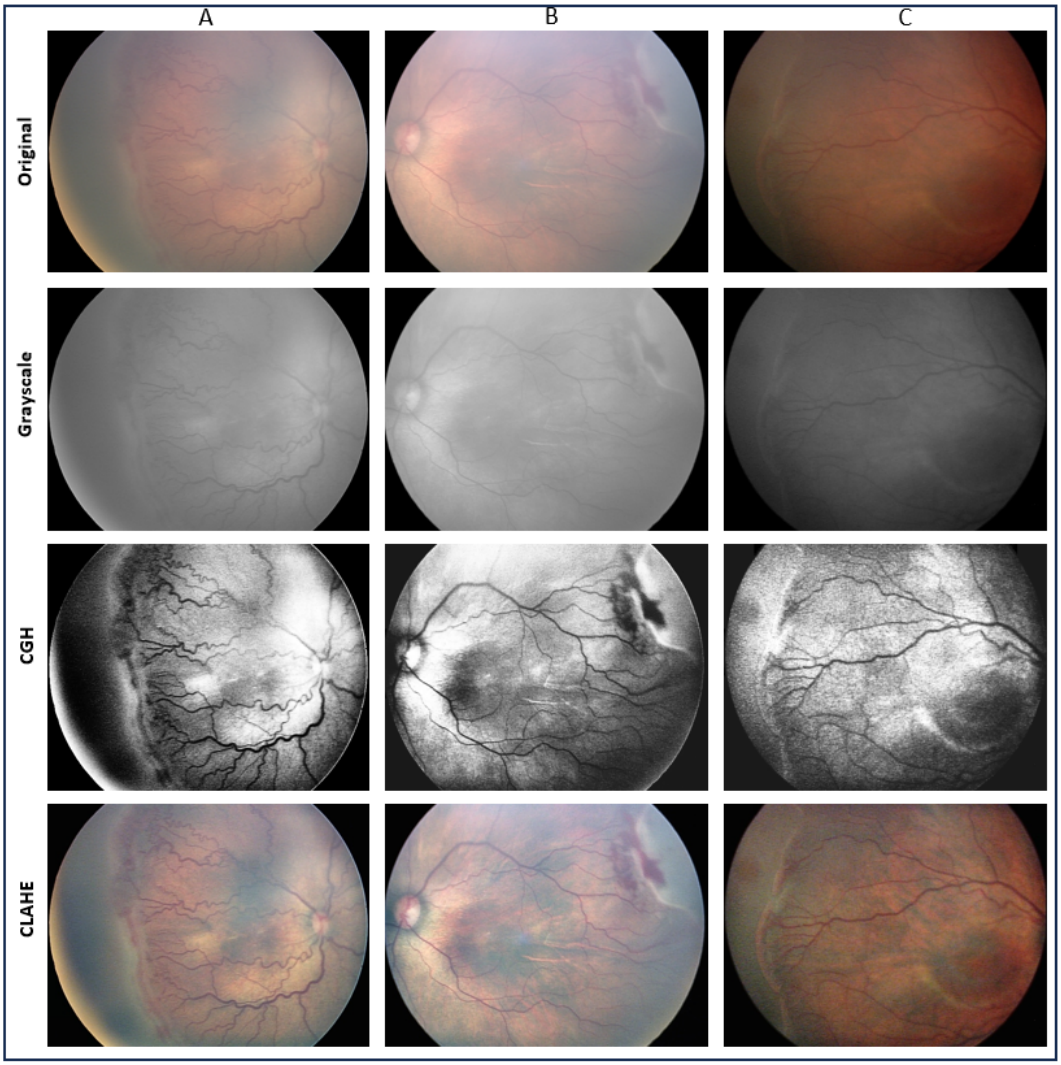
The results from our system for each classifier and method of pre-processing are discussed next. Using our limited dataset and augmentation techniques, the transfer learning based classifiers demonstrated improved classification results. These improvements reflected better outcomes when using the two improved restoration based image pre-processors in comparison to traditional methods. The results for each pre-processing method by classifiers are shown separately in terms of Plus disease, Stages, and Zones. A confusion matrix was used to breakdown the results for the best method for further system classification analysis. Confusion matrix showed the actual and predicted values that included true positives, false positives, true negatives and false negatives. The values, sensitivity, specificity, precision and accuracy were calculated based on Equation 4.1, Specificity Equation 4.2, Precision Equation 4.3, F1-Score Equation 4.4, and Accuracy Equation 4.5 is the percentage of accurately classified images. As noted before in Section 4.2, we focused on sensitivity, selectivity, precision and F1 Scores as our primarily metrics by seeking which method/CNN provided us the highest values together with precision.

These calculations allowed us to compare results with other ROP CNN based surveyed works as described in Section 3.4.3.

6.3.2 Pre-processing Method output

We first reviewed the image pre-processing outputs to summarise their key output features specific to ROP clinical features, namely blood vessels, optic disc, macula and demarcation line if present. In the case of image domain methods, grayscale images demonstrated even shade of gray with brightness evenly distributed with a low level contrast. This method has a degree of reflection present that caused inability to highlight ROP specific features. CGH method further improved on grayscale and provided a significantly better contrast for blood vessels. The demarcation line was also very visible including macula. 3 channel CLAHE colour output showed better colour contrast and sharper output than the original unprocessed image. This allowed good visualisation of ROP features. Figure 6.2 shows 3 ROP RetCam original sample images from McROP dataset are shown with respective grayscale, CGH and CLAHE results underneath the original.

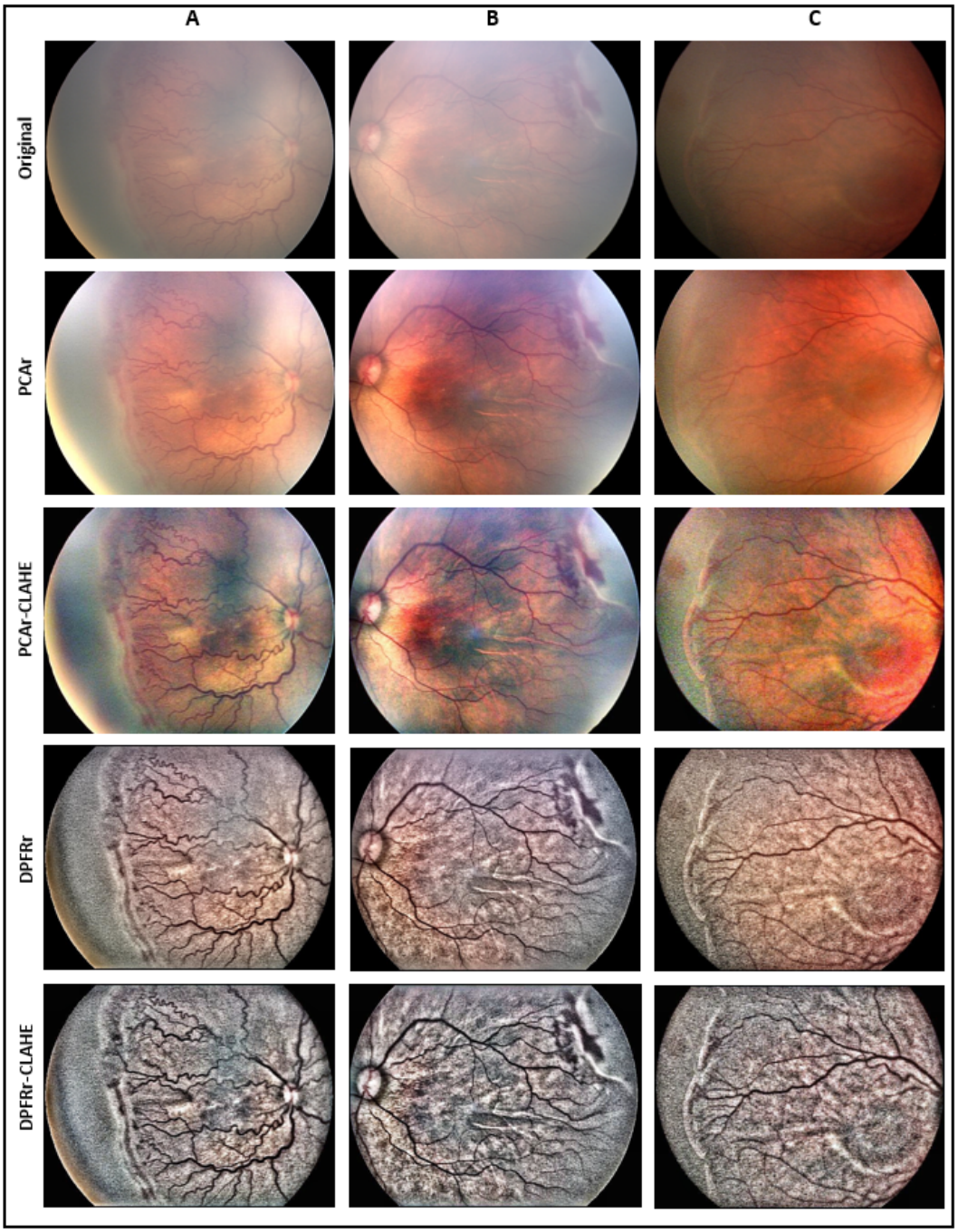
Figure 6.2: McROP RetCam samples using image domain based Pre-processing



In the case of the improved restoration methods, PCAr composites showed better contrasting than CLAHE. All ROP clinical features were visible, but there were challenges with red colour being amplified from the original image. This was resolved with the application of 3 channel CLAHE. The result of PCAr-CLAHE firstly reduced overall reflection and improved the contrast.

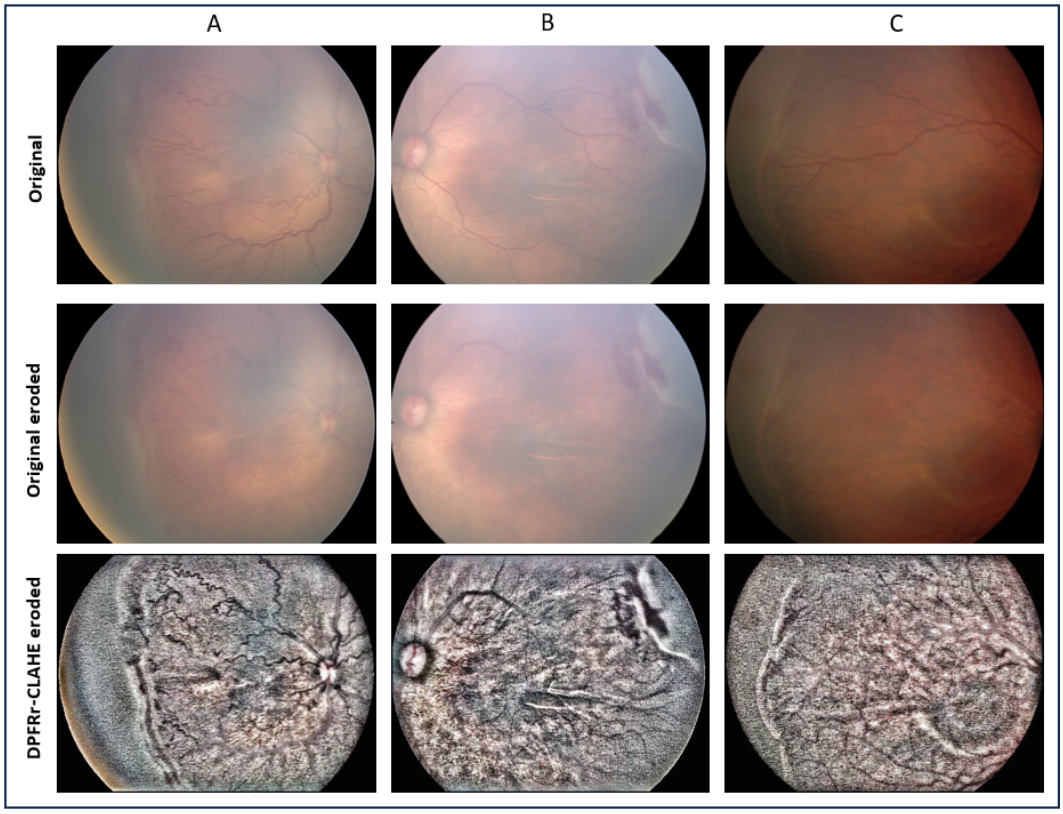
Blood vessels, macula, optic disc as well as demarcation line became very pronounced in terms of visibility. DPFRr for same original shows a remarkable cancellation of the reflection making the vessels more pronounced. This was much improved in comparison to PCAr-CLAHE. However reddish colour remains speckled across the image. By applying 3 channel CLAHE to DPFRr in DPRFr-CLAHE method, it cancelled this problem and presented an image where ROP features were sharper and pronounced. DPFRr-CLAHE as well as PCAr-CLAHE both allowed for good visualisation of ROP features. Figure 6.3 shows 3 ROP RetCam original sample images from McROP dataset are shown with respective PCAr, PCAr-CLAHE, DPRFr, and DPFRr-CLAHE results underneath the original.

Figure 6.3: McROP RetCam samples using restoration based based Pre-processing



We had also prepared eroded datasets. The first dataset was the original stages McROP dataset. We applied the erosion algorithm 5. The eroded original images show a complete erosion based off the generated maps. This image was then applied with DPFRr-CLAHE. The result showed the presence of the background where the vessels had been eroded. In other words, while erosion was not visible on the regular image, in DPRFr-CLAHE where reflection and colour contrasting has been resolved, the erosion effect is visible and therefore not a viable input into a CNN. The determination of viability is due to the visiblity of eroded lines which can challenge the classifier. Figure 6.4 shows 3 ROP RetCam original sample images from McROP dataset with respective original eroded images and DPFRr-CLAHE output of the eroded original image underneath respective original images.

Figure 6.4: McROP RetCam samples using Erosion based Pre-processing



We were able to see improvements in overall ROP RetCam image quality. The best outputs were from PCAr-CLAHE and DPFRr-CLAHE which provided significantly improved ROP clinical features in comparison to other methods. With the pre-processing methods in place, we discuss the results of each of these methods in terms of ROP Classification using CNNs in the next sections.

6.3.3 Plus Disease Classification

Using Plus/No Plus datasets, we trained and validated with Resnet50 and InceptionResv2 CNN classifiers for each image pre-processing methods as well as

with unprocessed images. In all cases, same original dataset pair was the baseline with which 7 pre-processed dataset pairs were generated. This ensured direct comparison. The results obtained for each classifier trained independently for paired dataset is illustrated in Tables 6.1, and 6.2. The best results are highlighted in yellow/bold. In ResNet50 CNN, for Plus disease, using the same data but pre-processed showed slight improvements based on Sensitivity, Specificity, Precision, and F1 Score. Both DPFR/DPFRr-CLAHE and showed best results over other methods with PCAr being tied as second best. With InceptionResv2, similar results to ResNet50 were found for Plus disease. Unprocessed (baseline) had lowest sensitivity with DPFRr and DPFRr-CLAHE showed the best results similar to ResNet50. PCAr/PCAr-CLAHE did not perform well here. With InceptionResv2, we are able to see a clear trend of improvements with each type of pre-processing method starting from grayscale and reach best outcome using restoration method of DPFRr/DPFRr-CLAHE. This improvement can be directly attributed by the improved visibility of blood vessels required for this classification.

	ResNet50 - Plus Disease					
Methods	**Sensitivity**	**Specificity**	**Precision**	**F1**	**Kappa**	**Accuracy**
Base	0.9764	0.9764	0.9764	0.9764	0.9529	0.9764
Gray	0.9764	0.9764	0.9764	0.9764	0.9529	0.9764
CLAHE	0.9764	0.9764	0.9764	0.9764	0.9529	0.9764
CGH	0.9647	0.9647	0.9647	0.9647	0.9294	0.9647
DPFRr	0.9764	0.9764	0.9764	0.9529	0.9529	0.9764
DPFRr_CLAHE	**0.9764**	**1.0000**	**1.0000**	**0.9764**	0.9529	**0.9764**
PCAr	0.9764	0.9764	0.9764	0.9529	0.8588	0.9764
PCAr_CLAHE	0.9351	1.0000	1.0000	0.9664	0.8824	0.9412

Table 6.1: ResNet50 results for Plus disease for each pre-processing method

The best overall image pre-processing result for Plus disease detection was evenly split between ResNet50 and InceptionResv2 CNNs with DPFRr-CLAHE pre-processing method providing best sensitivity/specificity. Using

	InceptionResv2 - Plus Disease					
Methods	**Sensitivity**	**Specificity**	**Precision**	**F1**	**Kappa**	**Accuracy**
Base	0.9111	1.0000	1.0000	0.9536	0.8353	0.9176
Gray	0.9474	1.0000	1.0000	0.9729	0.9059	0.9529
CLAHE	0.9474	1.0000	1.0000	0.9729	0.9059	0.9529
CGH	0.9333	0.8000	0.9722	0.9524	0.8353	0.9176
DPFRr	**0.9722**	**1.0000**	**1.0000**	**0.9859**	**0.9529**	**0.9759**
DPFRr_CLAHE	**0.9722**	**1.0000**	**1.0000**	**0.9859**	**0.9529**	**0.9759**
PCAr	0.9113	1.0000	1.0000	0.9536	0.8352	0.9176
PCAr_CLAHE	0.9113	1.0000	1.0000	0.9536	0.8352	0.9176

Table 6.2: InceptionResv2 results for Plus disease for each pre-processing method

	Performance of DPRFr-InceptionResv2 Architecture - Plus Disease					
Type	**Features**	**Sensitivity**	**Specificity**	**Precision**	**F1 Score**	**Accuracy**
No Plus	Normal	0.9700	1.0000	1.0000	0.9900	0.9765
Plus	Presence of Plus disease	1.0000	0.97	0.85	0.9200	0.9765

Table 6.3: Plus Classification breakdown for DPFRr-CLAHE-ResNet50

	Confusion Matrix	
Plus Disease	**No Plus**	**Plus**
No Plus	72	0
Plus	2	11

Table 6.4: Confusion Matrix for Plus Disease for DPFRr-CLAHE

the ResNet50/DPFRr-CLAHE as the best of the three top, the result was then further broken down by No Plus/Plus disease in Fig 6.3 using confusion matrix in Table 6.4 that resulted in overall accuracy of 97.65% which is marginally less than InceptionResv2/DPFR-CLAHE's accuracy of 98.59%. As both ResNet50 and InceptionResv2 for this method gave best results, we chose ResNet50 architecture. ResNet50 with DPFRr-CLAHE yielded the best results overall with sensitivity and specificity of 97.64% and 100%. Precision and F1-Score were 100%, 98.64% respectively. This was as of the point of writing, the only research where CNN based Plus disease classification had been performed using a restoration based image pre-processing.

6.3.4 Stages Classification

In terms of Stage disease, we were challenged with lack of large-scale ROP dataset with each case, especially for Stages 4, and 5. As noted previously, Stage 4 or Stage 5 data was not available, we we instead focused on Stages 0-3. Our approach differed from other ROP papers [133] which combined various conditions e.g. Stage, Zone and Plus. Instead, using previous findings of Mulay et al., [74], Ding et al., [18] and Tong et al., [125], this research focused on the impact of image pre-processing on ROP classifiers for Stages such that the presence of demarcation line was picked up. Our approach was to understand the problem of fundus reflection and then tuning relevant methods for ROP.

Using the seven (7) pre-processing datasets including unprocessed baseline ROP dataset, the results are highlighted in bold in Tables 6.5, 6.6 for Resnet50 and InceptionResv2 CNNs respectively. Using sensitivity, specificity, Precision and F1-Score as a primary measures, in ResNet50, PCAr performed best of all image pre-processing methods with sensitivity/specificity of 56.89%/93.44%. Precision and F1-Score were 63.46% and 60% respectively. PCAr-CLAHE was second best with scores of 53.44%/91.72%/56.36%/54.87% . In comparison, CLAHE and unprocessed images had the lowest value of 32.76%/98.96%/60.53%/47.92% and 39.83%/94.83%/61.90%/52%. With InceptionResv2, we saw DPFRr-CLAHE obtain the highest sensitivity/specificity of 72.41%/96.21%. Precision and F1-Score were 79.25%/75.67% respectively. DPFRr had the second best outcome with sensitivity/specificity of 63.79%/94.48%. Precision and F1-Score was 69.81%/66.67% respectively. The lowest results were from unprocessed fundus images with sensitivity/specificity of 39.65%/95.86%. Precision and F1-Score were 65.71%/49.46%.The results

demonstrated a pattern of improvements in ROP Stage detection when using our newer improved methods in comparison to unprocessed or standard traditional techniques. In the newer improved restoration methods, the demarcation line is more visible and additional noise relating to colour is suppressed especially when further processed with CLAHE. To illustrate this case, when we compare CLAHE with DPFRr, in either ResNet50 or InceptionResv2, we see sensitivity improve. When DPFRr-CLAHE which is DPFRr with CLAHE post processing, the sensitivity is much further improved. This is primarily because the colour factor in DPFRr is suppressed resulting in the demarcation line being better visible after processing.

	ResNet50 - Stages					
Methods	**Sensitivity**	**Specificity**	**Precision**	**F1**	**Kappa**	**Accuracy**
Base	0.3965	0.9483	0.6053	0.4792	0.3999	0.8563
Gray	0.4483	0.0448	0.6190	0.5200	0.4119	0.8621
CLAHE	0.3276	0.9896	0.8636	0.4750	0.422	0.8791
CGH	0.3448	0.9483	0.5714	0.4301	0.3483	0.8477
DPFRr	0.4482	0.9552	0.6667	0.5361	0.4642	0.8707
DPFRr_CLAHE	0.4482	0.9217	0.6500	0.5306	0.4567	0.8678
PCAr	**0.5689**	**0.9344**	**0.6346**	**0.6000**	**0.5252**	**0.8736**
PCAr_CLAHE	0.5344	0.9172	0.5636	0.5487	0.4617	0.8534

Table 6.5: Resnet50 results for Stages classification for each pre-processing method

	InceptionResv2 - Stages					
Methods	**Sensitivity**	**Specificity**	**Precision**	**F1**	**Kappa**	**Accuracy**
Base	0.3965	0.9586	0.6571	0.4946	0.4221	0.8649
Gray	0.5000	0.9241	0.6667	0.5321	0.4456	0.8534
CLAHE	0.6034	0.9276	0.6250	0.6140	0.5386	0.8736
CGH	0.4828	0.9207	0.5490	0.5138	0.4239	0.8477
DPFRr	0.6379	0.9448	0.6981	0.6667	0.6036	0.8936
DPFRr_CLAHE	**0.7241**	**0.9621**	**0.7925**	**0.7567**	**0.7107**	**0.9224**
PCAr	0.5344	0.9207	0.5740	0.5536	0.468	0.8563
PCAr_CLAHE	0.5172	0.9207	0.5660	0.5405	0.4535	0.8535

Table 6.6: InceptionResv2 results for Stage classification for each pre-processing method

	Performance of DPRFr-InceptionResv2 Architecture - Stages					
Type	**Features**	**Sensitivity**	**Specificity**	**Precision**	**F1 Score**	**Accuracy**
Stage 0	No Stage disease	0.6300	0.8800	0.6700	0.6500	0.8103
Stage 1	Presence of demarcation line	0.6000	0.9400	0.6700	0.6300	0.8793
Stage 2	Presence of ridge	0.8300	0.8800	0.7500	0.7900	0.8621
Stage 3	Ridge with additional features	0.8600	0.9500	0.8600	0.8600	0.9310

Table 6.7: Stages Classification breakdown for DPFRr-CLAHE-InceptionResv2

	Confusion Matrix			
STAGES	**Stage 0**	**Stage 1**	**Stage 2**	**Stage 3**
Stage 0	10	2	2	1
Stage 1	2	6	1	0
Stage 2	3	1	15	1
Stage 3	1	1	0	12

Table 6.8: Confusion Matrix for Stages result for DPFRr-CLAHE-InceptionResv2

Amongst all methods/CNN combination, we found DPFRr-CLAHE using InceptionResv2 CNN gave best sensitivity/specificity of 72.41%/96.21% with Precision and F1-Score being 79.25%/75.67% respectively. Its detailed breakdown is shown in Table 6.7 with its confusion matrix in Table 6.8. DPFRr-CLAHE/InceptionResv2 results breakdown showed that it provided a high sensitivity/specificity/precision/F1-Score for Stage 0 (63%/67%/88%/65%) followed by Stage 1 (60%/67%/94%/63%), Stage 2 (83%/75%/88%/79%) and Stage 3 (86%/86%/95%/86%) respectively. This can be explained with the following insights. Stage 1 classification was still challenged primarily due to the very fine grainy output which, though reduced significantly did cause challenges in detecting the faintest of demarcation line. The transition between Stage 0 and Stage 1 such that improved Stage 1 detection is part of future work of improvement. Stage 2 and Stage 3 show significant improvements in classification primarily because the demarcation line is now very visible. This was

as of the point of writing, the only research where Stage classification for ROP Retcam, had been performed using a restoration based image pre-processing.

6.3.5 Zones Classification

While Zone classification is a challenge to identify, we took the opportunity to attempt this as a proof of concept as there was insufficient validation data. The criteria for Zone eligible image is very stringent and a Data Qualifier process may be a valuable contribution to this body of knowledge. Using the seven (7) pre-processing datasets including an unprocessed baseline ROP dataset, the results are highlighted in bold in Tables 6.9, 6.10 and its breakdown for the best performer in Fig 6.11 and its respective confusion matrix in Table 6.12 resulting in overall accuracy of 85.37%.

	ResNet50 - Zones					
Methods	**Sensitivity**	**Specificity**	**Precision**	**F1**	**Kappa**	**Accuracy**
Base	0.7805	0.9146	0.8205	0.8000	0.7037	0.8699
Gray	0.7073	0.8536	0.7073	0.7073	0.5610	0.8699
CLAHE	0.7805	0.9024	0.8000	0.7912	0.6871	0.8618
CGH	0.7317	0.8902	0.7692	0.8373	0.6296	0.8374
DPFRr	0.7317	0.8780	0.7500	0.7407	0.6134	0.8293
DPFRr_CLAHE	**0.8537**	**0.9268**	**0.8536**	**0.8536**	**0.7804**	**0.9024**
PCAr	0.7561	0.8781	0.7561	0.7561	0.6341	0.8374
PCAr_CLAHE	0.7073	0.9024	0.7838	0.8374	0.6250	0.8374

Table 6.9: ResNet50 results for Zones classification for each pre-processing method

In our results, we observed that DPRFr-CLAHE using ResNet50 worked well here. It provided high sensitivity/specificity of 85.37%/92.68%. Precision and F1-Score were 85.36%/85.36%. Further breakdown analysis of DPFR-r/ResNet50 results identified Zone I/II with sensitivity/specificity/precision/F1-Score and overall accuracy. In terms of Zone I detection, it showed sensitivity/specificity/precision/F Score of 100%/97%/67%/80% (accuracy of 93%), and Zone II with 84%/100%/85%/91.67%

	InceptionResv2 -Zones					
Methods	**Sensitivity**	**Specificity**	**Precision**	**F1**	**Kappa**	**Accuracy**
Base	0.5854	0.8902	0.7273	0.6486	0.5000	0.7886
Gray	0.4146	0.8536	0.5862	0.4857	0.2894	0.7073
CLAHE	0.5609	0.8170	0.6052	0.5822	0.3850	0.7317
CGH	0.5854	0.8537	0.6667	0.6234	0.4528	0.7642
DPFRr	0.5121	0.8283	0.6000	0.5526	0.4528	0.7236
DPFRr_CLAHE	0.5122	0.8537	0.6363	0.5676	0.3846	0.7398
PCAr	0.5122	0.8049	0.5676	0.5385	0.3250	0.7073
PCAr_CLAHE	**0.6097**	**0.8095**	**0.6098**	**0.6098**	**0.4444**	**0.7440**

Table 6.10: InceptionResv2 results for Zones classification for each pre-processing method

	Performance of DPRFr-InceptionResv2 Architecture - Zones					
Type	**Features**	**Sensitivity**	**Specificity**	**Precision**	**F1 Score**	**Accuracy**
Zone I	Presence of disease in Zone I	1.0000	0.9744	0.6700	0.8000	0.9756
Zone II	Presence of disease in Zone II	0.8462	1.0000	1.0000	0.9167	0.8537
Zone III	Presence of disease in Zone III	0.0000	0.8780	0.0000	0.0000	0.8780

Table 6.11: Zones Classification breakdown for DPFRr-CLAHE ResNet50

Confusion Matrix			
Zones	**Zone I**	**Zone II**	**Zone III**
Zone I	2	1	0
Zone II	0	33	0
Zone III	0	5	0

Table 6.12: Confusion Matrix for Zones Classification

(accuracy of 85.37%) respectively. Zone III was challenged due to lack of sufficient data. Zhao [146] in comparison only obtained 91% for Zone I. He did not attempt Zone II, or Zone III. This work was, as far as we are aware, where Zone I and II had been shown to be classified successfully using deep learning. We note once more this is a proof of concept for feasibility with potentially more work required.

Overall, there was improvement noted for PCAr when combined with CLAHE. Similarly same effects were noted for DPFRr-CLAHE as well.

6.3.6 Discussion

In lieu of the results obtained and comparative analysis with other related works, we demonstrated that there were significant improvements as a result of improved image pre-processing that improved accuracy for Plus disease. It was very significantly improved for Stages and to some extent Zones.

For Plus disease classification, our improved novel ROP specific processes (PCAr, PCAr-CLAHE, DPFRr, DPFRr-CLAHE) showed improvements over other methods in classification when used with the transfer learning based CNNs. Both DPFRr and DPFRr-CLAHE showed improvements primarily due to the removal of the source of reflection as well as the reduction of the choroidal blood vessels. This allowed primary blood vessels where tortuosity was present to be more clearly visible. Applying CLAHE to the DPFRr reduced the colour factor further. The comparative analysis as noted in the table 6.13 below suggested that these new hybrids showed better results when compared to other research papers where R-CNN was used as well. Analysing our results for DPFRr-CLAHE/InceptionResv2 CNN results, a high precision, sensitivity, specificity for NoPlus (100%/97%/100%) and Plus (85%/100%/97%/65%) was obtained. This was similar to results obtained by iROP-DL [8], and better than Vinekar et al., [130].

	Results Comparison			
Measure	**McROP**	**Brown et al.,[8]**	**Tong et al.,[125]**	**Vinekar et al.,[130]**
Sensitivity	0.98	0.93	0.71	0.95
Specificity	1.00	0.94	0.91	1.00

Table 6.13: Comparison of McROP DPFRr-CLAHE InceptionResv2 for Plus Diseases

The described methods in conjunction with CLAHE did provide a very significant improvement over tradition and rival R-CNN methods by removing the source of the fundus image problem. Namely, reflection cancellation, and colour reduction to amplify the ROP clinical features more clearly for Plus disease. We propose that R-CNN methods can still be used to further improve the classifiers as further add on methods for Plus disease.

In terms of Stages classification, it was important to understand the difference of improvements between DPFRr and DPFRr-CLAHE. The analysis of results and pre-processed images showed two critical findings emerge. First, DPFRr significantly reduced reflection and improved retinal features for ROP but red coloration was still present which challenged the classifiers. Once the 3 channel based CLAHE applied, there was a significant further reduction in colours and made the demarcation line more pronounced. The 3 channel based CLAHE application to DPFRr provided a better ROP Stage featured image allowing the classifier to better determine the Stage classification. This was critical when it comes to Stages 2-3 in comparison to two comparative papers without leveraging R-CNN. This combination provided a very significant improvement over tradition and rival R-CNN methods by removing the source of the ROP RetCam image problem. As noted, addition of R-CNN further to this work should improve the results further.

	McROP				**Ding et al (Hybrid)**			
Stages	**Precision**	**Recall**	**Specificity**	**F1**	**Precision**	**Recall**	**Specificity**	**F1**
Stage 0	0.67	0.63	0.88	0.65	-	-	-	-
Stage 1	0.67	0.6	0.94	0.63	0.78	0.77	-	0.78
Stage 2	0.75	0.83	0.88	0.79	0.61	0.62	-	0.61
Stage 3	0.86	0.86	0.95	0.86	0.62	0.62	-	0.62

Table 6.14: Comparison of McROP DPFRr-CLAHE InceptionResv2 vs Ding et al., Hybrid

	McROP				Ding et al., Classifier Only			
Stages	**Precision**	**Recall**	**Specificity**	**F1**	**Precision**	**Recall**	**Specificity**	**F1**
Stage 0	0.67	0.63	0.88	0.65	-	-	-	-
Stage 1	0.67	0.6	0.94	0.63	0.98	0.36	-	0.53
Stage 2	0.75	0.83	0.88	0.79	0.45	0.86	-	0.59
Stage 3	0.86	0.86	0.95	0.86	0.59	0.36	-	0.45

Table 6.15: Comparison of McROP DPFRr-CLAHE InceptionResv2 vs Ding et al., Classifier Only

	McROP				Tong et al.,			
Stages	**Precision**	**Recall**	**Specificity**	**F1**	**Precision**	**Recall**	**Specificity**	**F1-Score**
Stage 0	0.67	0.63	0.88	0.65	-	-	-	-
Stage 1	0.67	0.6	0.94	0.63	0.86	0.77	0.88	0.81
Stage 2	0.75	0.83	0.88	0.79	0.49	0.55	0.97	0.52
Stage 3	0.86	0.86	0.95	0.86	0.55	0.47	0.98	0.51

Table 6.16: Comparison of McROP DPFRr-CLAHE InceptionResv2 vs Tong et al.,

With Stage only based comparison, our results in Stage 2 and Stage 3 showed significantly better outcomes than Ding et al., [18] Tong et al., [125]. Tables 6.14, 6.15, and 6.16 illustrate the comparisons between our results with theirs. Ding et al., [18] used R-CNN hybrid classifier and a pure classifier. Using an R-CNN hybrid, they achieved the following sensitivity/precision: 62%/61% for Stage 2, and 62%/62% in Stage 3 in terms of sensitivity/precision . Pure classifier results are 36%/98%, 86%/45%, and 36%/59%. In comparison, we performed better in classifier only in all stages while in the R-CNN hybrid, our results were better for Stage 2 and Stage 3. Ding's Stage 1 results were slightly better at 77%/78% using R-CNN hybrid. Further, we obtained better results using smaller 224x224 images whereas Ding et al., [18] used larger 299x299 images. Similarly Tong et al., [125] used 224x224 like us and noted challenges with Stage 0 and Stage 1 similar to our study. They did not state if their Stage 1 data was inclusive of faint demarcation line as they did not show Stage 0 either or it was only Stage 1 where the demarcation line

is visible. In our case, we did not distinguish between the intensity of the line pertaining to Stage 1. For Stage 2 and 3, we showed significant improvement in comparison. DPFR-CLAHE/InceptionResv2 CNN combination achieved in terms of sensitivity/specificity 83%/88% vs 55%/97%for Stage 2, 86%/96% vs 47%/98% for Stage 3.

As noted before Zones, classification was a proof of concept and this approach also offered a promising improvement which required further work to be able determine the geometric aspect zones. Further breakdown analysis of DPFRr/ResNet50 results identified Zone I/II with very high precision/sensitivity and overall accuracy. In terms of Zone I detection, it showed precision and sensitivity of 67%/100% (accuracy of 93%), and Zone II with 100%/85% (accuracy of 85.37%) respectively. Zhao [146] in comparison only obtained 91% for Zone I. This is, as far as we are aware, the only paper where Zone I and II have been shown to be classified successfully using deep learning and using restoration based methods. Zone III was challenged due to lack of sufficient data. Overall, there was improvements noted for PCAr when combined with CLAHE. Similar, same results were noted for DPFRr-CLAHE as well.

6.3.7 Limitations

This study had several limitations. We noted the challenges in obtaining large RetCam ROP datasets. Our CNNs were limited by the data we used for training including unavailable data for Stage 4 and 5 as well as Zones. Augmentation was used to generate further training data. We could not obtain independent ROP RetCam data to be used as a testing dataset, thereby limiting our ability to including testing. Our system at present allowed the classifica-

tion of Plus, Stages 0-3 and to a limited degree Zones. Our future work would include the ability to firstly qualify the incoming data in terms of quality and allow it to predict all three aspects at the same time. In other words a unified model of ROP pathology.

6.3.8 Miscellaneous findings

As part of this series of experiments, we also investigated the possibility of training Stages classifiers with Stages training RetCam ROP dataset pre-processed using one method with corresponding Stages RetCam fundus validation dataset pre-processed using an alternate method. For Stages, training was performed with Grayscale but validation with PCAr-CLAHE. The results are noted in Table 6.17 and its resulting confusion matrix in Table 6.18.

	InceptionResv2-Stages						
Methods	**Sensitivity**	**Specificity**	**Precision**	**F1**	**Kappa**	**AUC**	**Accuracy**
Gray/PCAr_Clahe	0.1724	0.8448	0.1818	0.7327	0.176	0.5025	0.7327

Table 6.17: Mixed pre-processor use with McROP InceptionResv2

Confusion Matrix				
Stages	**Stage 0**	**Stage 1**	**Stage 2**	**Stage 3**
Stage 0	1	13	0	1
Stage 1	1	7	0	1
Stage 2	1	18	1	2
Stage 3	0	11	0	3

Table 6.18: Confusion Matrix for mixed pre-processor usage with McROP InceptionResv2

The results were very poor and therefore leading to the recommendation that the same pre-processing method be used for both the Training and Vali-

dation datasets. As the resulting images were different for each pre-processing method, each of these methods can be used as additional methods for RetCam fundus image augmentation. In particular, Pixel Colour Amplification Illumination for ROP (PCAr) correction method we contributed here could be further extended to generate a set of images from the eight methods A-D and W-Z for only those where visible features are present while negating others. A wider spectrum of these now readily available methods could greatly contribute to meeting challenges of augmentation faced only within the ROP CNN arena but other fundus specific pathology areas.

6.4 Chapter Summary

This chapter summarised the experiments conducted using the limited dataset gathered in conjunction with an overall process named as McROP. Using unprocessed baseline image, data was processed using image pre-processing that included two extended novel methods for ROP and combination of traditional methods. Two specific classifiers pre-trained on Imagenet were further trained either in Plus disease, Stages, and Zones. The results were insightful in that they provided the following findings. There was not one particular method / CNN which was dominant but rather one may work best for the specifics e.g. Plus disease, Stage or Zone e.g. DPFRr-CLAHE was best for Stage and Plus disease using InceptionResv2 CNN and DPFRr-CLAHE was best suited for Zones using ResNet50 CNN. Once a larger dataset would be made available, it would serve to validate these findings further.

Chapter 7

Conclusions and Future Work

The development of any ROP RetCam related computer aid diagnosis system is a difficult challenge. This requires an indepth understanding of the ROP pathology and the physiology of the human eye. Advanced knowledge of image pre-processing methods is critical for solving the challenges related to ROP RetCam image quality and thereby improving CNN based diagnosis. With Covid-19 impact, there is a greater need for better imaging processes as ROP patient care was under stress. Research from this thesis may help in effective remote screening with improved image quality while providing input to a more robust automated diagnostic system for Plus disease, Stages and Zones.

This chapter concludes this thesis and summarises the research. Section 7.1 discusses the research contributions and how it has improved the outcomes of the diagnosis using CNNs. Section 7.2 describes potential future research areas based on this thesis research which can provide additional contribution to this challenging area. Section 7.3 provides the final summary and conclusion to this chapter.

7.1 Summary of Contributions

The thesis began with the understanding of the challenges faced in paediatric retinopathy. Lack of trained ROP specialists with ever increasing demands for ROP paediatric screening can be assisted by automated ROP screening especially for remote areas using Telehealth where possible. Pediatric patient's retina images are captured either with RetCam [44] or smartphone prior to being sent over to the ROP specialist for screening. The thesis surveyed diagnosis methods and types of digital cameras used for capturing patient's retinal images. RetCam [44] and recently, smartphones both offered these capabilities with each having their advantages and disadvantages. ROP RetCam images were studied in terms of quality issues which mandated image pre-processing to resolve them to improve ROP clinical features. These pre-processing methods reviewed included those used for adult digital fundus images for diabetic retinopathy based CNNs and paediatric ROP RetCam [44] based CNN classifications. The studies highlighted all published ROP based CNNs which classified Plus disease, Stages classification and one for Zones used traditional forms of image pre-processing. These methods were recognised as insufficient for improving the ROP features. Consequently, Stages detection remained a key problem area with difficulties in discerning each distinct Stage with a high degree of accuracy. Different methods of image pre-processing which included image domain methods, restoration including the use of segmentation were next studied. Restoration methods were considered as the ideal path for ROP RetCam [44] image pre-processing including leveraging LWNET to solve blood vessel erosion such that these features were seamlessly removed from the original image. For classifying different ROP conditions, convolutional

neural networks were studied with transfer learning using InceptionResv2 and Resnet50 frameworks.

This thesis research proposed and implemented two improved restoration based image pre-processing methods including combining with it additional 3 channel based CLAHE pre-processing step which resulted in better ROP RetCam[44] quality of features specific to Plus disease, Stages and with possible use in Zones. Another unique approach was also implemented using segmentation to erode blood vessels for clinical use. The CNN classification results demonstrated that these 2 improved restoration methods in conjunction with traditional methods contributed towards improved the image quality that highlighted ROP clinical features. This allowed the ROP RetCam pre-processed images to be used in improving CNNs classification of Plus disease, Stages 0-3, and Zones I-II.

The research can be summarised as follows:

1. This research as its core contribution improved the two image restoration methods we now referred to as PCAr and DPFRr that improved ROP RetCam images leading to better accuracy in the deep learning CNNs. The contribution firstly solved PCA method's illumination problem for single output image using a median based scoring method. This provided a balanced illumination image irrespective of the input image's illumination. Scoring results were further used to then generate a composite image from 3 of the 8 images. Secondly, DPFR was modified to be able to resolve granularity problems and reduce choroid blood vessels such that a ROP RetCam pre-processed image can highlight key ROP clinical features better. The resulting outputs from these contributions were then further enhanced with traditional methods

such as 3 channels based CLAHE that significantly improved the contrast and colour saturation problems.

2. Implemented a novel approach for vessel removal from the original ROP RetCam image by using a generated segmentation map of blood vessels from the original image using the U-Net based LWNET [31] method.

3. Implemented two CNNs - ResNet50 and InceptionResv2 leveraging transfer learning to train for Plus disease, Stages, and Zone classifications. A third CNN using Dense169 was also implemented for Stages as further validation. Transfer learning is useful when there is less training data as in our case.

4. Compared outcomes of 8 prepared McROP datasets pairs. Each pair comprised of a training and validation datasets. These included 1 original unprocessed images, 3 traditional pre-processing methods, 2 improvised restoration based image processing including extending these 2 with 3 channel CLAHE. The processed datasets used the same original unprocessed images dataset. These were input as training/validation datasets into 2 CNNs classifiers for Plus Disease, Stages, and Zones classifications.

5. Recommendation for resizing algorithm noted as Lanczos method that retains the ROP clinical features with minimal loss for use by CNNs.

6. Demonstrated that the same pre-processing method must be used for both training and validation datasets.

The above contributions combined form the core components of the McROP system.

The thesis demonstrated that the improved restoration image pre-processing and combining it with a traditional method, namely 3 channel CLAHE did improve the quality of ROP RetCam images significantly by highlighting key

ROP clinical features namely blood vessels, demarcation line, macula and optic disc. This improved classification results with high degree of accuracy from the CNNs. These CNNs were developed separately to diagnose Plus disease, Stages of ROP and Zones. This research has contributed to image processing knowledge as it pertains to Retinopathy of Prematurity.

7.2 Future Works

During this research work, many new novelties were identified. These are described as future work that can be further added to this body of knowledge. These are noted as follows:

1. Hybrid of PCAr and DPFRr

PCAr can be further improved by adding the estimation of the atmosphere and reflection cancellation. At present, atmosphere value can either be a 0 or 1 to brighten or darken. This will allow 4 priors outputs to be available which will enable the study of cancellation effect on other priors.

2. Use R-CNN mask to segment fundus image for Stage and Zones

In section 5.4, segmentation using LWNET was employed first to segment out the vasculature, which then fed into the vascular removal process. The premise of this contribution was to present an image as little of vasculature so that the training can be more focused on detection and classification of the demarcation line that is key for Stage categorisation. The alternative option was to completely partition out ROP RetCam image into area of demarcation vs rest of vasculature. With DPFRr and this studied modified Pixel Colour Amplification, R-CNN mask can be used split the image only having the de-

marcation line present. This allows the classifier CNN for Stage only to analyse this vs anything else. Another alternative is to segment out the demarcation line using LWNET directly.

This approach can be extended further to include Optic disc (OD) as well as a circle. Using OD circle and demarcation line area only, it will be possible to then overlap either solely with just training using augmented data only or add calculated circular lines for Zone I, II and III. Once the latter is done, it may be possible to very accurately point out the Zone as well.

3. Transformers and Foundational models for ROP

Transformers are the new state of the art models which will be reviewed for application towards ROP. It is very possible to employ these for the task for Plus Disease and demarcation line. Foundational models have shown promise in a limited study with DR.

4. McROP classifiers for Plus, Stage, Zone combined as a multi-instance classifier using Ensemble

At present the system overall has been designed to operate in separate streams of image processing and classifier data preparation. It is possible to create a unified model and use two separate CNN per diagnosis for Plus disease, Stages, and Zones in an ensemble mechanism to show the results across the pathology. Foundation models can also be studied for this as well.

5. Utilisation of Disqualified images

In this study, it was apparent that newly improved methods show significant improvements over traditional methods. These newer methods should also be reviewed if they may also qualify prior disqualified images using traditional image pre-processing methods, which were removed for quality reasons to be included, and thereby increasing the ROP RetCam dataset.

In this research, approximately 30% of images were of too poor quality to be considered for use. With these new improved restoration based methods, it may be possible to use this to improve the quality of fundus images. This is not only applicable for ROP datasets but in general in other fundus pathologies such as Diabetic Retinopathy.

6. Data Quality Filter

In the present works and literature survey, there is no research relating to automatic Data Quality Filter process for Plus disease, Stages and Zones. The purpose of this process is to automatically assess images to a prescribed agreed upon standard across the ROP community. A CNN based filtering approach is recommended which could leverage ROP standardised dataset with labelled ground truth such as McROP. For this thesis work, the effort required to clean and segregate the data took a long time. As larger datasets become available for further research, manual processing will be laborious and needs to be pragmatically assisted. This will make quality images available for manual ground truth verification and labelling.

7. Smartphone based capturing

Our research was based on ROP RetCam images. As noted earlier in Section 2.3, smartphone based capturing is increasingly being used. It provides an efficient and cost effective method but image type and quality differs from RetCam images. It will be ideal to have a uniform convergence output of image pre-processing results irrespective of image capture mechanism.

Each area noted above will significantly contribute towards the improvement of ROP classification and assist with generating robust ROP dataset in terms of standardisation and quality.

7.3 Chapter Summary

ROP can be life altering for the premature infant with severe complications which can be life long including possibility of vision loss. It is essential that early screening is accomplished such that timely treatment can be performed with high degree of success where necessary. As access to ROP pediatrics ophthalmologists can be limited, ROP RetCam images can play a crucial role in assisting with early diagnosis in pre-term infants. Tele-health infra-structure can help transfer and pre-process the images to clearly identify and diagnose the patient clinically. Deep learning trained classifiers can also be employed to get preliminary assessment on the ROP condition at the earliest point possible.

This research project reviewed the problem of image clarity in ROP RetCam image and studied how to improve the quality by researching the application of traditional image pre-processing methods while understanding new restoration based methods. The latter were extended for resolving for the ROP RetCam image quality challenges such that the ROP clinical features are clearly visible in the patient's RetCam image. The project also, explored segmentation problem to try and reduce the noise from the vasculature by removing them to a high degree using segmentation generated map. Finally utilising CNNs via transfer learning improved the accuracy in detecting Plus disease, Stages and Zones with small data. The research concluded by proposing two improved restoration based methods which can assist ophthalmologists identify high risk ROP patients either in a clinical setting or using CNNs. In conclusion, we demonstrated that these proposed methods improved ROP RetCam images quality by highlighting ROP specific pathology features that

resulted in improved CNN performance. These methods can contribute in the creation of accurate and reliable ROP screening systems.

Bibliography

[1] Agarwal Komal, S. J. (2018). Classification of retinopathy of prematurity: from then till now. *Community eye health*, 31101.

[2] Akram, M. U., Khalid, S., Tariq, A., Khan, S. A., and Azam, F. (2014). Detection and classification of retinal lesions for grading of diabetic retinopathy. *Computers in biology and medicine*, 45:161–171.

[3] Akram, M. U., Tariq, A., Nasir, S., and Khan, S. A. (2009). Gabor wavelet based vessel segmentation in retinal images. In *2009 IEEE Symposium on Computational Intelligence for Image Processing*, pages 116–119. IEEE.

[4] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., and Farhan, L. (2021). Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *Journal of big Data*, 8:1–74.

[5] Anila, V. and Thomas, S. (2016). Detection of diabetic retinopathy from fundus images through local binary patterns and artificial neural network. *Int J Eng Adv Technol (IJEAT)*, 6(1):5–9.

[6] Baldi, P. (2012). Autoencoders, unsupervised learning, and deep architectures. In *Proceedings of ICML workshop on unsupervised and transfer learning*, pages 37–49. JMLR Workshop and Conference Proceedings.

[7] Biten, H., Redd, T. K., Moleta, C., Campbell, J. P., Ostmo, S., Jonas, K., Chan, R. P., Chiang, M. F., et al. (2018). Diagnostic accuracy of ophthalmoscopy vs telemedicine in examinations for retinopathy of prematurity. *JAMA ophthalmology*, 136(5):498–504.

[8] Brown, J. M., Campbell, J. P., Beers, A., Chang, K., Ostmo, S., Chan, R. P., Dy, J., Erdogmus, D., Ioannidis, S., Kalpathy-Cramer, J., et al. (2018). Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA ophthalmology*, 136(7):803–810.

[9] Chaudhuri, S., Chatterjee, S., Katz, N., Nelson, M., and Goldbaum, M. (1989). Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Transactions on medical imaging*, 8(3):263–269.

[10] Chaudhury, S., Krishna, A. N., Gupta, S., Sankaran, K. S., Khan, S., Sau, K., Raghuvanshi, A., and Sammy, F. (2022). Effective image processing and segmentation-based machine learning techniques for diagnosis of breast cancer. *Computational and Mathematical Methods in Medicine*, 2022.

[11] Cheng, H.-D. and Shi, X. (2004). A simple and effective histogram equalization approach to image enhancement. *Digital signal processing*, 14(2):158–170.

[12] Cheng, J.-Z., Ni, D., Chou, Y.-H., Qin, J., Tiu, C.-M., Chang, Y.-C., Huang, C.-S., Shen, D., and Chen, C.-M. (2016). Computer-aided diagnosis

with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans. *Scientific reports*, 6(1):1–13.

[13] Coyner, A. S., Swan, R., Brown, J. M., Kalpathy-Cramer, J., Kim, S. J., Campbell, J. P., Jonas, K., Chan, R. P., Ostmo, S., and Chiang, M. F. (2018). Deep learning for image quality assessment of fundus images in retinopathy of prematurity. *Investigative Ophthalmology & Visual Science*, 59(9):2762–2762.

[14] D, N. (2022). *Unsupervised Machine Learning*. Github.

[15] Dalianis, H. (2018). *Clinical text mining: Secondary use of electronic patient records*. Springer Nature.

[16] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.

[17] Developers, T. P. (2023). pywavelets. *pywavelets*.

[18] Ding, A., Chen, Q., Cao, Y., and Liu, B. (2020). Retinopathy of prematurity stage diagnosis using object segmentation and convolutional neural networks. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6. IEEE.

[19] Dissopa, J., Kansomkeat, S., and Intajag, S. (2021). Enhance contrast and balance color of retinal image. *Symmetry*, 13(11):2089.

[20] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.

(2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

[21] Dridi, S. (2021). Supervised learning-a systematic literature review. *OSF Preprints*.

[22] Eck, D. J. (2016). *Introduction to Computer Graphics*. David J. Eck.

[23] Ergen, B. (2012). *Signal and image denoising using wavelet transform*. InTech London, UK.

[24] Farnell, D. J., Hatfield, F. N., Knox, P., Reakes, M., Spencer, S., Parry, D., and Harding, S. P. (2008). Enhancement of blood vessels in digital fundus photographs via the application of multiscale line operators. *Journal of the Franklin institute*, 345(7):748–765.

[25] Fierson, W. M., Chiang, M. F., Good, W., Phelps, D., Reynolds, J., Robbins, S. L., Karr, D. J., Bradford, G. E., Nischal, K., Roarty, J., et al. (2018). Screening examination of premature infants for retinopathy of prematurity. *Pediatrics*, 142(6).

[26] Forssard, D. (2023). Vgg16. *University of Toronto*.

[27] Foundation, P. R. R. (2023). Retinopathy of prematurity. *https://www.pediatricrrf.org/rop*.

[28] Franklin, S. W. and Rajan, S. E. (2014). Retinal vessel segmentation employing ann technique by gabor and moment invariants-based features. *Applied Soft Computing*, 22:94–100.

[29] Fraz, M. M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A. R., Owen, C. G., and Barman, S. A. (2012). Blood vessel segmentation

methodologies in retinal images–a survey. *Computer methods and programs in biomedicine*, 108(1):407–433.

[30] Fu, H., Xu, Y., Lin, S., Kee Wong, D. W., and Liu, J. (2016). Deepvessel: Retinal vessel segmentation via deep learning and conditional random field. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 132–139. Springer.

[31] Galdran, A., Anjos, A., Dolz, J., Chakor, H., Lombaert, H., and Ayed, I. B. (2022). State-of-the-art retinal vessel segmentation with minimalistic models. *Scientific Reports*, 12(1):6174.

[32] Gangwar, A. K. and Ravi, V. (2021). Diabetic retinopathy detection using transfer learning and deep learning. In *Evolution in Computational Intelligence: Frontiers in Intelligent Computing: Theory and Applications (FICTA 2020), Volume 1*, pages 679–689. Springer.

[33] Garcia, M., Hornero, R., Sánchez, C. I., López, M. I., and Díez, A. (2007). Feature extraction and selection for the automatic detection of hard exudates in retinal images. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4969–4972. IEEE.

[34] Gaudio, A., Smailagic, A., and Campilho, A. (2020). Enhancement of retinal fundus images via pixel color amplification. In *Image Analysis and Recognition: 17th International Conference, ICIAR 2020, Póvoa de Varzim, Portugal, June 24–26, 2020, Proceedings, Part II 17*, pages 299–312. Springer.

[35] Ghosh, R., Ghosh, K., and Maitra, S. (2017). Automatic detection and classification of diabetic retinopathy stages using cnn. In *2017 4th International Conference on Signal Processing and Integrated Networks (SPIN)*, pages 550–554. IEEE.

[36] Gojić, G., Petrović, V., Turović, R., Dragan, D., Oros, A., Gajić, D., and Horvat, N. (2020). Deep learning methods for retinal blood vessel segmentation: evaluation on images with retinopathy of prematurity. In *2020 IEEE 18th international symposium on intelligent systems and informatics (SISY)*, pages 131–136. IEEE.

[37] Gonzalez, R. C., Woods, R. E., and Eddins, S. L. (2004). Digital image processing using matlab. *(No Title)*.

[38] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.

[39] gradcam (2023). Gradcam. *Keras*.

[40] Graps, A. (1995). An introduction to wavelets. *IEEE computational science and engineering*, 2(2):50–61.

[41] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22):2402–2410.

[42] He, K., Sun, J., and Tang, X. (2010). Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353.

[43] Inc, M. R. C. (2023). Mii ret cam inc (smartphone based retina imaging device). *Online: https://www.medicalsdir.com/listing/mii-ret-cam-inc-smartphone-based-retina-imaging-device/*.

[44] Incorporated, N. M. (2023). Retcam. *https://natus.com/products-services/retcam-envision*.

[45] Jabbar, M. K., Yan, J., Xu, H., Ur Rehman, Z., and Jabbar, A. (2022). Transfer learning-based model for diabetic retinopathy diagnosis using retinal images. *Brain Sciences*, 12(5):535.

[46] Jefferies, A. L., Society, C. P., Fetus, and Committee, N. (2016). Retinopathy of prematurity: An update on screening and management. *Paediatrics & child health*, 21(2):101–104.

[47] Jobson, D. J., Rahman, Z.-u., and Woodell, G. A. (1997). Properties and performance of a center/surround retinex. *IEEE transactions on image processing*, 6(3):451–462.

[48] Joshi, S. and Karule, P. (2012). Retinal blood vessel segmentation. *International Journal of Engineering and Innovative Technology (IJEIT)*, 1(3):175–178.

[49] Joshua, A. O., Nelwamondo, F. V., and Mabuza-Hocquet, G. (2020). Blood vessel segmentation from fundus images using modified u-net convolutional neural network. *Journal of Image and Graphics*, 8(1):21–25.

[50] Kaggle (2023). Kaggle eyepacs data set. *Kaggle*.

[51] Karasulu, B. (2012). Automatic extraction of retinal blood vessels: a software implementation. *European Scientific Journal*, 8(30).

[52] Kauppi, T. et al. (2010). *Eye fundus image analysis for automatic detection of diabetic retinopathy*. Lappeenranta University of Technology.

[53] Kaur, G. (2023). components of neural network. *Online:medium*.

[54] keras (2023a). Densenet. *Keras*.

[55] keras (2023b). Developer guides/transfer learning & fine-tuning. *Keras*.

[56] keras (2023c). Inceptionresnetv2. *Keras*.

[57] keras (2023d). Keras examples. *Keras*.

[58] Keras (2023). Resnet and resnetv2. *Keras*.

[59] Khojasteh, P., Aliahmad, B., and Kumar, D. K. (2018). Fundus images analysis using deep features for detection of exudates, hemorrhages and microaneurysms. *BMC ophthalmology*, 18(1):1–13.

[60] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., et al. (2023). Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026.

[61] Kumar, V., Patel, H., Paul, K., and Azad, S. (2023). Deep learning-assisted retinopathy of prematurity (rop) screening. *ACM Trans. Comput. Healthcare*, 4(3).

[62] Le, P. and Zuidema, W. (2016). Quantifying the vanishing gradient and long distance dependency problem in recursive neural networks and recursive lstms. *arXiv preprint arXiv:1603.00423*.

[63] Lekha, T., Ramesh, S., Sharma, A., and Abinaya, G. (2019). Mii retcam assisted smartphone based fundus imaging for retinopathy of prematurity. *Indian journal of ophthalmology*, 67(6):834—839.

[64] Li, W., Xiong, X., Xia, P., Ju, L., and Ge, Z. (2024). Tp-drseg: Improving diabetic retinopathy lesion segmentation with explicit text-prompts assisted sam. *arXiv preprint arXiv:2406.15764*.

[65] Ma, J., Fan, X., Yang, S. X., Zhang, X., and Zhu, X. (2018). Contrast limited adaptive histogram equalization-based fusion in yiq and hsi color spaces for underwater image enhancement. *International Journal of Pattern Recognition and Artificial Intelligence*, 32(07):1854018.

[66] Ma, J., He, Y., Li, F., Han, L., You, C., and Wang, B. (2024). Segment anything in medical images. *Nature Communications*, 15(1):654.

[67] Madhu, K., John, S. M., Joseph, A., and Abraham, B. (2024). A study on retinopathy of prematurity (rop) screening using deep learning approaches and a blood vessel segmentation framework for rop affected eyes. In *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)*, pages 1–10. IEEE.

[68] Maji, D., Santara, A., Mitra, P., and Sheet, D. (2016). Ensemble of deep convolutional neural networks for learning to detect retinal vessels in fundus images. *arXiv preprint arXiv:1603.04833*.

[69] Matsuo, Y., LeCun, Y., Sahani, M., Precup, D., Silver, D., Sugiyama, M., Uchibe, E., and Morimoto, J. (2022). Deep learning, reinforcement learning, and world models. *Neural Networks*.

[70] Messidor (2023). Messidor data set. *Messidor*.

[71] Mo, M. (2022). Unsupervised learning - a systematic literature review. *OSF Preprints*.

[72] Mohan, N. J., Murugan, R., Goel, T., and Roy, P. (2022). Vit-dr: Vision transformers in diabetic retinopathy grading using fundus images. In *2022 IEEE 10th region 10 humanitarian technology conference (R10-HTC)*, pages 167–172. IEEE.

[73] Mookiah, M. R. K., Acharya, U. R., Chua, C. K., Lim, C. M., Ng, E., and Laude, A. (2013). Computer-aided diagnosis of diabetic retinopathy: A review. *Computers in biology and medicine*, 43(12):2136–2155.

[74] Mulay, S., Ram, K., Sivaprakasam, M., and Vinekar, A. (2019). Early detection of retinopathy of prematurity stage using deep learning approach. In *Medical Imaging 2019: Computer-Aided Diagnosis*, volume 10950, pages 758–764. SPIE.

[75] Nashat, A. A. and Hussain Hassan, N. (2016). Image compression based upon wavelet transform and a statistical threshold. In *2016 International Conference on Optoelectronics and Image Processing (ICOIP)*, pages 20–24.

[76] Nazih, W., Aseeri, A. O., Atallah, O. Y., and El-Sappagh, S. (2023). Vision transformer model for predicting the severity of diabetic retinopathy in fundus photography-based retina images. *IEEE Access*, 11:117546–117561.

[77] Nguyen, U. T., Bhuiyan, A., Park, L. A., and Ramamohanarao, K. (2013a). An effective retinal blood vessel segmentation method using multi-scale line detection. *Pattern recognition*, 46(3):703–715.

[78] Nguyen, U. T., Bhuiyan, A., Park, L. A. F., Kawasaki, R., Wong, T. Y., and Ramamohanarao, K. (2013b). Automatic detection of retinal vascular landmark features for colour fundus image matching and patient longitudinal study. In *2013 IEEE International Conference on Image Processing*, pages 616–620. IEEE.

[79] Nielsen, M. A. (2015). *Neural networks and deep learning*, volume 25. Determination press San Francisco, CA, USA.

[80] of Opthalmology, A. A. (2023). Eye Anatomy: Parts of the Eye and How We See — aao.org. *Online: https://www.aao.org/eye-health/anatomy/parts-of-eye/*.

[81] of Toronto, U. (2022). Preventing overfitting. *Lecture Notes*.

[82] on Ophthalmology, S., of Pediatrics, A. A., of Ophthalmology, A. A., for Pediatric Ophthalmology, A. A., and Strabismus (2006). Screening examination of premature infants for retinopathy of prematurity. *Pediatrics*, 117(2):572–576.

[83] opencv (2023a). Geometric image transformations. *OpenCV*.

[84] opencv (2023b). Image processing. *OpenCV*.

[85] Osareh, A., Mirmehdi, M., Thomas, B., and Markham, R. (2002). Classification and localisation of diabetic-related eye disease. In *Computer Vision—ECCV 2002: 7th European Conference on Computer Vision Copen-*

hagen, Denmark, May 28–31, 2002 Proceedings, Part IV 7, pages 502–516. Springer.

[86] OSAREH, A. and Shadgar, B. (2009). Automatic blood vessel segmentation in color images of retina. *IRANIAN JOURNAL OF SCIENCE AND TECHNOLOGY TRANSACTION B-ENGINEERING*.

[87] Ouali, Y., Hudelot, C., and Tami, M. (2020a). An overview of deep semi-supervised learning. *CoRR*, abs/2006.05278.

[88] Ouali, Y., Hudelot, C., and Tami, M. (2020b). An overview of deep semi-supervised learning. *arXiv preprint arXiv:2006.05278*.

[89] Pandey, S. K., Shukla, A., Bhatia, S., Gadekallu, T. R., Kumar, A., Mashat, A., Shah, M. A., and Janghel, R. R. (2023). Detection of arrhythmia heartbeats from ecg signal using wavelet transform-based cnn model. *International Journal of Computational Intelligence Systems*, 16(1):80.

[90] Peli, E. and Peli, T. (1989). Restoration of retinal images obtained through cataracts. *IEEE transactions on medical imaging*, 8(4):401–406.

[91] Picsellia (2023). Sam and foundation models in computer vision. *picsellia*.

[92] Pizer, S. M., Amburn, E. P., Austin, J. D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J. B., and Zuiderveld, K. (1987). Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368.

[93] Polikar, R. (2006). Engineers ultimate guide to wavelet analysis - wavelet tutorial. *Online: https://cseweb.ucsd.edu/ baden/Doc/wavelets/polikar-wavelets.pdf*.

[94] Poplin, R., Varadarajan, A. V., Blumer, K., Liu, Y., McConnell, M. V., Corrado, G. S., Peng, L., and Webster, D. R. (2018). Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature biomedical engineering*, 2(3):158–164.

[95] Priya, R. and Aruna, P. (2010). Automated classification system for early detection of diabetic retinopathy in fundus images. *International Journal of Applied Engineering Research*, 1(3).

[96] Quinn, G. E. (2016). Retinopathy of prematurity blindness worldwide: phenotypes in the third epidemic. *Eye and brain*, pages 31–36.

[97] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. (2021). Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR.

[98] Rahim, S., Sabri, K., Ells, A., Wassyng, A., Lawford, M., Chu, L., and He, W. (2024). Novel fundus image preprocessing for retcam images to improve deep learning classification of retinopathy of prematurity.

[99] Rahim, S. S., Palade, V., Shuttleworth, J., and Jayne, C. (2014). Automatic screening and classification of diabetic retinopathy fundus images. In *Engineering Applications of Neural Networks: 15th International Conference, EANN 2014, Sofia, Bulgaria, September 5-7, 2014. Proceedings 15*, pages 113–122. Springer.

[100] Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., and Sutskever, I. (2021). Zero-shot text-to-image generation. In *International conference on machine learning*, pages 8821–8831. Pmlr.

[101] Reid, J. E. and Eaton, E. (2019). Artificial intelligence for pediatric ophthalmology. *Current opinion in ophthalmology*, 30(5):337–346.

[102] Remington, L. A. (2012). Clinical anatomy and physiology of the visual system, ; remington.

[103] Ren, K., Chang, L., Wan, M., Gu, G., and Chen, Q. (2022). An improved u-net based retinal vessel image segmentation method. *Heliyon*, 8(10).

[104] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer.

[105] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252.

[106] Sabri, K. (2016). Global challenges in retinopathy of prematurity screening: modern solutions for modern times. *Pediatrics*, 137(1).

[107] Sadowsky, J. (1994). The continuous wavelet transform: A tool for signal investigation and understanding. *Johns Hopkins APL Technical Digest*, 15:306–306.

[108] Sami, S. B., Muqeet, A., and Tariq, H. (2019). A novel image dehazing and assessment method. In *2019 8th International Conference on Information and Communication Technologies (ICICT)*, pages 72–77. IEEE.

[109] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61:85–117.

[110] Sheet, S. S. M., Tan, T.-S., As'ari, M., Hitam, W. H. W., and Sia, J. S. (2022). Retinal disease identification using upgraded clahe filter and transfer convolution neural network. *ICT Express*, 8(1):142–150.

[111] Sifat, M. B. U., Pathan, N. S., Kibria, J. B., Mohammad, N., and Bhattacharjee, S. (2022). Retinal blood vessels segmentation from fundus images using u-net. In *2022 12th International Conference on Electrical and Computer Engineering (ICECE)*, pages 68–71. IEEE.

[112] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[113] Soares, J. V., Leandro, J. J., Cesar, R. M., Jelinek, H. F., and Cree, M. J. (2006). Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. *IEEE Transactions on medical Imaging*, 25(9):1214–1222.

[114] Sonka, M., Hlavac, V., and Boyle, R. (2014). *Image processing, analysis, and machine vision*. Cengage Learning.

[115] Stapor, K., Świtonski, A., Chrástek, R., and Michelson, G. (2004). Segmentation of fundus eye images using methods of mathematical morphology for glaucoma diagnosis. In *Computational Science-ICCS 2004: 4th International Conference, Kraków, Poland, June 6-9, 2004, Proceedings, Part IV 4*, pages 41–48. Springer.

[116] STARE (2023). Stare data set. *STARE*.

[117] Stephen, K. J. (2022). *2022-2023 Basic and Clinical Science Course, Section 12, Retina and Vitreous*. American Academy of Opthalmology.

[118] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9.

[119] Tan, M. and Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR.

[120] Taspinar, A. (2018). Ml fundamentals - a guide for using the wavelet transform in machine learning. *Online: https://ataspinar.com/2018/12/21/a-guide-for-using-the-wavelet-transform-in-machine-learning/*.

[121] Tatsumi, T. (2008). Clinical application of matched filter processing of digital color fundus images. *Investigative Ophthalmology & Visual Science*, 49(13):1840–1840.

[122] Team, O. (2023). Opencv documentation. *https://docs.opencv.org/2.4/doc/tutorials/imgproc/histograms/*.

[123] Tjandrasa, H., Wijayanti, A., and Suciati, N. (2012). Optic nerve head segmentation using hough transform and active contours. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 10(3):531–536.

[124] Tolias, Y. A. and Panas, S. M. (1998). A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering. *IEEE Transactions on Medical Imaging*, 17(2):263–273.

[125] Tong, Y., Lu, W., Deng, Q.-q., Chen, C., and Shen, Y. (2020). Automated identification of retinopathy of prematurity by image-based deep learning. *Eye and Vision*, 7:1–12.

[126] Valikodath, N., Cole, E., Chiang, M., Campbell, J., and Chan, R. (2019). Imaging in retinopathy of prematurity. *Asia-Pacific Journal of Ophthalmology*, 8(2):178–186.

[127] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

[128] Venturelli, M. (2023). The dangers behind image resizing. *Blog - Zuru Tech*.

[129] Vinekar, A. and Bhende, P. (2018). Innovations in technology and service delivery to improve retinopathy of prematurity care. *Community eye health*, 31:S20–S22.

[130] Vinekar, A., Jie, V. Y. W., Savoy, F. M., and Parthasarathy, D. R. (2021). Development and validation of a deep learning (dl)-based screening tool for 'plus disease'detection on retinal images captured through a tele-ophthalmology platform for retinopathy of prematurity (rop) in india. *Investigative Ophthalmology & Visual Science*, 62(8):3267–3267.

[131] Waheed, A., Akram, M. U., Khalid, S., Waheed, Z., Khan, M. A., and Shaukat, A. (2015). Hybrid features and mediods classification based robust segmentation of blood vessels. *Journal of medical systems*, 39:1–14.

[132] Wang, B., Wang, S., Qiu, S., Wei, W., Wang, H., and He, H. (2020). Csu-net: A context spatial u-net for accurate blood vessel segmentation in fundus images. *IEEE Journal of Biomedical and Health Informatics*, 25(4):1128–1138.

[133] Wang, J., Ju, R., Chen, Y., Zhang, L., Hu, J., Wu, Y., Dong, W., Zhong, J., and Yi, Z. (2018). Automated retinopathy of prematurity screening using deep neural networks. *EBioMedicine*, 35:361–368.

[134] WESTWAŃSKA, W., Archidiecezji, Z. S. S. R. K., im Kardynała, K., Hlonda, P. A., and RESPONDEK, J. S. (2018). Image preprocessing experiments for blood vessels detection in fundus eye images. *STUDIA INFORMATICA*, 39(3):136.

[135] Wikipedia (2023). Picture of the eye. *Online:Wikipedia*.

[136] Wongwai, P., Suwannaraj, S., and Asawaphureekorn, S. (2018). Diagnostic accuracy of a digital fundus photographic system for detection of retinopathy of prematurity requiring treatment (rop-rt). *Plos one*, 13(7):e0201544.

[137] Worrall, D. E., Wilson, C. M., and Brostow, G. J. (2016). Automated retinopathy of prematurity case detection with convolutional neural networks. In *Deep Learning and Data Labeling for Medical Applications: First International Workshop, LABELS 2016, and Second Interna-*

tional Workshop, DLMIA 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 21, 2016, Proceedings 1, pages 68–76. Springer.

[138] Wu, Q., Hu, Y., Mo, Z., Wu, R., Zhang, X., Yang, Y., Liu, B., Xiao, Y., Zeng, X., Lin, Z., et al. (2022). Development and validation of a deep learning model to predict the occurrence and severity of retinopathy of prematurity. *JAMA Network Open*, 5(6):e2217447–e2217447.

[139] Yamashita, R., Nishio, M., Do, R. K. G., and Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9:611–629.

[140] Yan, L. C., Yoshua, B., and Geoffrey, H. (2015). Deep learning. *nature*, 521(7553):436–444.

[141] Yang, Y., Li, T., Li, W., Wu, H., Fan, W., and Zhang, W. (2017). Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks. In *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20*, pages 533–540. Springer.

[142] Yildiz, V. M., Tian, P., Yildiz, I., Brown, J. M., Kalpathy-Cramer, J., Dy, J., Ioannidis, S., Erdogmus, D., Ostmo, S., Kim, S. J., et al. (2020). Plus disease in retinopathy of prematurity: convolutional neural network performance using a combined neural network and feature extraction approach. *Translational Vision Science & Technology*, 9(2):10–10.

[143] Yulong, D., Ke, D., Chunsheng, O., Yingshe, L., Yu, T., Jianyi, F., Wei, W., and Yaguang, D. (2023). Wavelets and curvelets transform for image

denoising to damage identification of thin plate. *Results in Engineering*, 17:100837.

[144] Zhang, C., Puspitasari, F. D., Zheng, S., Li, C., Qiao, Y., Kang, T., Shan, X., Zhang, C., Qin, C., Rameau, F., et al. (2023). A survey on segment anything model (sam): Vision foundation model meets prompt engineering. *arXiv preprint arXiv:2306.06211*.

[145] Zhang, S., Webers, C. A., and Berendschot, T. T. (2022). A double-pass fundus reflection model for efficient single retinal image enhancement. *Signal Processing*, 192:108400.

[146] Zhao, J., Lei, B., Wu, Z., Zhang, Y., Li, Y., Wang, L., Tian, R., Chen, Y., Ma, D., Wang, J., et al. (2019). A deep learning framework for identifying zone i in retcam images. *IEEE Access*, 7:103530–103537.

[147] Zhou, Y., Wang, B., Huang, L., Cui, S., and Shao, L. (2020). A benchmark for studying diabetic retinopathy: segmentation, grading, and transferability. *IEEE Transactions on Medical Imaging*, 40(3):818–828.

[148] Zhu, Q., Mai, J., and Shao, L. (2014). Single image dehazing using color attenuation prior. In *BMVC*.

[149] Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. *Graphics gems*, pages 474–485.

Appendix A

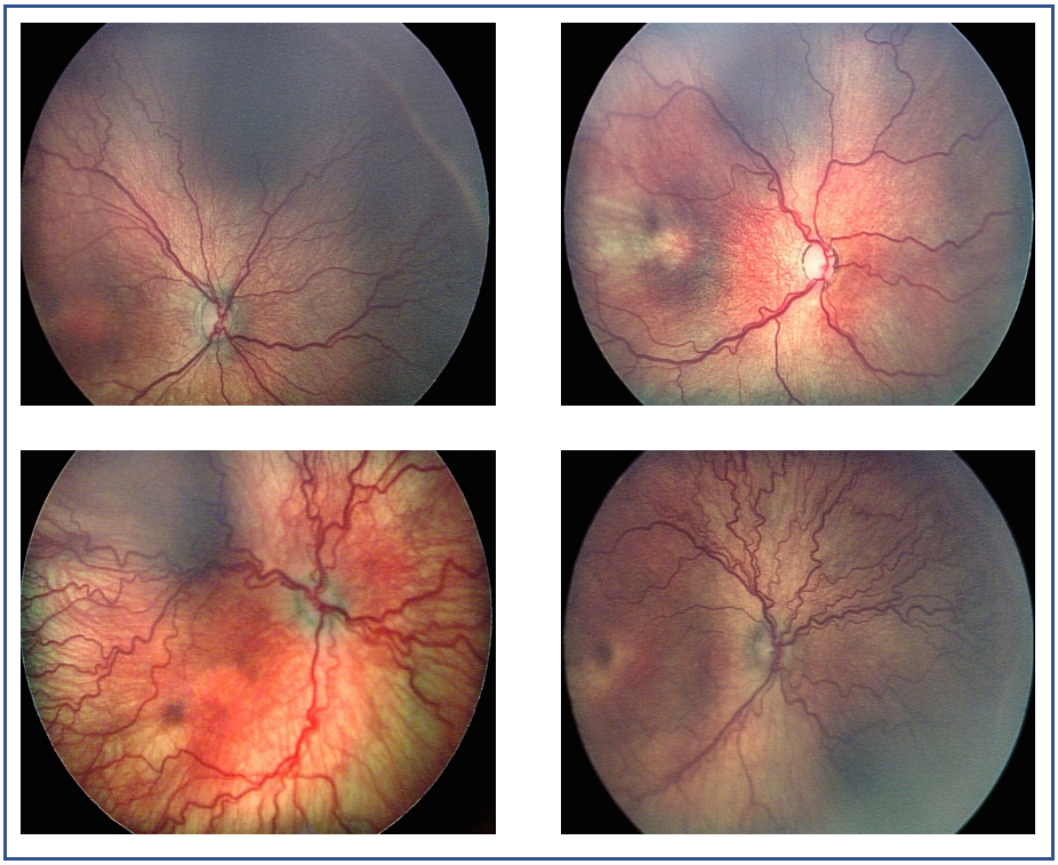
Appendix

A.1 Images samples - Plus diseases, Stages, Zones

Below are samples for Plus Disease, Stages 0-3 and Zones I-III that are part of the training dataset.

A.2 Plus Disease

Figure A.1: **Plus Disease**



A.3 Stages 0-3

Figure A.2: **Stage 0**

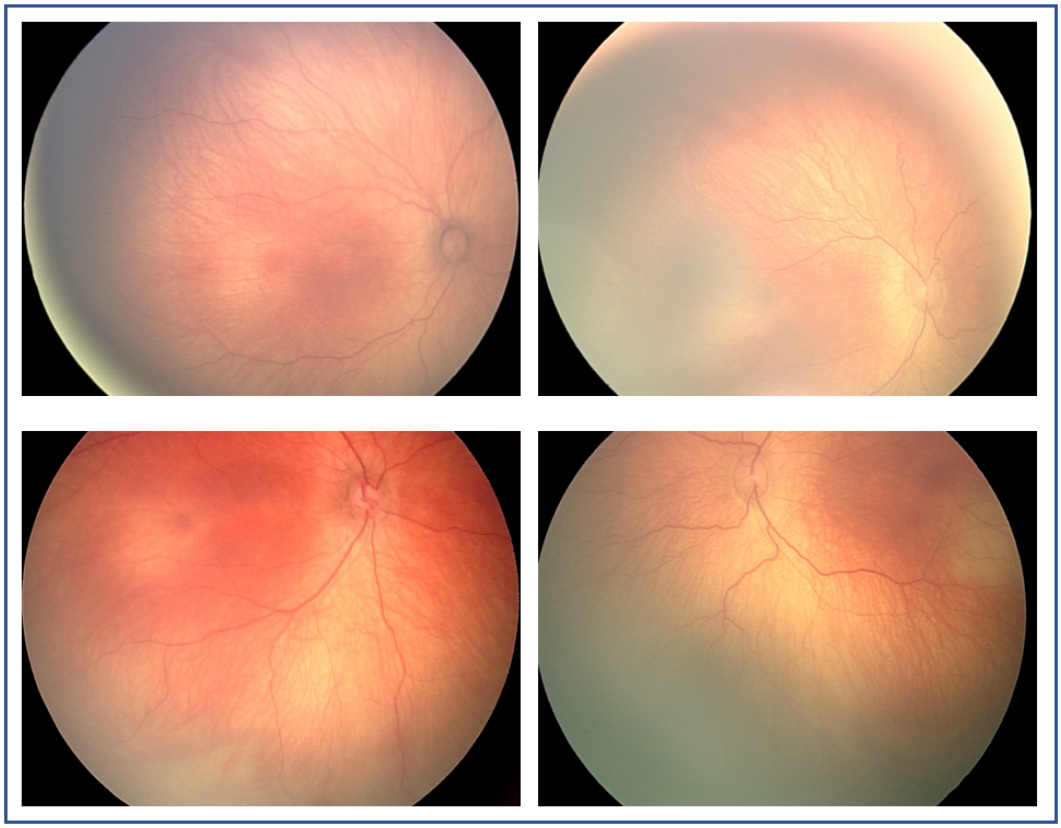


Figure A.3: **Stage 1**

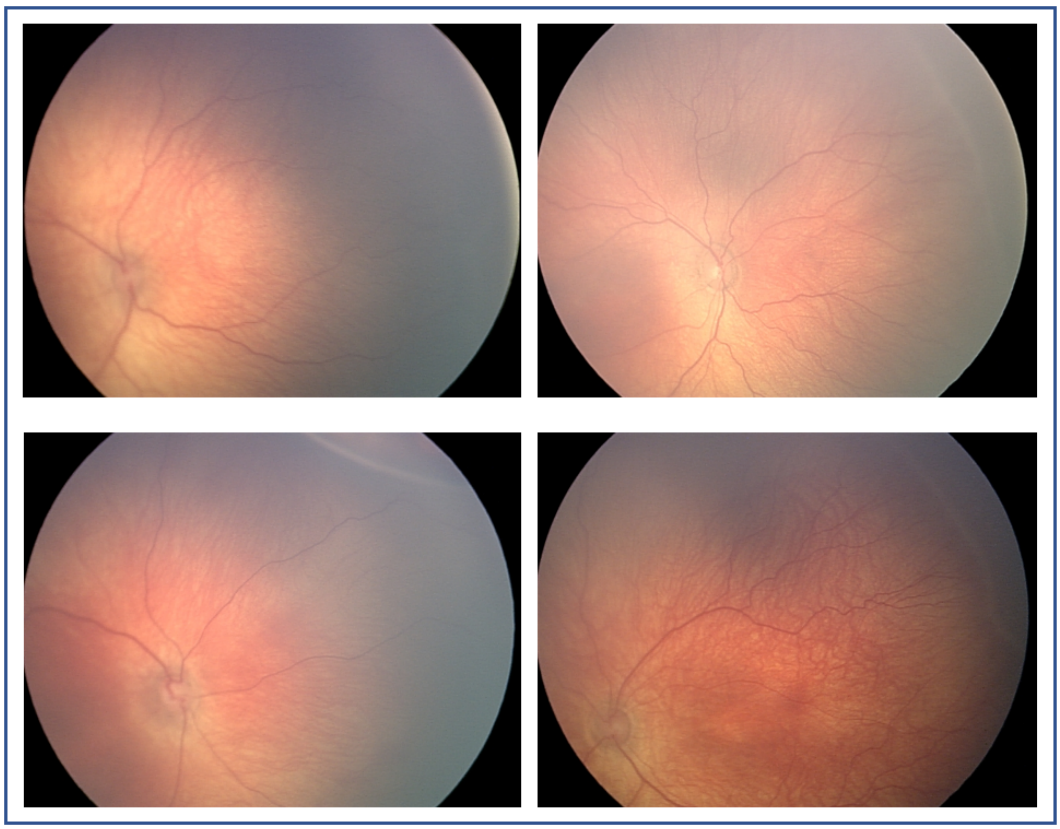


Figure A.4: **Stage 2**

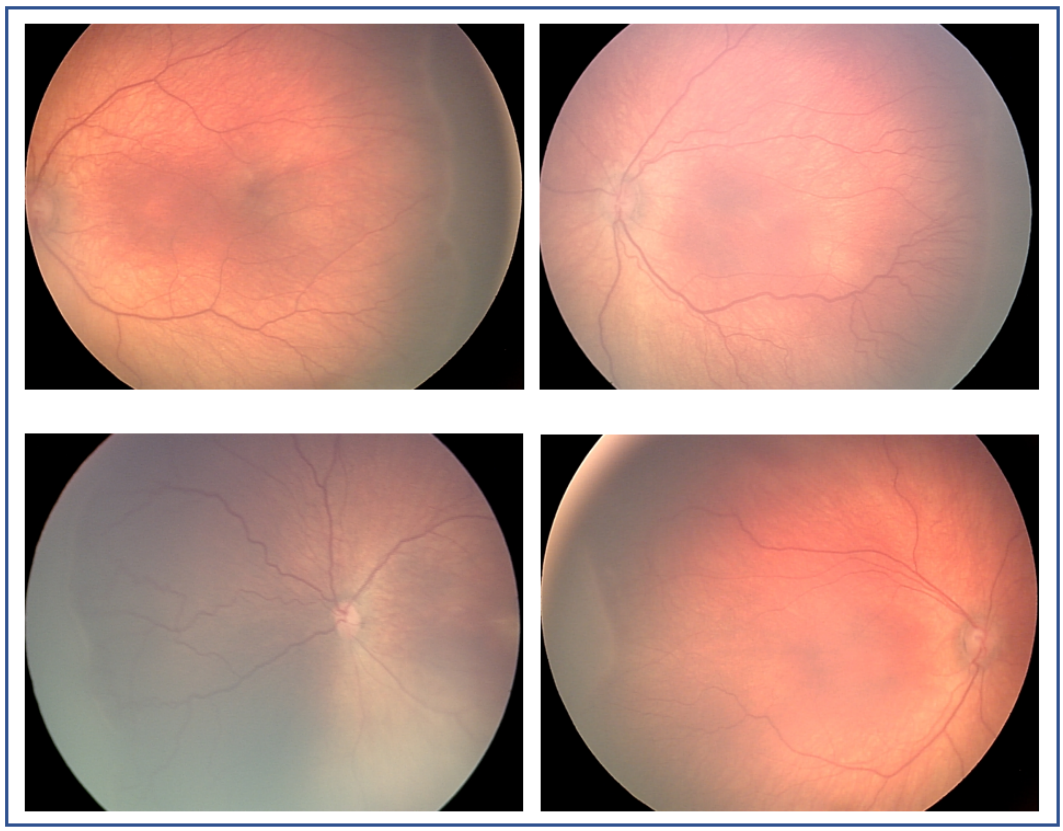
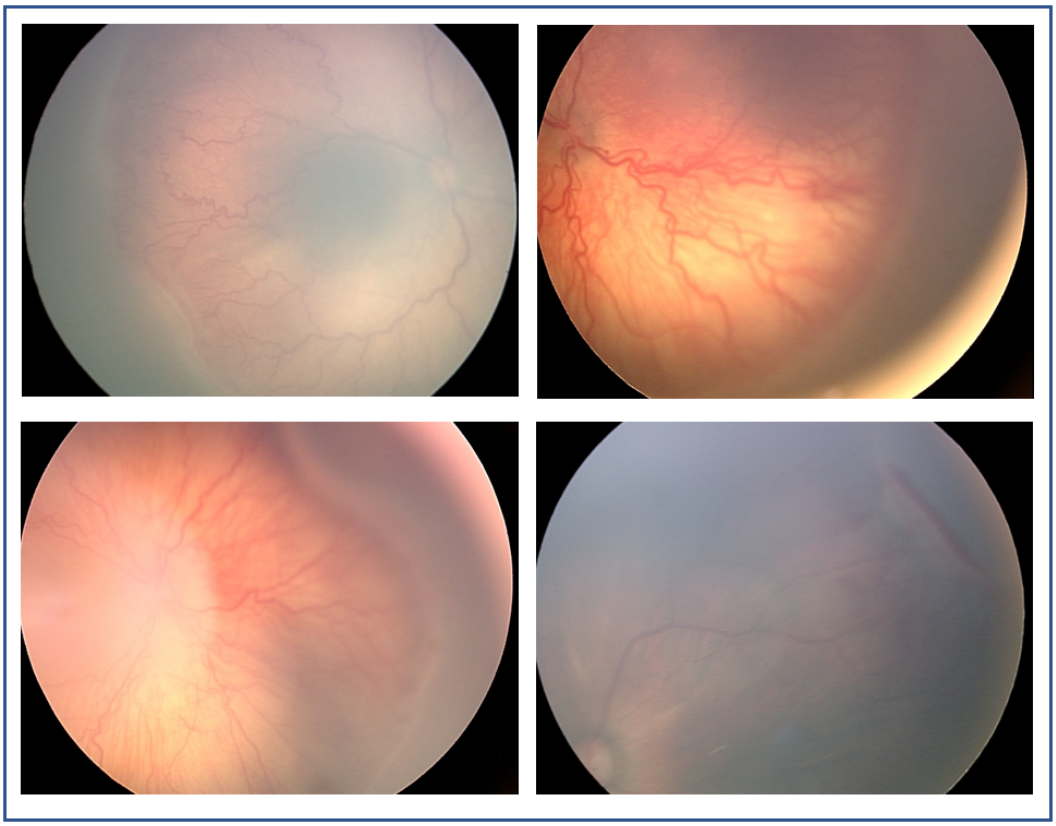


Figure A.5: **Stage 3**



A.4 Zones I-III

Figure A.6: **Zone I**

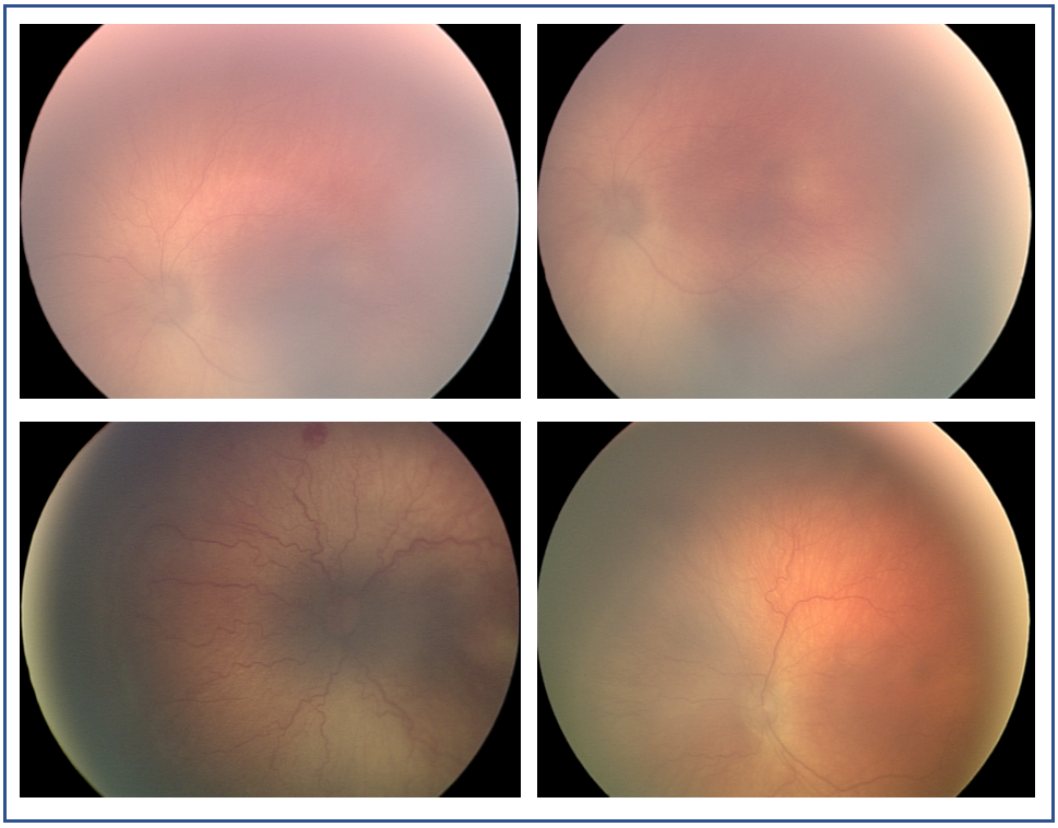


Figure A.7: **Zone II**

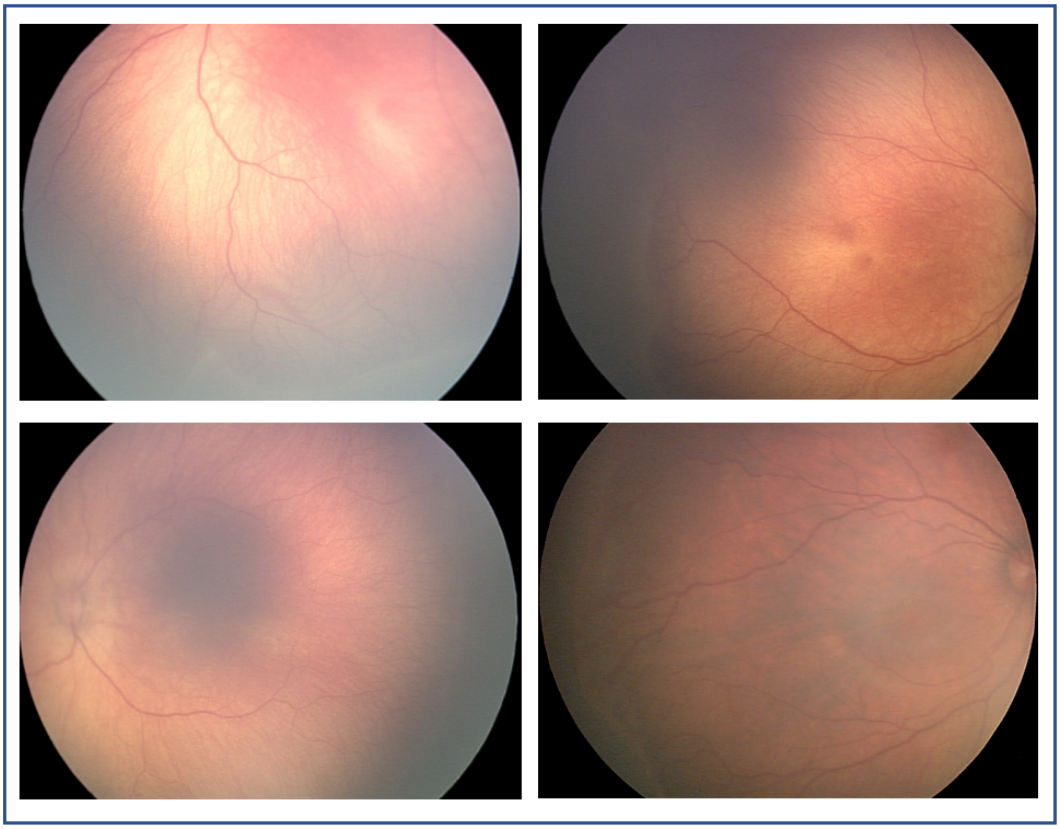
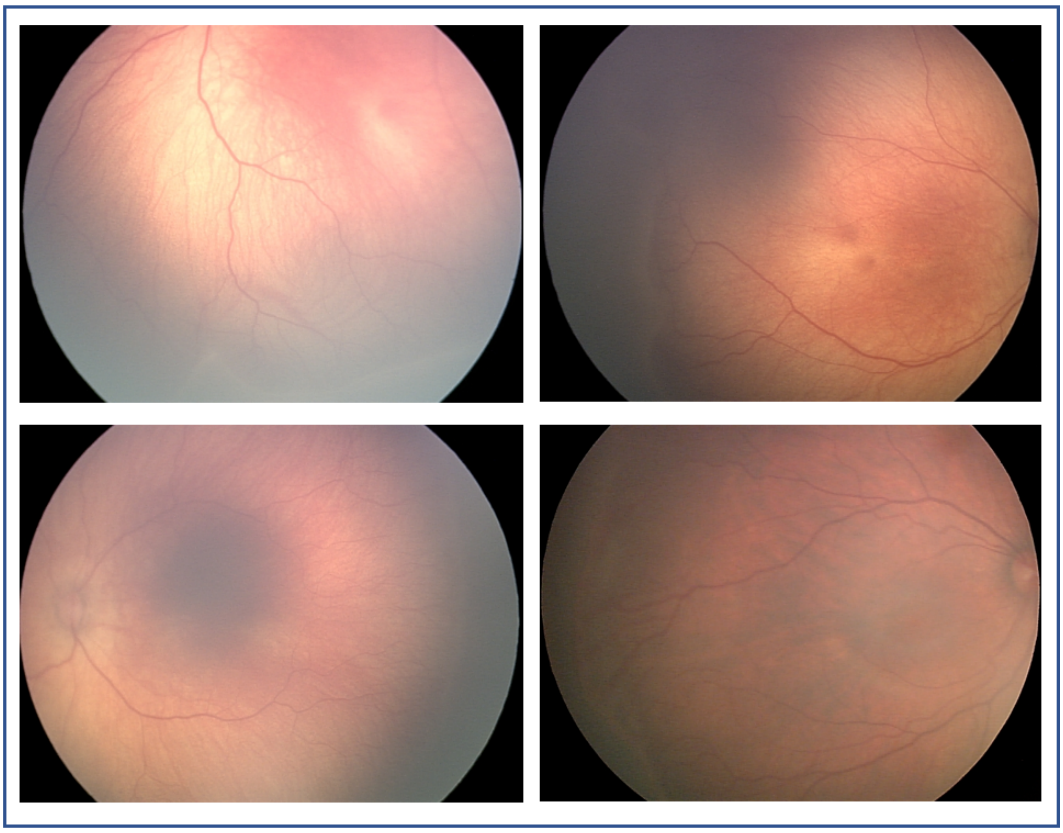


Figure A.8: **Zone III**



A.5 Section in Appendix

More text

A.6 Another Section

Even more text

Appendix B

Website Sample Appendix

Below is the website www.screenropimages.ca folder setup which has ground truth folders for:

- Plus + No Plus + Pre-plus images
- Stages: 0,1,2,3,4,5
- Zones: I, II, III

Figure B.1: **Website www.screenropimages.ca**

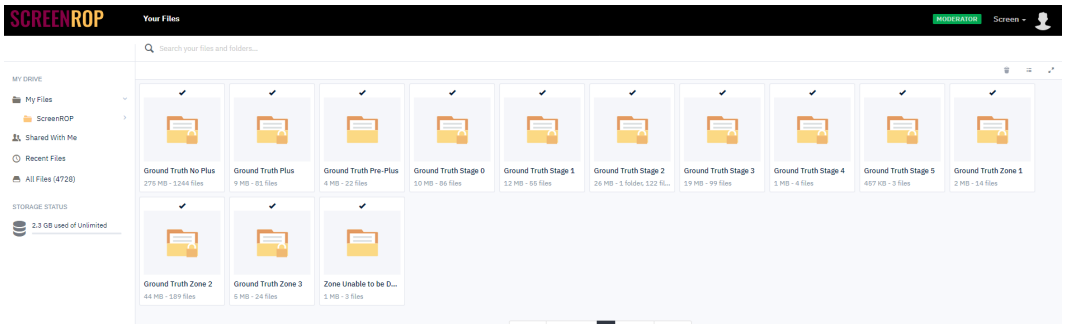
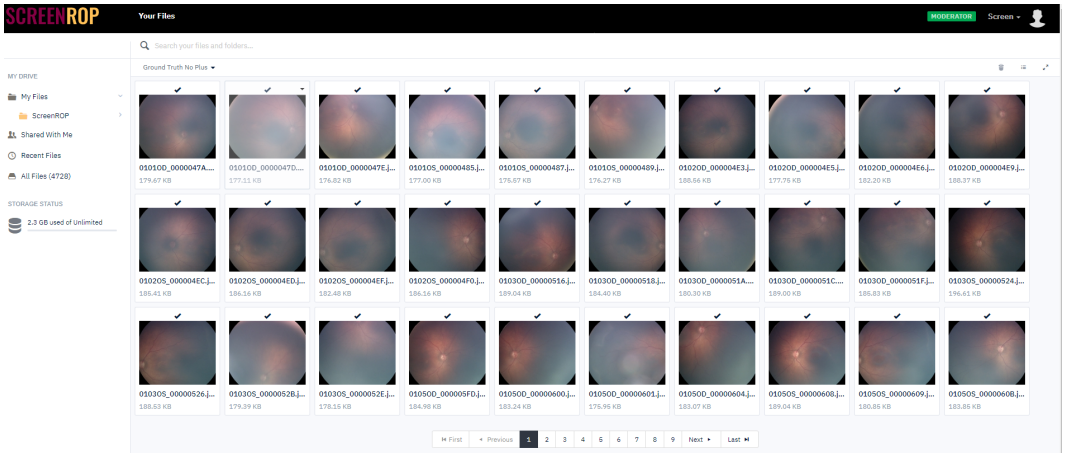


Figure B.2: **Website - No Plus disease folder details**



Appendix C

Code Appendix

Code for this project is uploaded on following repository.
https://gitlab.cas.mcmaster.ca/rahims9/retinopathy-of-premature/code