Prognosis Prediction in Retinal Vein Occlusion

By

Sumeia Ahmed A Elkazza

Student Number: 190614795

A thesis submitted in partial fulfilment of the requirements of the Newcastle University for the degree of Doctor of Philosophy.

Research undertaken in the School of Engineering/ Newcastle University.

Under the supervision of:

Prof. Satnam Dlay

Dr. Jeffry Hogg

May 2023

DECLARATION

I, Sumeia Ahmed A Elkazza, hereby declare that this thesis is my own work and effort and that it has not been submitted anywhere for any award.

Student name: Sumeia Ahmed A Elkazza.

Student signature: Sumeia Elkazza.

Date: 18/05/2023

Abstract

Retinal vein occlusion (RVO) represents the second highest cause of retinal vascular blindness after diabetic retinopathy. The most common cause of vision loss in eyes associated with RVO is macular oedema (MO). In many cases, MO can be successfully treated by intravitreal injections of VEGF agents which has become widespread to improve clinical outcomes in RVO patients. Nevertheless, many RVO patients show a well response, poorly response, or no response to the treatment. This has serious implications, as the treatment is costly and has risks to blindness. It would be better for patients and more efficient use of National Health Service (NHS) funds to avoid unnecessary injections. Making-decisions to continue the treatment is a challenging. Most previous studies have focused on diagnosing and classifying RVO. This work presents four novel methods that can help using computer aids to make decisions in RVO. These novel methods predict visual acuity (VA) after one year of anti-VEGF treatment for RVO patients. In the first method, linear regression and Random Forest regression (RF) were performed to improve treatment discussions with patients using Electronic Medical Records (EMRs). It is proved that linear regression produced a model accounting for 57% of the accuracy seen in 1-year VA. Using the same data, RF surpassed this, with the model accounting for 62% of the accuracy. The second method investigated different machine learning techniques to compare with linear regression, and two feature extraction techniques were considered to analyse Optical Coherence Tomography (OCT) images and combined with EMRs. It has been shown that machine learning methods overcome linear regression, and RF Regressor performs best with its highest score, R^2= 0.75. Thirdly, pre-processing OCT images were presented using morphological transformation and sharpened images. Then, OCT scans were analysed using feature extraction techniques and fused with EMR to generate the input data fed to the CNN model. To evaluate the model's performance, 11 ophthalmology doctors provided similar VA forecasts for two subsets (n=41) of the model testing data set (n=82). The CNN proposed method outperformed all the doctors, with an MAE of 13.65. Fourthly, two binary classifications were trained to predict the probability of an eye showing High and Low anti-VEGF demand using three machine learning methods. In addition, the Random Forest regressor model is analysed and designed to predict anti-VEGF demand after 12 months. Furthermore, High-demand treatment was determined with reasonable accuracy already at baseline before the start of treatment with AUC=0.78 and AUC =0.64 for Low-demand using a Random Forest classifier which demonstrated a higher AUC for prediction using baseline information. In summary, novel methods to address the problem are proposed using several machine learning techniques and the CNN model for treatment discussion. By evaluating the results with the number of ophthalmologists, this work can be considered a dependable prognostic technique for patients deciding whether to commit to treatment.

Research Achievements

- Elkazza SA, Hogg HJ, Di Simplicio S, Dlay SS. *Comment on: Ranibizumab and aflibercept intravitreal injection for treatment naïve and refractory macular oedema in branch retinal vein occlusion. European Journal of Ophthalmology*. 2022;32(1):NP303-NP304. doi:10.1177/1120672120945103.

- Elkazza SA, Hogg HJ, Di Simplicio S, Dlay SS. *Machine learning techniques increase the accuracy of visual outcome prediction in retinal vein occlusion complicated by macular oedema. Oxford Ophthalmological Congress,* OOC 2020 Abstracts e-Booklet. P.36, available online at: https://www.ooc.uk.com/pdf/Oxford-Ophthalmological-Congress-e-Booklet.pdf.

- Elkazza SA, Grace George, North East Trainee Research in Ophthalmology Network, Di Simplicio S, Hogg HJ, Dlay SS. *Random Forest model matches clinician performance in forecasting visual outcomes from anti-VEGF treatment in retinal vein occlusion. Revolutionary eye and vision research (ARVO) 2021*. May 2021, available online at : https://arvo2021.arvo.org/people/mTSBfsDjDasdLL2D5.

- Elkazza SA, Hogg, HDJ, Di Simplicio S, Dlay SS, *Machine learning predicts three-quarters of the variability of individual prognosis from anti-VEGF treatment in retinal vein occlusion*, EURETINA 2020, available online at: https://euretina.conference2web.com/#!resourcegroups/order=primary_event_starts_at&page=1&query=elkazza.

- Submitted in Journal: “Artificial Intelligence in Medicine”. Title: “*Hybrid Method and CNN Regressor for OCT images to Predict Visual Acuity of Retinal Vein Occlusion Patients After 12 Months of Treatment.*

Acknowledgements

Firstly, I would like to convey my deep gratitude to my lord, Almighty ALLAH, for generously giving me endless precious bounties. Without his guidance and mercy, this work would have never achieved this stage. I would like to state genuine gratitude to my supervisors Prof. Satnam Dlay and Dr. Jeffry Hogg for the generous provision of their time and advice, which paved the way during my study. To my mother, a strong and gentle soul who taught me to trust in Allah and believe in hard work. I could never have done this without your support and constant encouragement. Thank you for teaching me to believe in myself and my dream. To my father for earning an honest living for us and supporting and encouraging me to believe in myself. I know you would be so proud if you could see me now. I owe a very important debt to my mother-in-law, who has given me endless love and support throughout my study. My heartfelt appreciation goes to my husband Dr. Ashref Lawgaly, for his support and great patience at all times, and my special thanks to my sons, Adel, Hafeed, Ahmed, Asem; from them, I stole great moments and special days for the sake of study. My lovely family has been a constant source of love, motivation, and energy. My special thanks to my sisters and brothers for their support and encouragement. Last but not least, I would like to thank all my friends for their essential help and encouragement.

Contents

List of Acronyms

AI	Artificial Intelligence.
AMD	Age-related Macular Degeneration.
Anti-VEGF	Anti-vascular endothelial growth factor.
AUC	Area Under the Curve.
BCVA	Best corrected visual acuity.
BRAO	Branch retinal artery occlusion.
BRVO	Branch retinal vein occlusion.
CatBoost	Gradient categorical boosting Classifier.
CFT	Central Foveal Thickness.
CLAHE	Contrast-limited adaptive histogram equalization.
CME	Central macular oedema.
CNN	Convolutional neural networks.
CNV	Choroidal neovascularization.
CRAO	Central retinal artery occlusion.
CRUISE	Central retinal vein occlusion study: evaluation of efficacy and safety.
CRVO	Central retinal vein occlusion.
CSF	Central SubField.
CRVO	Compute unified device architecture.
DME	Diabetic macular edema.
DICOM	Digital imaging and communications.
DL	Deep learning.
DR	Diabetic Retinopathy.
DRIVE	Digital retinal images for vessel extraction.

DS	Data structure.
EMRs	Electronic Medical Records.
ETDRS	Early Treatment Diabetic Retinopathy Study.
EZ	Ellipsoid zone.
FA	Fluorescein-angiography.
GC	Ganglion cell layer.
GENEVA	Global evaluation of implantable dexamethasone in retinal vein occlusion with macular edema.
GH	Gradient Horizontal.
GLCM	Gray-level co-occurrence matrix.
GV	Gradient Vertical.
HLBP	Hierarchical binary pattern.
HOG	Histogram of Oriented Gradients.
HRVO	Hemiretinal vein occlusion.
HvO	High versus others.
ILM	Internal limiting membrane.
IPL	Inner plexiform layer.
IVI	Intravitreal injection.
LASSO	Least Absolute Shrinkage Selector Operator.
LBP	Local Binary Patterns.
LDA	Linear Discriminant Analysis.
LDF	Laser doppler flowmetry.
LDV	Laser doppler velocimetry.
LSFG	Laser Speckle Flow graphy.
LvO	Low versus others.

MAE	Mean Absolute Error.
ML	Machine learning.
MLP	Multi-layer Perceptron regression.
MLR	Multiple linear regression.
MO	Macular odema.
MSE	Mean Square Error.
NHS	National Health Service.
NPDR	Non-proliferative DR.
NPV	Negative predictive value.
OCT	Optical coherence tomography.
OCTA	OCT angiography.
OLS	Ordinary Least Squares regression.
ONH	Optic Nerve Head analysis.
ONL	Outer nuclear layer.
PDR	Proliferative diabetic retinopathy.
PCA	Principle Component Analysis.
PLS	Partial Least Squares Regression.
PPV	Positive predictive value.
PRP	Panretinal photocoagulation.
qCSF	Quantitative contrast sensitivity function.
QDA	Quadratic Discriminant Analysis. R
RAO	Retinal artery occlusion.
RBV	Retinal Blood Vessel.
RCT	Randomized clinical trial.

Relus	Rectified Linear Units.
RF	Random Forest.
RNFL	Retinal nerve fiber layer analysis.
RPE	Retinal pigment epithelium.
RRF	Retinal reflectivity.
RSS	Residual sum of squares.
RVI	Royal Victoria Infirmary.
RVO	Retinal vein occlusion.
SCORE	Standard care versus corticosteroid for retinal vein occlusion.
STARE	Structured analysis of the retina.
SVM	Support vector machine.
TDD	Test Driven Development.
TPA	Tissue plasminogen activator.
TIFF	Tagged Image File Format.
VA	Visual acuity.
VB	Vitreous body.
VMA	Vitreomacular adhesion.

List of Figures

List of Tables

Chapter 1 : INTRODUCTION

1.1 Motivation

Retinal Vein Occlusion (RVO) is the most common retinal vascular occlusive disorder and is usually associated with visual loss to variable degrees. The most common cause of vision loss in eyes associated with RVO is macular oedema. In many cases, macular oedema (MO) can be successfully treated or managed with intravitreal injections of anti-VEGF agents[1, 2]. However, while many patients with RVO show an excellent response to treatment, others show no or only a partial response. Ophthalmologists at present have no way of predicting who will improve and who will not. This has profound implications, as treatment is costly and an injection into the eyeball carries a risk of blinding the eye in every 1:1000 cases. It would be better for patients and make more efficient use of NHS funds to avoid injections that are not necessary. Following an extensive literature review, it appears that most previous studies have focused on the diagnosis and classification of RVO. This may not be particularly useful since it is usually an easy diagnosis to make clinically[3]. Moreover, in many recent studies, the volumes of data exploited[4, 5]were relatively small, leading to low statistical power in confirming any robust prediction of vision. Additionally, methods involved in informing risk prediction are relatively complex. Therefore, it is unrealistic to expect ophthalmologists to replicate these methodologies for each patient, and such methods may not benefit them in their current state. Alternatively, if each optical coherence tomography (OCT) comes to a clinician with information about risk in terms of biomarkers, this could be more useful. Clinicians widely use OCT images, and they play an essential role in predicting the visual outcome in RVO. The proposed study applies novel methods using EMR and OCT scans for prognosis prediction that can help clinicians in decision-making regarding RVO. The novel proposed methods are investigated which uses feature extraction methods to predict prognosis.

1.2 Scope of Research

The present research concerns prognosis prediction that can help clinicians in decision making regarding RVO according to regression of visual acuity (VA) after 12 months of anti-VEGF treatment for RVO patients at the Newcastle Eye Centre, Royal Victoria Infirmary, Newcastle Upon Tyne, UK. Two types of RVO, central and branch, are considered in this study. Several machine learning algorithms and CNN algorithms are used in conducting the research.

1.3 Aims and objectives of the Work

The aim of this thesis is to investigate a novel deep learning architecture for prognosis prediction that can help clinicians in decision-making regarding RVO treatment. The objectives of this thesis are as follow:

1. To accumulate OCT scans from the eye clinic at the Newcastle Eye Centre at the Royal Victoria Infirmary, Newcastle Upon Tyne, with OCT images acquired during routine treatment with anti-VEGF injections and graded at months 1, 6, and 12 each patient's visit. This objective is covered in Chapter 4.
2. To analyse and design a new method using learning regression and machine learning technique to predict 1-year VA during anti-VEGF treatment for RVO patients from baseline data only. This objective is covered in Chapter 5.
3. To examine several machine learning approaches, including Ridge Regression[6], LASSO Regression[7], Multi-layer Perceptron (MLP) Regression Neural Network[8], PLS Regression[9], and Random Forest Regression[10], and feature extraction methods to analyse biomarkers in OCT scans and electronic medical record (EMR) datasets. This objective is covered in Chapter 6.
4. To develop a new method to improve the regression performance for the prediction of 1-year VA in anti-VEGF treatment for RVO patients using the CNN regression approach and a hybrid feature extraction method in the analysis of OCT scans. This objective is covered in Chapter 7.
5. To analyse and design a new method using machine learning approaches to predict high and low anti-VEGF demand after 12 months of treatment for RVO patients. This objective is covered in Chapter 8.

1.4 Contributions

This thesis offers main contributions that can be summarised as follows:

- This study introduces a new method to predict 1-year VA at baseline, which increases the accuracy of prognosis in RVO complicated by MO and improves treatment discussions with patients. Linear regression and Random Forest regressions were performed to predict 1-year VA from baseline data.

- Another new method is used to predict 1-year VA using five machine learning techniques. Optical coherence tomography (OCT) stacks taken at baseline were collected for each case, and OCT slice was identified for analysis using histograms of oriented gradients (HOG) and local binary patterns (LBP). OCT slices were then vectorised and added for analysis. Linear regression of all independent variables established the performance of traditional statistical techniques to predict 1-year visual outcomes. The following machine learning (ML) techniques were then applied: Random Forest regression, Ridge regression, the multi-layer perceptron (MLP) regression neural network, Partial least squares regression (PLS), and LASSO regression.

- A new method is used to investigate a CNN technique for the prediction of VA after 12 months of anti-VEGF treatment. OCTs scans were pre-processed. A hybrid method using HOG, LBP, and a grey level co-occurrence matrix (GLCM) to extract features from OCT images was examined. The regression system was studied by employing different local feature representations of OCT images. These features were then combined to improve the recognition rate. The hybrid feature provides a good representation of OCT images to complement EMR. It is input to the CNN regressor, which generates a forecast of VA after one year of anti-VEGF treatment.

- A new machine learning method predicts high and low anti-VEGF demand in patients with RVO. OCT images were cropped and changed to grayscale, and LBP, HOG, and GLCM extraction methods were used for OCT analysis. Hybrid features and EMR were merged and used as input for training and testing the model, where the anti-VEGF injections after one year represent the target of the model. Random Forest regression

was investigated to predict anti-VEGF demand after 12 months of therapy. In addition, two binary classifications were trained to predict the probability of an eye showing high demand (HvO) and low demand (LvO). Several classification methods were implemented, including Logistic regression and the Random Forest and CatBoost classifiers.

1.5 Thesis Outline

This dissertation is organised as follows. Chapter 1 includes the research motivation and scope, the aims and objectives of the work, and the research contributions. The rest of the thesis is arranged in eight chapters as follows.

Chapter 2: Research Background: The Eye and the Retina

This chapter presents an overview of the human eye's anatomical components, and the retina's structure and function are explained. A description of the retina vascular occlusion decreases, including RVO is provided. Risk factors for RVO are also reported and its clinical presentations are shown before diagnostic imaging techniques and treatments for RVO are discussed.

Chapter 3: Literature Review

This chapter presents the existing techniques for the detection and analysis of RVO. These techniques are categorised to base on the techniques used for RVO analysis medical studies, machine learning, deep learning, machine learning technique for prediction Anti-VEGF treatment demand, existing OCT database, and CNN.

Chapter 4: Data Set

This chapter describes the data sets utilised in the research which represent EMR and optical coherence tomography (OCT) scans.

Chapter 5: Visual Prognosis Prediction in Retinal Vein Occlusion using Learning Regression and Machine Learning Approach

This chapter briefly introduces learning regression (LR) and Random Forest regression (RF). Challenges in the prognosis prediction of VA after 1-year of anti-VEGF treatment for RVO patients using LR and RF models is evaluated using a baseline data set.

Chapter 6: Predicting the Visual Acuity of RVO Patients after 1 Year using Several Machine Learning Techniques and Feature Extraction methods

In this chapter, a brief introduction is given of several machine learning approaches. Random Forest regression, Ridge regression, the multi-layer perceptron (MLP) regression neural network, Partial least squares regression (PLS), and LASSO regression are described. An overview of HOG and LBP feature extraction methods are also provided. Finally, an evaluation of these machine learning techniques for prognosis prediction of VA after 12 months of anti-VEGF treatment for RVO patients is discussed, and their performance with LR is compared.

Chapter 7: Predicting the Visual Acuity of RVO Patients after 12 Months of Treatment using the LBP-HOG-GLCM Hybrid and CNN

This chapter briefly introduces gray level co-occurrence matrix (GLCM) feature extraction method. The architectures of convolutional neural networks (CNN) are then explored in detail. Image processing techniques, morphological transformation and sharpened images are briefly introduced. Pre-processing OCT images are analysed and a hybrid method for extracting features from OCT scans is proposed. Finally, a CNN architecture is proposed for the prognosis prediction of VA after 1 year of VEGF treatment for RVO patients.

Chapter 8: Machine Learning for the Prediction of Anti-VEGF Treatment Demand for RVO Patients

In this chapter, a brief introduction to Random Forest classifiers is provided. Then, LBP, HOG, and GLCM feature extractions are used to extract features from OCT scans and associated clinical data. Next, a Random Forest regression model to predict VEGF demand after 12 months is analysed and designed. Also, two binary classifications are trained to predict the probability of an eye showing high demand versus others (HvO) and low demand versus others (LvO).

Chapter 9: Conclusions and Future Work

This chapter summarises the main findings and conclusions drawn in this thesis and discusses future research directions that could be investigated based on this work.

Chapter 2 : RESEARCH BACKGROUND: THE EYE AND THE RETINA

2.1 Introduction

The retina is the light-sensitive tissue covering the eye's interior surface: the cornea and the lens direct light rays onto the retina. The retina then converts the light received into electrical impulses and sends them to the brain via the optic nerve. After that, a person interprets those impulses as images. The cornea and lens in the eye behave like the lens of a camera, while the retina is analogous to the film[11].

This chapter gives a brief overview of anatomy of the eye, the function of the retina and retinal vascular occlusions. In addition, pathogenesis of retinal vein occlusions (RVO), diagnostic imaging techniques for RVOs and its types, and the treatments for the RVOs will be discussed. Finally, the motivation, aims, and the contributions of this work are listed, and the structure of the thesis is outlined.

2.2 Anatomy of the Eye

Visual perception represents light that passes through the pupil of the eye. The appropriate amount of light is directed towards the lens with help of the eye's structures such as the tear film, cornea, and iris. [12]. As in Figure 2.1, the conjunctiva and cornea direct contact with the environment, their role is to protect the eye from chemical or mechanical trauma. The iris, anterior chamber, and posterior chamber lie behind the cornea, all these components are filled with or in contact with the aqueous fluid. The cornea is transparent because of its microscopic architecture: there are no blood vessels, thin-diameter collagen is laid out in a highly regular fashion, and water is actively transported out of the tissue. Whilst its main role is refraction of light, the lens separates the aqueous fluid from the vitreous, which is a gel-like fluid that occupies a majority of the eye posterior to the lens [13].

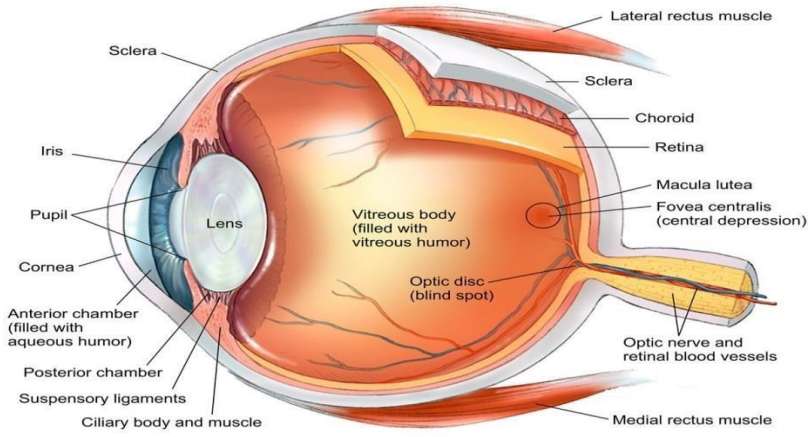


Figure 2.1:Gross anatomy of the eye [14].

2.3 Structure of the Retina

The retina includes many different types of nerve cells, a pigmented epithelium and blood vessels. It covers the internal surface of the posterior part of the eye. The retina has several layers, as shown in Figure 2.2 which can be ordered anatomically from inner-most to outer-most: The inner limiting membrane is the boundary between the vitreous body and the retina. The nerve fibre layer contains axons of the ganglion cell bodies. The inner plexiform layer contains the synapses between the dendrites and bipolar cell axons of the ganglion and amacrine cells. The inner nuclear layer contains bipolar cells, horizontal cells, and amacrine cells. The outer plexiform layer represents the first area of neuropil, where the joining between cones and rod and vertically running bipolar cells, and horizontally cells occur. The outer nuclear layer includes the cell bodies of the cones and rods. The outer limiting membrane (or external limiting membrane) is one of the ten distinct retina layers formed from adherens junctions between Müller cells and photoreceptor cell inner segments. This layer has a network-like structure and is located at the bases of the cones and rods. The Photoreceptor layer contains small photoreceptors cells located in the retina and play a vital role in night vision and affect how the eye sees colour. The photoreceptor cells are anchored to retinal pigment epithelium cells, which provide nutrients and phagocytose light-damaged cells. Due to the photoreceptors, including the outermost layer of the neurosensory retina, any loss of retinal transparency in the inner layers, for example, a breach of the blood-retinal barrier, can negatively affect vision[14].

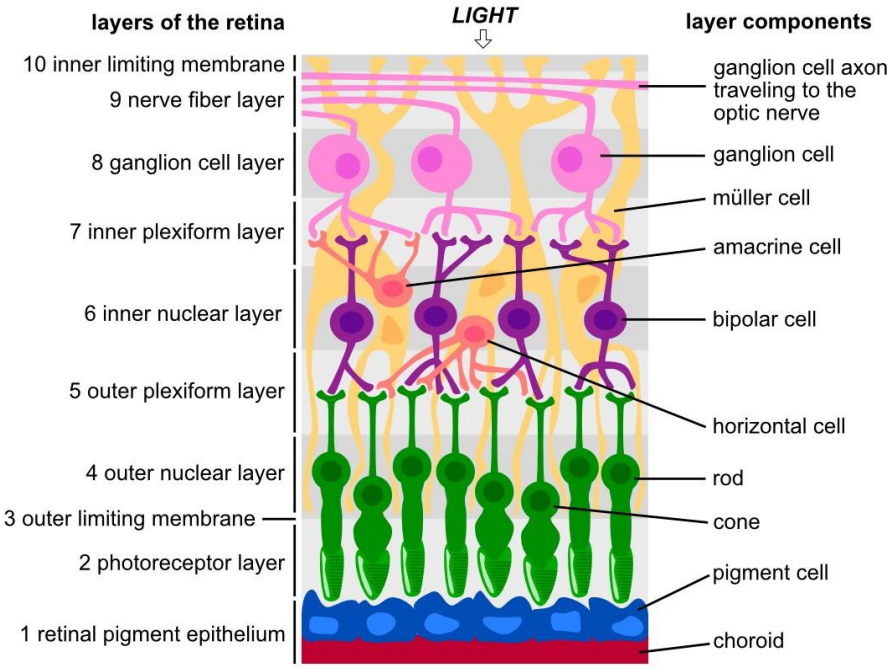


Figure 2.2 : Schematic of the microscopic structure of the retina [14].

2.4 Function of the Retina

As shown in Figure 2.2, the retina consists of millions of cells which work together to detect light. Of the previously mentioned cells in the neurosensory retina, the photoreceptor cells are responsible for transforming light into the electrical impulses which allow vision. There are two types of photoreceptor cells and both kinds transform light energy into electrical energy through structural and chemical intracellular changes. These photoreceptor types are called rods and cones. Together they permit differences between dark and light, as well as colours to be perceived. Rods are responsible for vision and recognition of movement in low light condition, whereas cones function optimally in a bright environment and are responsible for high acuity colour vision[15]. The location of these photoreceptor types also affects vision. All cones are located in the central region of the retina called the macula and are the only photoreceptor type in the central 500 microns known as the fovea. In contrast, rods are located throughout the retina except in the fovea. The function of the macula is to support central vision, while the rest of the retina provides circumferential vision. The optic disc is the exit point for ganglion cell axons, which come together to form the optic nerve as they leave the eye. The optic disc acts as the beginning of the optic nerve and is the point where the axons of retinal ganglion cell joined together.

2.5 Retinal Vasculature

Before birth, the most superficial or inner-most retinal vascular layer is created first, starting from the optic nerve head and growing toward the peripheral edge of the retina. The vessels reach the nasal extreme of the retina by 36 weeks of gestation. When this most superficial layer is complete, retina vessels develop deeper/outward into the retina to form the deep retinal vascular layer at the outer-most aspect of the outer plexiform layer. The retina's blood supply arrives from two sources derived from the ophthalmic artery. The posterior outer third of retinal layers and the entire fovea thickness are supplied by the network of capillaries within the choriocapillaris of the choroid. The central retinal artery supplies nutrients and oxygen to the anterior, inner two-thirds of the retina from the nerve fibre layer to the outer plexiform layer. The main retinal artery enters the eye within the optic nerve and ultimately branches into capillary networks, which are eventually drained by the central retinal vein. The central retinal vein exits the eye adjacent to the main retinal artery in the optic nerve. A non-fenestrated epithelium bounds the capillaries within these networks between the retinal vein and artery with tight junctions to form the inner blood-retinal barrier [16].

2.6 Retinal Vascular Occlusions

The vascular system consists of blood vessels called veins and arteries, the role of the vessels is to transport blood throughout the body, including the eyes. To make the cells within the retina obtain enough oxygen and nutrients, the retina requires a constant supply of blood. However, in some pathological circumstances these vessels carrying blood from or to the retina can become blocked. This is known as a vascular occlusion. Retinal vascular occlusions are separated to two main categories determined by the vessel types: Retinal Arterial Occlusions and Retinal Vein Occlusions[17].

2.6.1 Pathogenesis of Retinal Arterial Occlusion (RAO)

Retinal Artery Occlusion is the blockage of one of the retinal arteries and is the ocular analogue of cerebral stroke. The blockage is usually caused by thrombus (blood clot) or a small piece of cholesterol which blocks blood flow and is associated with painless sudden catastrophic visual loss. The blockage is divided to two categories, the more damaging of which is central retinal artery occlusion (CRAO) which occurs when the main artery supplying blood to the eye is blocked. CRAO is an ocular emergency. Patients usually present with acute, deep, painless monocular visual loss[18]. Another type is branch retinal artery occlusion (BRAO). It occurs when the branches of the arterial supply to the retina becomes occluded. The affected area is not in the centre of the retina and is comparatively small. Emboli secondary to either cardiac or carotid plaques are the most common cause by BRAO [19].

2.6.2 Pathogenesis of Retinal Vein Occlusion (RVO)

RVO occurs when a vein in the retina becomes blocked. This can lead to backflow of blood in the retina. The retina then becomes deprived of oxygen and nutrients, which can result in permanent damage[20]. RVO is the second most common retinal vascular disease after diabetic retinopathy, affecting around 0.5% of adults over 30 years of age [20] and often associated with variable degrees of loss of vision. There are two main types of the condition: branch retinal vein occlusion (BRVO) and central retinal vein occlusion (CRVO). It has been estimated in population-based studies that the prevalence of CRVO is between 0.1 to 0.4%. The most common complaint in patients with CRVO is a sudden painless worsening of vision. It may present as mild, transient obscurations, and persistent blurring [21].

Meanwhile presenting symptoms in BRVO depend on the severity and site of occlusion. Sudden painless loss of vision is the commonest presentation [3].

2.6.2.1 Risk factors for retinal vein occlusion

Retinal vein occlusion occurs due to occlusion within retinal veins, but it remains unclear whether this is a primary or secondary effect. Established cardiovascular risk factors are the predominant medical associations for both central and branch vein occlusions, as summarised below [22, 23]:

- *Hypertension*: This is the predominant risk factor, with up to 64% of patients having hypertension in the older age group (more than 50 years). Inadequately controlled hypertension is associated with the recurrence of RVO [24].
- *High cholesterol*: This is the main association in the younger age group (less than 65 years) of patients with retinal vein occlusion [25].
- *Diabetes mellitus*: This association with retinal vein occlusion may be due to an increase in other cardiovascular risk factors (for example, where 70% of type II diabetics are hypertensive) [23].
- *Glaucoma:* Prospective surveys showed that glaucoma is an important risk factor for the development of RVO, and especially CRVO [26, 27].

Additionally, heamatological, vascular diseases and advanced age are considered as the other main risk factors for RVO [28]. A recent meta-analysis retrieved epidemiological studies of RVO prevalence and risk factors. It identified hypertension and age as a most robust risk factors for the disease. The study reported that in 2015 the RVO prevalence in people aged 30 to 89 years was 0.77%. Stroke history, heart attack history, high cholesterol and higher level of creatinine were also significantly correlated with RVO[29].

2.6.2.2 Types of Retinal Vein Occlusions

❖ **Central retinal vein occlusion (CRVO)**

CRVO is due to an obstruction of the main retinal vein. It is commonly seen in elderly people, with incidence rising sharply above 65 years of age. However, prevalence does not

vary with gender. This severe event that can cause persistent visual loss through macular oedema. Vision loss is also derived from damage to the inner retinal cells carrying visual information from the macula, ultimately causing irreparable damage to the macula's photoreceptors [17]. A subcategory of CRVO can be labelled as ischaemic, which is associated with an inferior visual prognosis and carries a risk of iris neovascularisation (rubeosis), which results in high ocular pressures if untreated (rubeotic glaucoma), which leaves the eye painful as well as blind [30].

- **Branch retinal vein occlusion (BRVO)**

This second type of RVO is caused by the occlusion of one of the branches of the central retinal vein which drains a subsection of the retina[31]. BRVO is more commonly found in the temporal and superior quadrants of the retina than in the nasal and inferior quadrants. In addition to sight loss, patients with BRVO can present with floaters if a vitreous haemorrhage occurs. Although less common than in CRVO, retinal or iris neovascularization may also develop [32].

- **Hemiretinal Vein Occlusion (HRVO)**

HRVO was first described by Hayreh and Hayreh [33] as it is a clinical variant of CRVO which occurs less frequently RVO. The authors stated that a two-trunked CRVO in the anterior part of the optic nerve persists in several human beings as an anatomic variant. As in CRVO, one of the two trunks may become occluded to produce an HRVO [34]. It remains unclear in the era of intravitreal injections whether HRVO ought to be considered as a CRVO, BRVO, or as a different entity but with modern treatment protocols this is largely an academic distinction. According to the clinical findings of HRVO, the occlusion occurs at the optic disc, involving either the inferior or superior hemifield, but is otherwise observed as a mild form of CRVO with the similar history and the same underlying aetiology[35].

2.7 Clinical Presentation of RVO

According to [36], RVO is a relatively common, sight threatening retinal vascular eye disorder. A recent pooled study of data from many countries estimated the global of RVO to be 16.4 million adults. Of these patients, 13.9 million are thought to have BRVO, while 2.5 million are thought to have CRVO[20, 29]. Ophthalmoscopic examinations reveal differing degrees of tortuous and dilated retinal veins, retinal oedema, intraretinal

haemorrhages, exudates, and cotton wool spots. In the absence of a prior diagnosis, chronic RVO may be difficult to identify on clinical evaluation , though this study [37] has suggested several clinical signs which may be present including vascular sheathing and venous collateral formation. Fluorescein angiography (FA) (described later in 2.8) can be used to detect delayed retinal ischemia and the presence of retinal neovascularization with fluorescein leakage in chronic and acute RVO [38]. It is also suggested that RVO patient should be assessed for diabetes during examination as the clinical manifestations of diabetic retinopathy can be quite similar to RVO[[39]. Among subtypes of RVO, CRVO usually has the greater severity of symptoms, it causes greater degrees of vision loss acutely and in general carries a worse visual prognosis with a greater risk of neovasularisation [39], BRVO patients tend to present with mild symptoms compared with CRVO and it is likely that many cases never present to a clinician to be diagnosed [40].

2.8 Diagnostic Imaging Techniques for RVO

There are several diagnostic imaging techniques that ophthalmologists can use in the diagnosis of RVO. A brief overview of imaging techniques will be provided in this section.

2.8.1 Fluorescein angiography (FA)

FA has been valuable in the evaluation of human retinal circulation since its development in the 1960s [41] and has become the standard and classic method for the assessment of chorioretinal vascular disease. Ophthalmologists put drops into the eyes in order to dilate the pupil. The yellow-coloured fluorescein dye is injected into the arm of the patient, and after a few seconds the dye can be observed in the retina. Following the injection of the dye, A fundus camera captures serial photographs. Although, FA can provide a way of visualising areas of vascular leakage and occlusion, it is an invasive method and hard to replicate imaging studies precisely [42]. In occasional cases, FA may also cause severe side effects [43].

2.8.2 Optical coherence tomography (OCT)

Optical coherence tomography (OCT) is a non-invasive imaging modality which can discern the structure of the retina in vivo. It was developed in 1991 and has become crucial in ophthalmology for the diagnosis and assessment of many vision-threatening conditions. This imaging technology uses light waves to take cross-sectional images of the retina[44].

Moreover, spectral domain OCT (Heidelberg Engineering, Inc., Heidelberg, Germany), which became available for clinical use in 2005, can provide detailed anatomical data allowing the ophthalmologist to see each of the retina's distinctive layers. This allows the measurement and mapping of their thickness [45]. These measurements have a central role in diagnosis and OCT images are widely used by clinicians in the decision-making process concerning Anti-Vascular Endothelial Growth Factor (Anti-VEGF) treatment [46]. OCT also provides the fast identification of macular oedema, epiretinal membranes, macular holes, and morphological changes in the layers of the retina[47]. Swept-source OCT was introduced to clinical practice in 2012 to improve on spectral domain OCT by allowing a greater depth of tissue penetration and faster data acquisition [48]. Further extensions to OCT technique include Doppler Fourier-domain. This technology uses Doppler shift for retinal blood flow measurement [49]. Another recent improvement in OCT is OCT angiography (OCTA), this a technique that provides a high-resolution map of the retinal vasculature and detects ocular blood flow [50].

2.8.3 Laser Speckle Flowgraphy (LSFG)

LSFG is non-invasive technology allows for the quantitative characterizing blood flow in the retina and optic nerve head using the laser speckle phenomenon. The laser speckle phenomenon is an interference event noticed when coherent light sources such as lasers are scattered by a diffusing surface. Comparing with other laser techniques, the measurement of the LSFG can cover a much larger field and provides two-dimensional observation of the overall hemodynamic condition of the tissue. LSFG-NAVI (Softcare Co., Ltd., Fukuoka, Japan) is an update version of LSFG, it was approved by the Japanese Pharmaceuticals and Medical Devices Agency in 2008. They demonstrated a superior spatial resolution of the blood flow map of ocular fundus. Using this improved technique can expand the observation area to 24 times larger than the original model[51]. However, it has little clinical use as the major focus of RVO management is anti-VEGF injection which is prescribed on the basis of the central part of the retina, the macula.

2.8.4 Other Technologies

Numerous other technologies have been utilized for the analysis of retinal blood flow and vasculature [52]. Several non-invasive techniques include laser doppler velocimetry (LDV) and blue-field entoptic simulation. The LDV technique is used for red blood cell

measurement and is based on the optical Doppler effect [53], while the blue-field entoptic simulation technique measures the number and velocity of leukocytes in the human macula based on the entoptic phenomenon of tracking the movement of leukocytes when looking into a blue background [54]. Laser Doppler flowmetry (LDF) is a relatively new technique which is similar in principle to LSFG. It is a non-invasive measurement of the tissue blood flow. The main advantage of LDF is its ability to measure the microcirculatory flux of the retina, their non-invasiveness, and fast changes of perfusion during light exposure. However, the major disadvantages are motion artifact noise, the influence of the optical properties of tissues, a lack of quantitative units for perfusion, and lack of knowledge of the depth of measurement and the biological zero signal[55].

For this study, OCT is well suited because the anatomical location of imaging is entirely reproducible as the software allows images to be taken in the same position using retinal vasculature as landmarks. It is quick, does not require pupil dilation, and is non-invasive. OCT has been recommended in the monitoring, diagnosis and assessing treatment response of MO secondary to RVO[56]. In addition, OCT has also been established as standard of care for monitoring treatment response to anti-VEGF agents, in fixed interval, treat-and-extend and pro re nata (PRN) treatment regimes[57-60].

2.9 Treatments for Retinal Vein Occlusion

Once patient present with RVO, they receive anti-vascular endothelial growth factor (anti-VEGF) injections and laser treatment for prevent reducing in VA and increase in macular oedema. Further treatment options include gas inhalation therapies, close observation, and surgery [61].

2.9.1 Injection

Anti-VEGF and corticosteroids are the two primary intravitreal injections for treating RVO patients. Retinal tissue hypoxia resulting from RVO leads to the upregulation of VEGF, which promotes vascular leakage. This Fluid from leaking vessels leads to macular oedema [62]. Anti-VEGF injections are used to prevent this leakage through the retinal blood barrier. Many studies have addressed the effectiveness of three anti-VEGF agents: bevacizumab, ranibizumab, and aflibercept. Bevacizumab injection can temporarily improve central

macular oedema and VA in RVO patients over a 3-to-9-week period [63, 64]. Bevacizumab has also been shown to reduce central macular oedema in patients with ischemic CRVO, but without improvement in visual outcomes [65]. There continues to be significant controversy about which of these anti-VEGF agents should be used in clinical practice as the market cost varies approximately 20-fold. A landmark trial evaluated the three anti-VEGF agents' treatment for macular oedema due to CRVO at 100 weeks. They suggested that mean changes in visual outcomes after treatment with ranibizumab were non-inferior compared to treatments with aflibercept. Mean changes in vision using bevacizumab compared to ranibizumab appeared equivocal but did not meet the statistically significant threshold for non-inferiority. Pragmatically this means there is little evidence to support a strong preference for any of these three anti-VEGF agents[66]. Corticosteroids work differently from anti-VEGF agents by broadly targeting the inflammatory pathway. This anti-inflammatory role reduces vascular permeability, suppressing homing and migration of inflammatory cells, inhibiting leukocyte movement, and stabilizing tight endothelial junctions[67]. Triamcinolone acetonide is an injectable corticosteroid for improving macular oedema over a 10–12-month duration in patients with CRVO [68]. However, triamcinolone is a less effective than bevacizumab for RVO patients[69]. Between anti-VEGF and corticosteroid, anti-VEGF is the standard of care for RVO treatment due to is more efficiency and superior risk profile as teroid increases the rate of cataract development and risks of ocular hypertension and glaucoma[70]

2.9.2 Laser

Laser photocoagulation is the standard therapy for other retinal vascular diseases, most notably proliferative diabetic retinopathy. Panretinal photocoagulation (PRP) is also established for treating RVO patients with neovascularisation or marked retinal ischaemia [71]. PRP cannot improve macular oedema in patients with BRVO and CRVO. In fact, it may worsen it acutely [72]. Its prominent role is in ischemic RVO to prevent or regress neovascularization of the retina and iris [71]. The use of macular grid laser therapy does not appear to improve macular oedema in BRVO patients, but there is no complete consensus, and some experts continue to use it [73]. According to [74]this technique may destroy retinal neurons and cause side effects, for example, impaired night vision and reduced VA, so careful balancing of risks and benefits required.

2.9.3 Gas Inhalation

Blockage of a retinal vein prevents an area of tissue from blood flow and the delivery of oxygen. It was found that lower oxygen saturation in the occluded retinal vein is associated with the level of ischemia in CRVO, representing a potential treatment mechanism [75]. Several studies have examined this effect of hyperoxia to help resolve macular oedema. The reasons behind this are that oxygen causes constriction of vessels, which may decrease leakage and that oxygen can reduce tissue ischemia. Further study in a mouse model of ischemic retinopathy confirmed oxygen can downregulate the VEGF pathway, which can help to reduce macular oedema [74]. Carbogen is an alternative to pure 100% oxygen, a mixture of 95% oxygen and 5% carbon. It may counteract oxygen-induced vasoconstriction and allow for maximal oxygen delivery to the retina. Coupling carbogen inhalation with acetazolamide injection in a porcine model of ischemic retinopathy help to decrease ischemia due to BRVO [76]. This treatment remains experimental and seems unlikely to supersede current standards of care soon.

2.9.4 Other Therapies

As RVOs can be caused by thrombosis, several studies have tested the efficiency of systemic anticoagulant and antiplatelet therapies with drugs including ticlopidine, streptokinase, and heparin [77-79]. Surgical intervention and the intravitreal injection of tissue plasminogen activator (tPA) have also been investigated on animal models and humans [80, 81]. However, although these therapies can in theory deliver a way to address a substantial cause of RVO, the risks of haematological disorders with ticlopidine and vitreous haemorrhage with streptokinase seem to exceed the advantages of these medicines [82]. Also, studies have investigated the benefits of combining injection with laser therapies [83]. In addition, recent improvements in mass-spectrometry have allowed scientists to uncover changes in protein dynamics associated with RVO. This, in turn, has led to the development of new treatments. [84].

According to [46] the recent standard guidelines in the UK for RVO treatment, Intravitreal injections of licensed anti-VEGF are recommended for treatment of MO secondary to RVO. In addition, OCT is recommended in diagnosing, monitoring, and assessing treatment response of MO secondary to RVO. The guidelines stated that just over a third of patients

require three anti-VEGF injections to reach the highest VA, whereases other patients may require six consecutive anti-VEGF injections. However, some patients show no improvement in VA or OCT central subfield thickness after three loading injections at monthly intervals. Therefore, treatment is not recommended if no response occurs after six injections.

In many cases, MO can be successfully managed with intravitreal injections of VEGF agents. However, treatment response is variable. It is costly and carries a risk of blinding. Additionally, humans have no way of reliably predicting good responses. It would be great if there were a standardised approach in clinical practice. This approach can help ophthalmologists and patients making a decision regarding RVO treatment.

Chapter 3 LITERATURE REVIEW

While large number of studies have demonstrated great promise in the classification and detection of many eye diseases including RVO disease which anti-VEGF is a standard treatment for such disease, less research has been conducted towards patient response to anti-VEGF therapy [85-89]. One of these models is to predict the anti-VEGF demand. These models predict whether a patient will show good or poor response, with low or high treatment demand respectively[90]. To my knowledge only one study has used machine learning classification technique to predict the anti-VEGF demand for RVO disease [90]. This study used OCT volumes at baseline and after two consecutive visits. The literature review below is divided into five sections regarding the techniques that related to the proposed work including medical studies, machine learning, deep learning, convolutional neural networks, prediction techniques for Anti-VEGF treatment demand, existing OCT databases.

3.1 Medical Studies

An active learning approach has been proposed to evaluate the use of the quantitative contrast sensitivity function (qCSF) in patients with RVO [4] in an observational, prospective study. The patients involved in the study had a history of RVO in one or both eyes. During regularly scheduled visits, they were tested using the Manifold Platform (Adaptive Sensory Technology, San Diego, CA) and SD-OCT. The study estimated a CSF model using an information-gain strategy to provide a global functional vision metric through the area under the CSF (AULCSF). In addition, contrast sensitivity was compared with previous data collected for 62 eyes from age-matched healthy controls. Twenty-one eyes with RVO were tested. The presence of MO markedly reduced contrast sensitivity relative to eyes with RVO but no macular oedema. The effect of anti-VEGF injection was also measured for a small set of eyes, and it was found that AULCSF improved. At the same time, logMAR VA comprised of rows of letters used by ophthalmologists to estimate VA did not show equivalent significant improvements. This study confirmed the occurrence of reduced contrast thresholds in patients with RVO.

A retrospective cohort study was conducted using sequential data from patients obtained from EMR at the Newcastle Eye Centre at the Royal Victoria Infirmary, Newcastle upon Tyne. The study aimed to determine the real-world visual outcomes of intravitreal injection (IVI) of anti-VEGF for CRVO complicated by macular oedema. Two-hundred-and-thirty-one eyes with CRVO complicated by MO were treated, with 6-48 months of follow-up. The

mean gain in VA after two years was 8.9 (SD 19.0) Early Treatment of Diabetic Retinopathy Study (ETDRS) letters. The authors observed that there were improvements in vision following anti-VEGF treatment; but more modest than those reported in clinical trials[91].

A further study evaluated the status of the ellipsoid zone (EZ) as measured by optical coherence tomography (SD-OCT) and its association with VA [92]. The research looked at eyes with MO secondary to CRVO or HRVO in a large clinical trial of RVO called SCORE2. SD-OCT macular cube scans of a randomly selected subset of 75 eyes were stratified by baseline. VA data and the EZ layer was segmented using customised software. At month 1, 42 eyes were found to have gradable images and the scans from these eyes were graded at months 1, 6, and 12. EZ layer thickness was measured within the central subfield (CSF) from the top of the EZ layer to the top of the retinal pigment epithelium (RPE). It was found that, at month 1, thinner EZ and larger areas of EZ defect were associated with worse VA in eyes with secondary macular oedema. Also, no significant relationship was found between changes in EZ thickness or the area of EZ defect and changes in VA from months 1 to 6. However, there was a considerable association between changes in EZ thickness and those in VA between months 1 to 12.

Work on retinal reflectivity (RRF) in healthy eyes and eyes with retinal artery occlusions (RAOs) and ischemic or non-ischemic RVO has also been carried out [93]. This study evaluated 100 eyes with branch or central RAO and 34 eyes with branch or central RVO. Fluorescein-angiography (FAG) was used to distinguish between ischemic and non-ischemic RVO. SD-OCT scans were treated as black-on-white JPEG images in Adobe Photoshop. Areas of vitreous body (VB), ganglion cell layer (GC), outer nuclear layer (ONL) and inner plexiform layer (IPL) 1000 µm away from the fovea were marked and mean grayscale was calculated to quantify RRF in which 0: black (high reflectivity), and 255: white (low reflectivity). SD-OCT measured RRF changes in inner retinal layers according to the type and severity of occlusions. According to the results, OCT may detect different levels of ischemic changes in the retina, while a weak association was found between RRF changes and ischemic areas.

Another review analysed data collected from a number of randomized clinical trials (RCTs) where anti-VEGFs were considered the primary treatment for CRVO patients, including CRUISE [94], RETAIN [95] and HORIZON [96]. Additionally, the authors looked at GENEVA and SCORE which used intravitreal steroids as a first-line therapy rather than

anti-VEGF. Alternative treatment regimens such as switching between anti-VEGF agents and/or steroids and laser therapy were also explored. Finally, a simplified modified treatment algorithm for CRVO patients was proposed. It showed that not all visual gains are maintained beyond the first year. Moreover, patients exhibit varying behaviour patterns. Some showed a full response with few recurrences, whereas other patients showed partial or even no response with several recurrences of anti-VEGFs. Based on the data, there are two main groups of patient response: early responders, and partial- or non-responders. It was concluded that robust clinical trials show that anti-VEGF alone is the standard treatment and forms the basis of modern treatment protocols.

Work has also been undertaken to identify a relationship between initial and final VA with changes in choroidal thickness [97]. A retrospective clinical analysis of macular ischemia was performed using fundus examination, clinical records, OCT, and fluorescein angiography in 35 eyes of patients with RVO from January 2011 to December 2013 who were followed over 6 months. The study evaluated the range and location of macular oedema, macular ischemia, choroidal thickness, initial and final VA, and treatment. In addition, changes in sub-foveal choroidal thickness were analysed. According to the results, choroidal thickness increased in RVO eyes; however, 6 months after treatment, average choroidal thickness had decreased but was still thicker than in fellow eyes. Also, a significant correlation was found between quantitative differences in the choroidal thickness of RVO eyes and fellow eyes with final visual outcome.

A recent study [98] aimed to examine the impact of ranibizumab and aflibercept IVIs in a real world of patients with BRVO complicated by macular oedema, and to evaluate the benefit of switching from ranibizumab to aflibercept in refractory cases. A cohort of 259 treatment naive eyes from 258 patients receiving ranibizumab, aflibercept or a combination of both from 2013 to 2018 with 6 months or more follow-ups were examined. The main outcome measure of this study was change in VA from baseline at 12 months following treatment initiation. Secondary outcomes involved VA at a range of time points, and the presence or absence of MO and type and number of IVIs given. The conclusion drawn in this study was that this real-world cohort study of BRVO complicated by MO presented more modest improvements in VA from anti-VEGF IVIs than those reported in clinical trials. Also, it was shown that, although aflibercept produced superior anatomical outcomes, ranibizumab was similar to aflibercept in terms of visual outcomes.

A retrospective study [99] aimed to determine whether or not foveal bulge was significantly correlated with VA after resolution of the MO associated with BRVO. A foveal bulge is a term utilised for the bulge in the inner segment-outer segment (IS-OS) line at the centre of the fovea seen on SD-OCT images. (Figure 3.1)

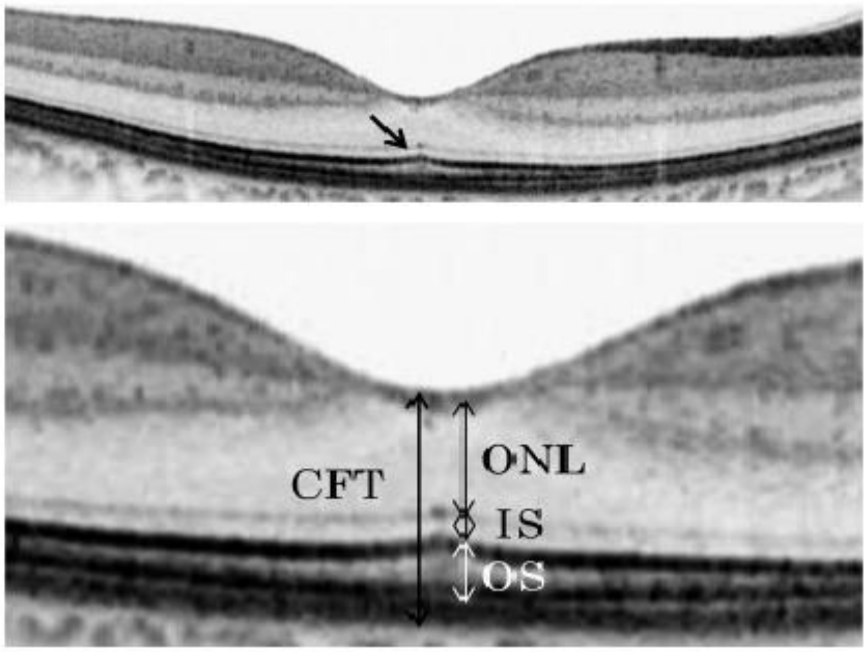


Figure 3.1: Foveal bulge in Optical Coherence Tomography[99]

Also, the macular status was investigated at the first visit and the factors influencing VA after resolution of the MO were assessed. The researchers examined the medical records of patients with resolved MO and IS/OS line at the central fovea in the SD-OCT images. A set of 31 eyes with MO associated with BRVO were studied, and 31 unaffected fellow eyes were controls. Of the 31 eyes with BRVO, 27 received an intravitreal injection of bevacizumab and a posterior sub-Tenon injection of triamcinolone acetonide for treatment. The horizontal cross-sectional images that were recorded at the initial and final visits after resolution of the MO were evaluated. The central foveal thickness (CFT), which is the distance between the internal limiting membrane (ILM) and the outer border of the retinal pigment epithelium (RPE) at the central fovea, was measured. Furthermore, the CFT was measured automatically with the caliper measurement tool embedded in the SD-OCT system. The findings indicated that the presence of the foveal bulge in the SD-OCT image is a good indicator of better BCVA after the resolution of MO associated with BRVO. The percentage of eyes with foveal bulge after resolution of the MO was only 22.6% in the BRVO group but was 100% in the control group. In addition, the BCVA in the BRVO group was better in the eyes with foveal bulges than those without. Thus,

according to this study, the foveal bulge can be considered as a good marker of the functional properties of the fovea in eyes with resolved MO associated with BRVO. However, foveal bulge is only assessed after the MO has resolved. Therefore, this marker may have a little prognostic value at the initial presentation.

A prospective study [100]examined 25 patients with unilateral acute major BRVO in the Department of Ophthalmology at Kyoto University Hospital during 2011 and 2012. This study aimed to illustrate the pathogenesis of BRVO by examining the clinical relevance of the morphological changes of retinal vasculature in BRVO and studying the morphologic changes of BRVO-affected retinal veins by sequential thin sectioning with OCT. All of the eyes had BRVO symptoms within six months and the BRVO was diagnosed based on fundus examination and the fluorescein angiography findings of 2 retina specialists. Arteries, veins, and arteriovenous (A/V) crossing was examined using sequential thin sectioning by Heidelberg Retinal Angiograph (HRA) and OCT. Additionally, retinal blood flow was simulated in vitro and scanned with HRA and OCT. In OCT images of healthy eyes, most retinal vessels are noticed as oval configurations with heterogeneous reflectivity and usually with four distinctive hyperreflectivities in vitro. To simulate the blood flow in the retinal vessels in vitro, blood samples were pumped through a fine glass tube at various flow velocities. The OCT of the glass tube without blood flow showed homogeneous internal reflectivity. Their findings suggest that BRVO can occur through different mechanisms depending on the relative anatomic positions of crossing vessels. Additionally, the authors detected the clinical relevance of the presence of thrombus in eyes with acute BRVO that was significantly correlated with the retinal perfusion status (ischemic BRVO).

3.2 Machine Learning

Another paper [85] investigated 5 machine learning algorithms (gradient boosting, AdaBoost.R2, extremely randomized trees, Random Forest, and LASSO) in order to predict the outcome of VA (VA) after 1 year of anti-VEGF treatment in patients with diagnosed neovascular age-related macular degeneration (AMD). The clinical dataset came from EMR and OCT measurement data, but not OCT images. The dataset included measurement values such as VA and an aggregated version with different timeframes of 3 months and 12 months and different aggregation functions using mean, variance, minimum, and maximum. To evaluate the performance of the system 10-fold cross validation was used and the data were

randomly split into 10 equally sized datasets. This system aimed to predict VA in patients with AMD undergoing anti-VEGF injection after three months. By comparing the ML techniques, the best performing algorithm was LASSO.

Another research group has evaluated vitreomacular adhesion (VMA) in patients receiving ranibizumab for MO secondary to branch and central RVO by using the fully automated computational segmentation and classification of SD-OCT images [86]. In this study, they investigated the effect of VMA on anti-VEGF therapy for RVO. Machine learning and a graph cut segmentation algorithm were designed with data from 391 CRVO and BRVO patients were included in the study. The results suggested that machine learning represents a promising path in assessing imaging biomarkers in OCT.

3.3 Deep Learning

Other researchers have investigated the performance of two machine learning methods in detecting CRVO using ultra-wide-field fundus images from 202 normal retina and 125 CRVO patients [87]. Training was provided to the CNN algorithms using ultra-wide-field fundus images. The diagnostic abilities of the support vector machine (SVM) algorithm and deep learning (DL) were determined by assessing their specificity and sensitivity. The results showed that the deep learning model outperformed the SVM in the diagnosis of CRVO. DL exhibited a sensitivity of 98.4%, in contrast to 84.0% for SVM, while DL had a specificity of 97.9% whereas for the SVM it was 87.5%. It is worth mentioning that specificity and sensitivity techniques are used to measure the diagnosis accuracy. Theoretically, the perfect diagnostic technique can discriminate cases with and without disease. Subjects with a perfect test above the cut-off diagnose disease, while the values below indicate no disease. To explain the calculation of these measurements, Table 3.1 below shows the population of examined subjects with and without disease divided into four subgroups: true positive (TP) represents the subjects of disease with the value of a parameter of interest above the cut-off. False positive (FP) represents the subjects without disease above the cut-off. True negative (TN) represents subjects without disease below the cut-off. False negative (FN) represents subjects with the disease indicated below the cut-off. Sensitivity defines the percentage of true positive subjects with the disease over a whole group of subjects with disease (TP/TP+FN). At the same time, Specificity describes a proportion of subjects without the disease with a negative test in a total of subjects without disease (TN/TN+FP)[101].

	Subjects with the disease	Subjects without the disease
Positive	TP	FP
Negative	FN	TN

Table 3.1: illustrates the population of subjects with and without disease.

Further research by a previous group used a similar methodology with DL and the SVM to detect BRVO[88]. The study included 237 images from 236 patients with BRVO and 229 images from 176 non-BRVO healthy subjects. In order to construct the DL model, the CNN was trained using ultra-wide field fundus images. The positive predictive value (PPV), sensitivity, specificity, negative predictive value (NPV), and area under the curve (AUC) were calculated in comparing the diagnostic abilities of the DL and SVM models. Where the PPV defines a proportion of patients with positive test result in total of states with positive results (TP/TP+FP). Whereas NPP =(TN/TN+FN). AUC used for estimation of how high the discriminative power of a test is. When AUC =1 indicates a perfect diagnostic test, while a non-discrimination test has an area 0.5. Figure 3.2 explains how AUC works; a higher x-axis value indicates a higher number of false positive than true negative. Nevertheless, higher y-axis value represents a higher number of true positive than false negative. Where TPR is true positive rate, and FPR is false positive rate[101].

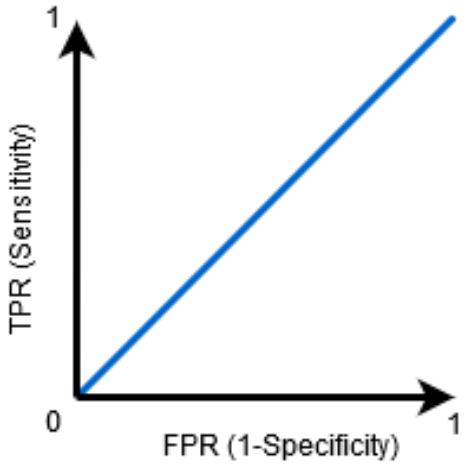


Figure 3.2: Area under the curve (AUC).

From the results, deep learning outperformed the SVM model for all of these parameters. They also tested the effects in the two patient groups of sex, age, right or left eyes, hypertension, and diabetes. No significant differences in terms of these factors were detected. However, there were significantly higher rates of diabetes, hypertension or arterial

occlusive disorders in the BRVO images than in the non-BRVO images [88]. This led to the conclusion that diabetes, hypertension, and arterial occlusive disorders are risk factors for BRVO.

A retrospective study presented a novel deep learning model that analyses OCT scans and makes referral suggestions that is comparable to clinical experts. A deep learning architecture has been applied to a clinically heterogeneous set of OCT scans obtained from patients referred to Moorfields Eye Hospital, London [89]. A total of 14,884 OCT scans were used for training, and a deep segmentation network was implemented to create device-independent tissue-segmentation mapping. A three-dimensional U-Net architecture was used in the segmentation network to translate the raw OCT scan into a tissue map with 15 different tissue classes. Then, a deep classification network was applied to analyse the segmentation map and to provide diagnoses and referral suggestions. The decisions generated used a framework with four clinical referral suggestions: urgent, semi-urgent, routine, and observation only. The model was compared with the two highest performing retina specialists; it performed comparably to the two best retina specialists without making serious wrong decisions. Model performance was reported with AUCs that were 99% for most of the diseases.

In another study, a DL algorithm was applied for the automated detection of retinal vein occlusion (RVO) features in fundus photographs [5]. The algorithm was trained using a retrospective dataset of fundus images graded independently from 1 to 49 for the presence of RVO by a group of 122 ophthalmologists. One hundred images were used for validation which were acquired from 86 eyes of 90 patients with a mean age of 58 years. The performance results achieved were sensitivity at 81.36% and specificity 82.76% at the selected operating point. Deep learning has also been utilised to discriminate between normal OCT images and images of patients with age-related macular degeneration (AMD) [102]. OCT images were extracted using an extraction tool from the Heidelberg Spectralis imaging database (Heidelberg Engineering, Heidelberg, Germany). These images were merged later with data from patient medical records. Cross-validation was achieved using a random subset of patients. Receiver operating characteristic (ROC) curves were constructed at macular OCT level, independent image level, and patient level. The deep learning network investigated achieved AUC of 93.83% with an accuracy of 88.98%. At patient level, the study achieved AUC of 97.45% with an accuracy of 93.45%, and at image level, accuracy

was 87.63% AUC of 92.78%. According to this study, deep learning techniques represent promising and reliable methods as image classification techniques.

Other variations of such techniques have been added to deep leaning methodologies and CNN architectures. Most notably, the hierarchical local binary pattern LBP (HLBP) was introduced in a study which claimed to be the first to achieve the automatic recognition of BRVO using fundus images[103]. The architecture of HLBP is based on the LBP feature representation method which acts as the coding operator for BRVO recognition. The dataset comprised a total of 200 fundus images obtained from 200 persons, 100 normal fundus images and 100 BRVO fundus images. The accuracy of the method was found to be 97%.

3.4 Convolutional Neural Networks (CNNs)

A pilot study evaluated two retinal blood vessel (RBV) segmentation algorithms, both of which are based on directional response vector similarity and region growing [104]. Fundus images were evaluated for both healthy retinas and those affected by RVO disease. The study presented two methods. The first used a RBV segmentation algorithm which generates a vascular binary map of the fundus image utilised as input. It carries out region growing and makes the use of directional response vector similarity instead of relying on the maximal or other aggregated measures such as averaged or summed, and stores and handles all the directional vectors of each pixel. In the second method, calculations of response vectors and the template machine were carried out with a generalised form of Gabor filters. Instead of using matched Gabor filtering with only a few specific Gabor filters, this method depends on a large set of such templates in different sections of the retinal vascular system. The fractal properties of blood vessels were computed and compared to distinguish normal retinae and those affected by RVO. The methods were trained on images from the DRIVE and STARE databases. Experimental results showed that one of these methods was computationally effective and suitable for real-time computer-aided diagnosis applications, and the other method was robust against unexpected inputs. This advantage renders it potentially useful in retinopathy screening systems[104].

Convolutional neural networks (CNN) have been used for BRVO recognition, and image-based as well as image patch-based voting methods have been proposed for recognition [105]. Augmentation of the number of images was achieved by the addition of noise flipping and image rotation. Green channel images were used for processing because haemoglobin absorbs light more around the green spectrum [106]. Global and local enhancement were

tested, with the former performing histogram equalisation on whole images, while the latter processed the image first and divided it into patches. Histogram equalization was implemented on each patch, and then the patches were joined with some overlap to form the whole image again. The same dataset was then used in a further study [103] of 100 BRVO and 100 normal fundus images. In the image-based method, image pre-processing was implemented to resize the images to 60*60 for CNN recognition. In addition, the CNN input size was 50*50 for the patch-based method. Both methods used a CNN with 3 convolutional layers. A comparison with the hierarchical local binary pattern (HLBP) recognition method was conducted to demonstrate the superior performance of the proposed method, which yielded an accuracy rate of 97% according to experimental results.

A CNN method for the recognition of CRVO has also been proposed [11]. The authors collected 108 CRVO images from the STARE, DRIVE and Dr Hossein Rabbani databases and the Retinal Image Bank. The CNN used was based on a pre-trained network, LeNet [107]. The images were pre-processed to enhance their quality since they were collected from different databases containing different sized images in different formats. They were then converted into TIFF format and resized to a standard size of 60×60. In addition, contrast-limited adaptive histogram equalization (CLAHE) was applied in order to enhance the contrast of grayscale retina images. The proposed CNN consists of 12 layers, including three convolution layers, three pooling layers, 4 rectified linear units (ReLUs) and two fully connected layers. From two classes of normal and CRVO images, an accuracy rate of 97.56% was obtained.

Other advances in the field include a pilot study to assess a proposed multi-categorical classification of retinal diseases [108]. The STARE retinal image database was used to train a CNN using MatConvNet for the automated detection of multiple retinal diseases from fundus images. The database consists of 397 images divided into 10 categories, including normal retinas and nine retinal diseases including BRVO and CRVO. Four types of deep learning models were compared: transfer learning with an SVM-based Gaussian kernel, transfer learning with Random Forest, VGG-19, and AlexNet. It was found that the performance of the deep learning models decreased as the number of categories increased, and the size of datasets was small regarding classification performance. This led to the conclusion that the deep learning techniques used were ineffective for clinical application due to the considerable numbers of patients suffering from different types of retinal diseases visiting for diagnosis and treatment.

Another study proposed a CNN architecture to work specifically with retina images and to detect diabetic retinopathy (DR) and RVO [109]. It was found that the CNN can detect DR at the earliest stages and graded as mild, moderate, non-proliferative DR (NPDR), severe NPDR, or proliferative DR (PDR). The model was shown to detect DR in the early stages or mild-NPDR, with accuracy rates of 98.11% for the Messidor and 96.6% for the Kaggle databases. Additionally, both types of RVO (CRVO and BRVO) were detected. The images were collected from STARE and the Retinal Image Bank. The algorithm can classify normal, CRVO and BRVO images with 97% accuracy. A cascaded convolutional neural network (CCNN) was used in this study, which is a chain of three CNNs with the same configuration.

3.5 Machine Learning Techniques for Prediction Anti-VEGF Treatment Demand

Various algorithms have been suggested to predict low and high anti-VEGF treatment demand. A retrospective cohort study was used to assess the potential of machine learning for the prediction high and low treatment demand in real life in patients with three eye diseases: diabetic MO, neovascular age-related macular degeneration (nAMD), and retinal vein occlusion (RVO) [90]. Patients were treated according to a treat-and-extend regimen (TER). According to a predefined TER from 2014 to 2018 at the University Hospital of Bern, Switzerland, 340 patients with nAMD, 155 eyes with RVO related MO, and 178 eyes with DMO were included. The study considered ME-related RVO and DME in a single pathological group named the retinal vascular diseases group. The study aimed to observe the practical feasibility of such a prediction tool in practice to detect the long-term demand for anti-VEGF medication at the early stage of a 1-year TER in a routine clinical setting. Two binary classifiers were used: high versus others (HvO), which predicts the probability of an eye showing high treatment demand; and low versus others (LvO) to predict the probability of an eye showing low treatment demand. A Random Forest classifier was used for both LvO and HvO. Patient data, including age and sex, and the OCT volume data of the registered eyes gained at each visit were included. A morphological technique was used to extract features from OCT volumes. Existing retina and layer segmentation algorithms and biomarker presence detection algorithms were utilised to construct meaningful and compact morphological input features. The HvO and LvO classifiers were computed at three-time points: firstly, using baseline information only; secondly, from baseline information and subsequent consecutive visits; and thirdly from the baseline information and two consecutive

visits. Several limitations were encountered in this study, including the consideration of MO-related RVO and DMO as a single group since their specific behaviour and demand for the VEGF injections may differ. In addition, patients involved in the study were elderly, which may prevent the generalisation of the algorithm to younger patients. The results showed that machine learning could predict 12-month VEGF treatment demand. The nAMD-trained models produced over the 10-fold crossovers gave AUCs of 0.79 and 0.79 for high and low demand respectively. Models for RVO and DMO demonstrated matching results, with mean values of AUC of 0.76 and 0.78 for low and high demand respectively.

A deep learning strategy has also been investigated to predict the response of DME patients to anti-VEGF after every three months of injection treatment using 49 B-scan volume OCTS for each eye at baseline visits [110]. 127 subjects treated for DME with anti-VEGF injections were recruited in this retrospective study. The classification of the responder classes was based on the change in the thickness of the retina relative to the third injection. Patients were categorised in two groups: a responsive group with a decrease of 10% of retinal thickness at the fourth visit; and a non-responsive group for the remaining patients. The proposed algorithm reached an average AUC of 0.866 in distinguishing between responsive and non-responsive patients. However, the sample size was relatively small and divided for training and testing. Additionally, the study considered only OCT scans as input while using complementary features, for example, gender, age, and duration of diabetic disease. Furthermore, the responder classification only depends on the information at the baseline and the fourth visit.

Feng et al. [111] utilized a technique based on pre-trained CNN to predict the effectiveness of anti-VEGF treatment for choroidal neovascularization (CNV) and cystoid MO patients using OCT images. The study investigated 228 patients assessed at the Second Affiliated Hospital of Xi'an Jiaotong University, China, from 2017 to 2019 with diagnoses of CNV, MO or symptoms of both. Their method workflow comprised three stages: pre-processing in order to remove noise from the OCT scans, followed by augmentation of the dataset using rotation, flipping, translation, and Gaussian blurring. As a result of this step, the dataset increased from 228 images to 912 images. Finally, four CNNs pre-trained were fine-tuned in the last layers, including AlexNet, VGG-16, GoogLeNet, and ResNet-50. They found that accuracy was similar for the dataset with or without augmentation. Among the four CNNs, ResNet-50 had the highest AUC of 0.81, which means that esNet-50 is more appropriate to use for the prediction of the effectiveness of anti-VEGF.

Another study [112] used a machine learning approach to predict low and high anti-VEGF injection demand for eyes with nAMD, utilising OCT images obtained during a PRN treatment regimen. OCT scans were analysed at baseline, after 1 month, and after 2 months. Demographic characteristics were employed, including age, gender, and baseline visual acuity. A Random Forest classifier was trained to predict treatment categories and the results were evaluated using cross-validation. Three hundred and seventeen eyes were used, and 71 patients presented low, 176 medium, and 70 high injection requirements from months 3 to 23. The study predicted high and low anti-VEGF treatment needs, with AUCs of 80% and 70% respectively.

In another study by Sil Kar et al.[113], texture-based radiomics features were identified within an individual fluid and retinal tissue compartments of baseline spectral-domain optical coherence tomography (SD-OCT) images, and the specific spatial compartments that contributed to the most pertinent features were investigated in order to predict therapeutic response in MO secondary to RVO disease. The study extracted 962 texture-based radiomics features from each of the fluid and retinal tissue compartments of OCT scans, which were acquired from the PERMEATE study. After selecting top-performing features using different feature selection methods, four machine learning classifiers, Random Forest (RF), quadratic discriminant analysis (QDA), linear discriminant analysis (LDA), and the support vector machine (SVM), were investigated with a cross-validation approach in predicting responses to anti-VEGF therapy. The principal findings were that texture-based radiomics features were most closely related to responses to anti-VEGF injections. Also, the IRF sub compartment's texture features distinguished rebounders from non-rebounders. Gathering features of fluid and retinal tissue provoked a cross-validated area under receiver operating characteristic curve (AUC) of 0.78±0.08 in distinguishing rebounders from non-rebounders applied using the LDA classifier.

A pre-trained deep learning algorithm has also been examined [114] for the prediction of anti-VEGF treatment indications based on central retinal OCT scans of eyes with or without retinal disease. A total of 183,402 retinal OCT B-scans obtained between 2008 and 2016 were collected from the image archive of a university hospital. OCT images were pre-processed and split in a 9:1 ratio into training and testing datasets. OCT images following intravitreal injection during the first 21 days after image acquisition were labelled as the 'injection' group, while 'no injection' groups were assigned to the same amount of OCT scans without intravitreal injections. The accuracy of the CNN classifier correctly predicted

16,166 of the 17,112 images in the test dataset as either belonging to the ‘injection’ or ‘no injection’ groups, giving a prediction accuracy of 94.5%.

3.6 Existing OCT Databases

Publicly available databases are potentially valuable for research and health care innovation. The following brief overview describes several popular datasets used for various eye diseases. According to a global review [115], only 94 of the 140 unique datasets were open access so that the raw data can be downloaded, while the rest of the datasets are categorised as open access with barriers which can include governance, cost, and time barriers. In addition, usability issues include data format and quality barriers as well as image labelling barriers. Ophthalmological conditions represented in the datasets were diabetic eye disease (37%), glaucoma (20%), age-related macular degeneration (16%), hypertensive retinopathy (6%), cataracts (4%) and healthy eyes (62%). Furthermore, the main imaging modalities were retinal fundus photographs (57%), OCT (19%) containing 2- and 3-dimensional images, external eye photographs (7%), in vivo confocal microscopy (5%), scanning and adaptive optics-scanning laser ophthalmoscopy (5%), and fluorescein angiography (4%). Table 3.2 illustrates examples of publicly available datasets for OCT imaging.

A comprehensive open-access database containing more than 500 high resolution OCT images has been created which categorises different pathological conditions[116]. These images include age-related macular degeneration, macular hole, normal, diabetic retinopathy, and central serous retinopathy. The images were obtained from a raster scan protocol with a 2mm scan length and 512x1024 pixel resolution at Sankara Nethralaya (SN) Eye Hospital, Chennai, India. Each dataset has images at different stages in each of the diseases, including less severe, medium severe, and more severe phases, and 25 normal images along with the corresponding ground truth delineations that had already been determined by an expert. The study provided a user-friendly graphical user interface (GUI) which can be used by clinicians and researchers for manual or semi-automated segmentation.

Another study [117] launched open-access raw optic disc tilt (ODT) OCT images for the analysis of new images processed using segmentation algorithms. This dataset of ODT images includes both vertical and horizontal cross-sectional images obtained using spectral-domain optical coherence tomography (SD-OCT). All images are aligned with the centre of the optic nerve head. The dataset contains 121 each of clear and manually marked images. This public open access database is available online at the ICPSR website of the University

of Michigan. The dataset also consists of optic disc images from 67 subjects (33 male, 34 female) imaged by OCT covering images of both clinically normal and myopic subjects.

A new rat OCT database has also been presented [118], and the segmentation of the retinal layers was manually reviewed by experts. The proposed database contains two image batches, the first for use in performance evaluation and the second to analyse the generalisation ability of the method when applied to an image set with different conditions to be used for training. The images in the first batch of the database include follow-up 3, 7 and 14 days after injection performed on 10 rats, totalling 129 OCT images of 1024 × 1024 pixels with a resolution of 0.9775 μm/pixel. For the second batch, follow-up 3 and 7 days after injection was applied to 12 rats, obtaining a total of 115 OCT images with the same size and resolution.

Merging three-dimensional (3D)-optical coherence tomography (OCT) and fundus colour images has also been proposed [119]. Firstly, these two modalities are combined to register colour fundus images with an en-face representation of OCT. A curvelet transform is applied to extract vessels for both modalities. In the second step, the extracted vessels from two modalities are registered together in two phases. At first, scaling and translation transformations are used for image registration. Then a quadratic transformation model is applied between two pairs of images, and 22 pairs of images (5 peripapillary and 17 macular) were obtained from random patients with a variety of retinal diseases. Each pair includes one OCT image acquired with the Topcon 3D OCT-1000 instrument and a colour fundus image. OCT images consist of 650 different slices with a size of 650 × 512 × 128 voxels and a voxel resolution of 3.125 μm × 3.125 μm × 7 μm.

Another collection of OCT image data can be obtained from the Dongaonkar eye hospital [120]. These images were captured in clinic using the Cirrus HD-OCT (Carl Zeiss Meditec, Dublin, CA). The images were analysed using optic nerve head analysis (ONH) and retinal nerve fibre layer analysis (RNFL). The database contains 10 patients' OCT images which are two-dimensional with a size of 526×790. Then, after converting OCT colour into grayscale, various noise removal techniques were applied to the images. Furthermore, segmentation and classification were also investigated.

A publicly available retinal OCT-angiography vessel segmentation (ROSE) dataset has also been introduced [121]. This dataset comprises two subsets named ROSE-1 and ROSE-2

which were acquired using two different devices. ROSE-1 set contains of a total of 117 OCTA images from 26 subjects with Alzheimer's disease (AD) and 13 healthy controls, with a mean age of 68.4 ± 7.4 years for the AD group and 63.0 ± 10.2 years for the control group. Participants with eye diseases included age-related macular degeneration, glaucoma, and high myopia. All the OCTA scans were captured by the RTVue XR Avanti SD-OCT system (Optovue, USA) equipped with AngioVue software, with an image resolution of 304×304 pixels. The scan area was 3 × 3 mm2 centred on the fovea within an annular zone of 0.6 mm-2.5 mm diameter around the foveal centre. The ROSE-2 subset contains a total of 112 OCTA images taken from 112 eyes, acquired by a Heidelberg OCT2 system with Spectralis software (Heidelberg Engineering, Heidelberg,Germany) from eyes with various macular diseases. All the images in this dataset are en-face angiograms of the SVC within a 3×3 mm2 area centred at the fovea. These images have been reconstructed from 512 × 512 repeated A-scans, with the Heidelberg automated real time (ART) and Trutrack system employed to reduce noise.

Dataset Name	**Number of Images**	**Country of Origin**	**Image Type**	**Disease included**
BIOMISA Retinal Image Database	2,561	Pakistan	Fundus and OCT photography	Glaucoma, AMD, macular oedema, and healthy eyes
BIOMISA AMD	6,900	Pakistan	Fundus and OCT photography	AMD and healthy eyes, diabetic eye Disease.
BIOMISA Glaucoma Fundus and OCT	132	Pakistan	Fundus and OCT photography	Glaucoma and healthy eyes
Retinal OCT Fluid Challenge	112	Canada	OCT	AMD; retinal vein occlusion
Detection Of diabetic retinopathy in primary care offices	819	USA	Fundus and OCT photography	Diabetic eye disease and healthy eyes
Hong Kong OCT	4,877	China	OCT	Glaucoma and healthy eyes

Moorfields	16,990	UK	OCT	Diabetic eye disease; AMD; Choroidal neovascularisation; multiple retinal pathologies and healthy eyes
Yang OCT	10,100	South Korea	OCT	AMD and Healthy eyes
Contact Lens Anterior Segment–Optical Coherence Tomography Understanding Dataset	112	Spain	OCT	NR
Srinivasan 2014	3231	USA	OCT	Diabetic eye disease, age-related macular degeneration, healthy eyes
Corneal Heidelberg OCT	579	Iran	OCT	Healthy eyes
Retinal Fundus and OCT	44	Iran	OCT and fundus photography	Various retinal diseases
Fang 2013	195	USA	OCT	Healthy eyes
Fang 2012	51	USA	OCT	Age-related macular degeneration and healthy eyes
Duke OCT	38400	USA	OCT	Age-related macular degeneration and healthy eye
Kermany/ Guangzhou	109312	USA China	OCT	Diabetic eye disease, drusen, choroidal neovascularisation, healthy eyes
Noor Hospital	4142	Iran	OCT	Diabetic eye disease, age-related macular degeneration, healthy eyes
Chiu 2015	10	USA	OCT	Diabetic eye disease

Healthy OCT and Fundus	100	NR	OCT and fundus photography	Healthy eyes
OCT Glaucoma Detection	1100	NR	OCT	Glaucoma and healthy eyes
Optical Coherence Tomography Retinal Image Analysis 3D	10	NR	OCT	Healthy eyes
Canada OCT Retinal Images	470	India	OCT	Diabetic eye disease, age-related macular degeneration, macular hole, central serous retinopathy, healthy eyes
Project Maculopathy Unveiled by Laminar Analysis	239	USA	OCT and fundus photography	Age-related macular degeneration and healthy eyes
Retinal Optical Coherence Tomography Classification Challenge	165	Iran	OCT	Diabetic eye disease and healthy eyes
Investigative Ophthalmology and Visual Science; Chiu 2011	220	USA	OCT	Age-related macular degeneration

Table 3.2: Examples of publicly available datasets for OCT imaging [115].

Chapter 4 DATA SET

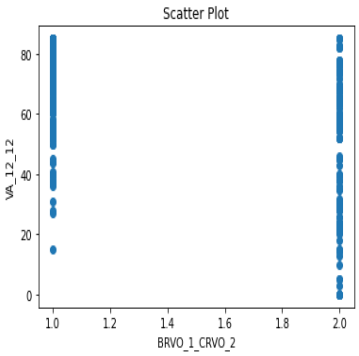
4.1 Electronic Medical Records (EMRs)

This section describes the EMR at the Royal Victoria Infirmary (RVI), Newcastle upon Tyne, which was interrogated for cases that received anti-VEGF intravitreal injections over 1 year. 244 eligible eyes were identified (129 left eye, 115 right eye, 110 central, 134 branch, 133 male, 111 female, mean age 71.94, mean delay between diagnosis and treatment 117.49 days) with a mean VA at treatment initiation of 49.85 ETDRS letters and a mean acuity at 1-year of 58.20 letters following treatment.

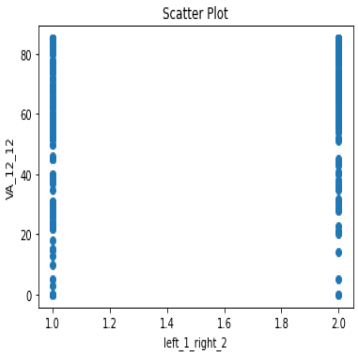
L=1, R=2	BRVO=1, CRVO=2	F=1, M=2	Age	VA at 1st Injection	Days Of Treatment Delay	VA One Year
1	2	1	75.89	68	53	60
1	2	2	82.64	37	0	52
2	1	2	69.51	35	0	36
2	1	1	79.75	63	20	79
2	2	1	71.7	75	0	38

Table 4.1: Sample of electronic medical records at the Royal Victoria Infirmary.

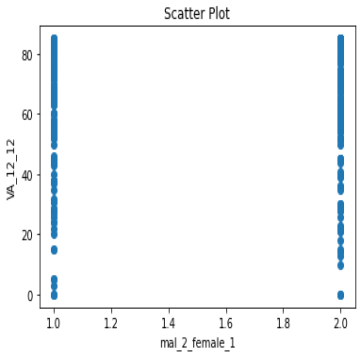
Table 4.1 shows a sample of EMR for RVO patients including 6 features: left and right eye (L=1, R=2), RVO types (BRVO=1, CRVO=2), gender (F=1, M=2), age, VA at the first injection (VA at first injection), and delay between diagnosis and treatment (days of treatment delay). The target represents VA after 12 months of the treatment (VA one year). Figure 4.1 shows the relationship between the features of the EMR with the target VA after 12 months of treatment. It is clear in Figure 4.1(a) that there were more BRVO than CRVO cases, with CRVO representation 47% and BRVO 53%. From Figure 4.1(b), left eyes were more affected at 52%, while right eyes affected with RVO were 51% female and 49% male. However, increases were found among elderly people, as Figure 4.1(d) shows RVO prevalence representation ranging with age from 45-90 years old. It is observed from Figure 4.1(e) those patients with high VA recorded in the first visit have more ability to improve their VA after 12 months of VEGF treatment. Additionally, it is clear from Figure 4.1(f) that delay to the start of treatment can affect the responses of RVO patients to the treatment.



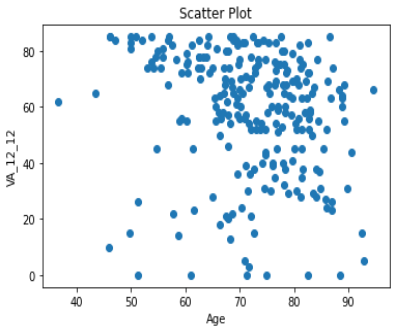
(a) The occurrence of Branch and centre retinal vein occlusion.



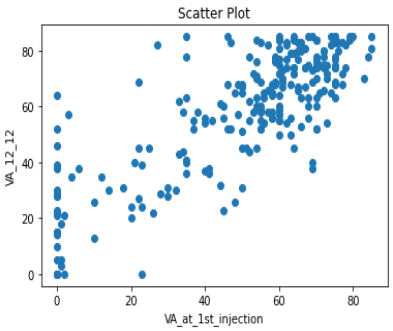
(b) Occurrence in the left and right eyes of the retinal vein occlusion patients.



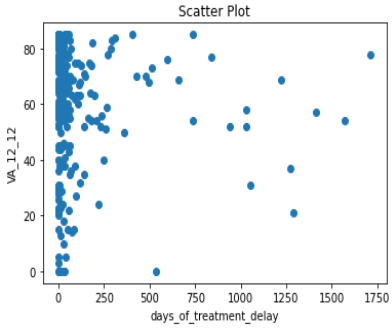
(c) Occurrence of RVO patient's sex.



(d) Relationship between the RVO patient 's age and VA after 12 months



(e) Relationship between the VA in the first visit and VA after 12 months treatment.



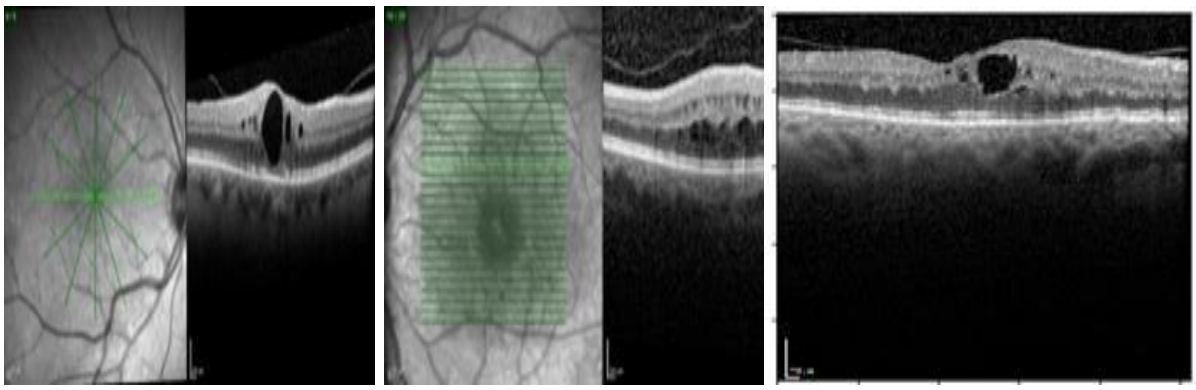
(f) Relationship between the delay between diagnosis and treatment and VA after 12 months

Figure 4.1: Distribution of EMR features over the target of VA after 12 months of VEGF therapy.

4.2 Optical Coherence Tomography (OCT)

Each row in the structure of EMR data has a unique number which represents the identification number for a specific patient's information. The role of this ID is to connect the patient's information with their images. In preparing the dataset, patients with missing OCT data were identified and excluded. Additionally, each image has two parts: an en-face and a cross-sectional image. Each OCT scan set contains either 6 or 25 different cross-sectional images arranged radially or orthogonally and centred on the fovea. The images are labelled from 0-5 or 0-24 respectively. These two styles represent a transition in OCT

protocols which occurred during the period in which the patients presented to the RVI. Images with the numbers 3 and 15 on each scale were chosen, as shown in Figure 4.2 (a) and (b) respectively, and these images were cropped to obtain only the retina part, as shown in Figures 4.2 (c). Several recent studies have found potential biomarkers for prognosis from OCT scans which can help to predict the vision outcome in cases of RVO. From the literature, a number of automated studies have focused on detection and classification for one or both types of RVO using OCT scans. Experimental results from these studies demonstrate good diagnostic accuracy; however, there is no published work using such an approach to predict prognosis or inform treatment plans.



(a) OCT image with scale 3. (b) OCT image with scale 15. (c) OCT cropped image.

Figure 4.2: OCT images containing 6 and 25 different cross-sectional images arranged radially and orthogonally: (a) OCT image with scale 3 chosen from 6 cross-sectional images; (b) OCT image with scale 15 chosen from 25 cross-sectional images; (c) OCT image after cropping.

Chapter 5 VISUAL PROGNOSIS PREDICTION IN RETINAL VEIN OCCLUSION USING LEARNING REGRESSION AND A MACHINE LEARNING APPROACH

RVO is the second most common retinal vascular disease and can cause sight loss through MO [61]. With anti-VEGF intravitreal injections (IVI), many patients regain vision but around a third suffer from persistent MO and limited visual improvement despite treatment in real-world settings[1]. Given the expense, inconvenience to patients and clinical risk from IVI it would be valuable to identify patients who are unlikely to gain visually from treatment. Over the past two decades, machine learning has progressed dramatically. It has been widely employed in both imaging, categorical and scalar data analysis [122-124]. Most recent publications on ophthalmic data analysis have focused on machine learning, which outperforms legacy algorithms in different applications in the grading and detection of diseases, where classification and segmentation are the two problems most commonly addressed[125]. The aim of this chapter is to apply a machine learning technique and compare its performance with that of linear regression using an EMR dataset to improve prognostication in RVO complicated by MO and to better inform treatment discussions with patients.

This chapter gives a brief overview of linear regression and machine learning, and particularly the Random Forest regression technique. Then, experiments and their results are illustrated. Finally, a summary concludes the chapter.

5.1 Linear Regression

Linear regression is one of the most commonly used techniques to determine correlations between two or more variables and to make predictions for a topic by utilising such relations (Figure 5.1). It helps to find linear relationships between a target and one or more predictors. Linear regression is popular in many applications because it provides the most straightforward technique to model a regression function as a linear combination of predictors. Moreover, the model parameters are easily interpretable. In addition, linear regression is the foundation for many modern modelling techniques. Notably, it provides a satisfactory approximation of the underlying regression function when the sample size is small[126, 127]. The basic idea of regression is to observe two things: firstly, does a set of predictor variables do a good job in predicting an outcome (dependent) variable? Secondly, which variables can be significant predictors of the outcome variable? Of two types of linear regression, simple linear regression is used to find the relationship between an independent variable and a dependent variable. Regression using a single independent variable is called

univariate regression, whereas using two or more independent variables is called multivariate regression [128]. Simple linear regression is as stated in equation 5.1:

$$y = \beta_0 + \beta_1 x + \in \qquad (5.1)$$

where y is the dependent variable or target, x is the independent variable or predictor, β_0 is the y intercept or constant, and β_1 is the slope of x that can be interpreted as the change in the mean of y for a unit change in x. The values β_0 and β_1 should be selected so as to minimise error and the goal is to obtain a line that best reduces the error. Ordinary least squares regression (OLS) is the standard approach for the estimation of the coefficients ($\beta_0$and β_1) of linear regression, describing the relationship between one or more independent quantitative variables and a dependent variable. The OLS technique attempts to minimise the residual sum of squares (RSS), which is a measure of the variance between the actual data and estimated data. The goal of OLS is to minimise the differences between the observed responses in the dataset and the responses predicted by the linear approximation of the data [129]. Thus, the objective of OLS is to minimize:

$$RSS = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \qquad (5.2)$$

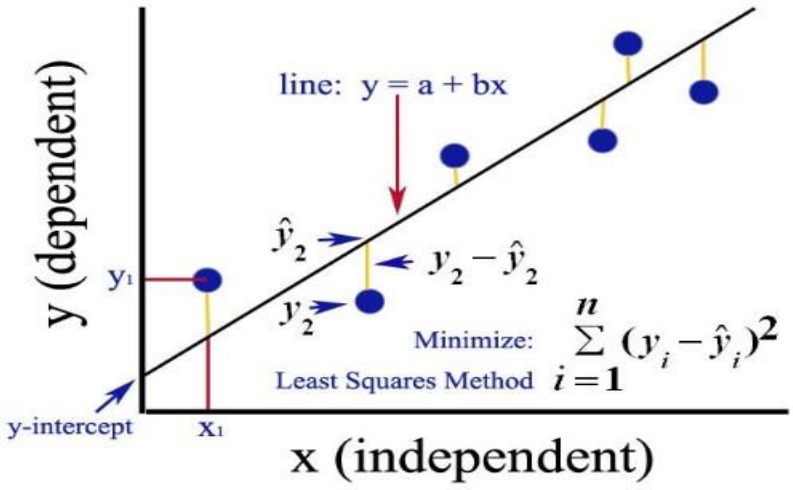


Figure 5.1: Linear regression and the least squares method [130].

The least-squares criteria are explained in equations 5.3 and 5.4:

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \tag{5.3}$$

$$\beta_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{5.4}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{5.5}$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{5.6}$$

where $\bar{x}$ and $\bar{y}$ (equations 5.5 and 5.6 respectively) are the means of the independent variable x and dependent variables y respectively for i= 1, 2, ..., n.

Multiple linear regression (MLR) is a statistical technique that utilises several dependent variables to predict the outcome of an independent variable. In fact, MLR is based on several assumptions, including that there is a linear relationship between dependent variables and independent variables [131]. Also, the independent variables may not be highly correlated with each other. The general equation for MLR is:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon \tag{5.7}$$

The purpose of most applications that can be applied using linear regression fall into the following three categories:

1. The establishment of a causal relationship between a response variable y and independent variables $x_1, x_2, \ldots, x_n$.

2. In cases where the target is prediction, linear regression is used to fit a predictive model to an observed dataset of y and x variables. Having completed the development of such a model, for any additional value of x given without its accompanying value of y, the fitted model can be used to make a prediction of the value of y.

3. Independent variables $x_1, x_2, \ldots, x_n$ are filtered to identify which are more important than others to help determine variable y more accurately and efficiently. This means

that linear regression can be applied to quantify the strength of the relationship between x variables and y .

5.2 Machine Learning

It is a subset of artificial intelligence (AI) and computer science that enables systems to learn and improve automatically through experience without being explicitly programmed. It uses data and algorithms to imitate how humans learn, gradually improving its accuracy. ML has dramatically progressed over the past two decades and has emerged as a method of choice in the development of practical software for computer vision, natural language processing, speech recognition, robot control, and many other applications. ML algorithms build a model based on sample data known as training data to make predictions or classifications without human intervention or assistance. To perform machine learning, firstly, we must have a data set and $D = x_{1,}\, x_2 \ldots, x_m$ is a data set where each record contains a description of an object or event, and *m* represents several records also called instances or samples. The object's description is called the attributes *d* of features, and their values are called attribute values. Each instance $x_i = (x_{i1}, x_{i2}, \ldots, x_{id})\, \epsilon\, x$ is a vector in the *d*-dimensional sample space x, where d is called the dimensionality of the instance x_i, and x_{ij} is the value of the jth attribute of the instance x_i. Learning or training is the mechanism used by ML algorithms to build models from data. The data used in the training stage is called training data, and the set of all training examples is called a training set. The learned model corresponds to the rules about the data, and these rules are called a hypothesis and the actual underlying rules are called ground truth or facts. For training an effective prediction model, outcome data must be available. The outcome of a sample is called a label, which can be written for the ith sample as (x_i, y_i), where $y_i \in y$ is the label of the sample x_i, and y_i is the set of all labels[132]. The goal of ML algorithms is to detect or approximate ground truth. In the field of ML, there are two fundamental learning types, namely supervised and unsupervised methods. In supervised learning, the data set contains the input or features as well as the output labels (also known as the target responses). This type of learning method seeks to learn a function that approximates the relationship between the feature values and the labels so that good predictions can be generated for all possible inputs. However, unsupervised learning is appropriate when labels are not provided and is suitable for problems that demand algorithms to find and extract similarities among inputs so that similar inputs will be

categorised together. Unsupervised learning can be further grouped into two types: clustering and association. In the clustering method, objects are divided into clusters depending on their similarities and dissimilarities. Association is a rule-based ML technique used to discover the possibility of the co-occurrence of items in a collection. It is found that most practical ML utilises supervised learning in most applications, which can be separated into classification and regression problems. In case the prediction output is discrete, it is called a classification problem, whereas when the prediction output is continuous it is called a regression problem[133]. In the present study, supervised learning and regression algorithms have been used. In particular, the Random Forest regressor has been applied to detect VA after 12 months of VEGF treatment.

5.3.1 Random Forest (RF)

A Random Forest is a collection of tree predictors. The fundamental idea behind this approach is to combine multiple decision trees in determining the final output instead of relying on an individual decision tree [10]. RF is a 'bagging' technique and supervised learning algorithm that utilises ensemble learning for regression and classification. In other words, RF uses multiple decision trees and a technique called bootstrap and aggregation (Figure 5.2), usually known as bagging, to perform both classification and regression tasks. Random row and feature sampling are performed from the dataset to generate sample datasets for every model, and this technique is called bootstrapping. In the case of a classification problem, the final output is taken by using the majority voting classifier for all the outputs generated from the bootstrap. In the case of a regression problem the final output is the mean of all the outputs and this stage is called aggregation. Every decision tree has high variance; however, when all of them are combined together using the RF technique in parallel then the resulting variance is low since each decision tree is trained on that particular sample data. Hence the output does not depend on one decision tree but on multiple decision trees[134]. Given a training set $X = x_1, x_2, \ldots, x_n$ with target $Y = y_1, y_2, \ldots, y_n$ repeatedly bagged *B* times, selects a random sample with replacement of the training set and fits trees to these samples. For $b = 1, \ldots, B$ in which, at first, sample with replacement, n training samples from *X, Y*, call these X_b, Y_b. Then, a classification or regression tree f_b on X_b, Y_b. After finishing the previous steps, predictions for unseen samples x' can be found by averaging the predictions from all individual regression trees on x', as shown in equation 5.8:

$$\hat{f} = \frac{1}{B}\sum_{b=1}^{B} f_b(x') \quad (5.8)$$

Alternatively, a majority vote can be taken in the case of classification trees. This procedure can lead to superior performance because it reduces the variance of the model without increasing the bias (Figure 5.3). One of the advantages of Random Forest is that it ensures that the model does not rely heavily on any individual feature, and the number of features which can be separated to each node is limited to some percentage of the total (known as the hyperparameter). Additionally, each tree generates a random sample from the original data set when producing its splits, adding extra random elements which can help in preventing overfitting. These advantages help avoid the tree from being highly correlated. It is very simple and parallelised, and relatively strong against outliers and noise [135]. A further advantage is that Random Forest can be considered as one of the most accurate learning algorithms available, generating highly accurate detectors for many datasets. Also, it provides estimates of what variables are important in regression or classification[136]. RF generated feature importance by calculating the decrease in node impurity using Gini index that results from splitting on a particular feature. When a random forest is built, at each spilt in each tree, a random subset of features is considered for the split. For each feature considered, the Gini index is calculated before and after the spilt. The difference in Gini index, averaged over all trees in the forest, is used to determine the importance of that feature.

Feature	Feature importance
Baseline visual acuity	0.6303
Age	0.1881
Delay between diagnosis and treatment	0.1186
Central and branch RVO diagnosis	0.0267
Left and right eye	0.0197
Male/female	0.0126

Table 5.1:Feature importance of the main six features.

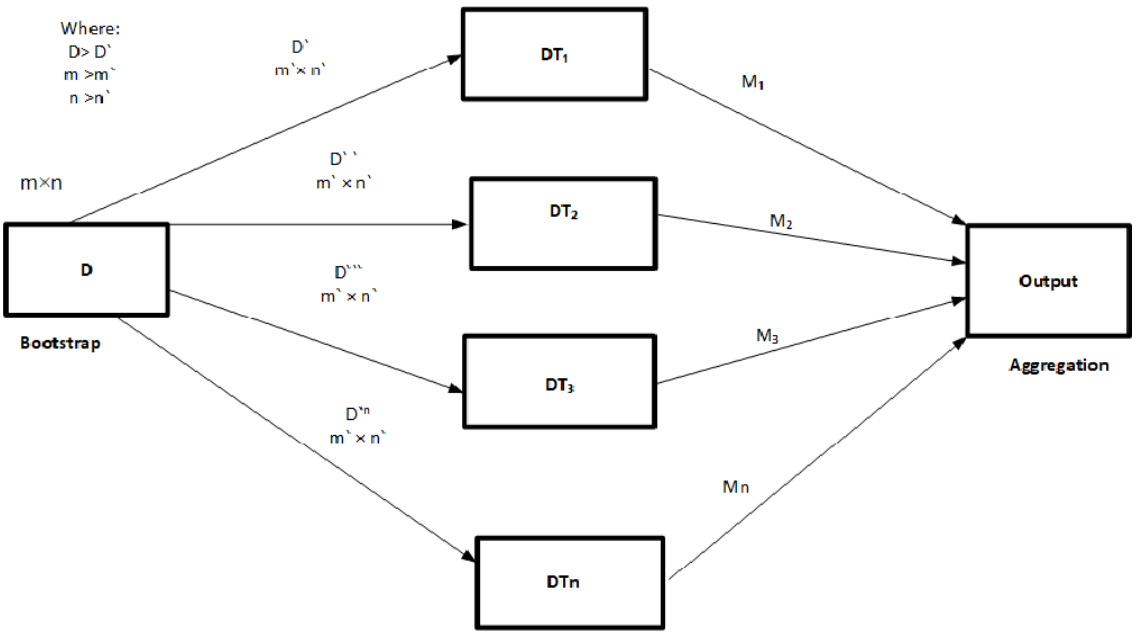


Figure 5.2: Bootstrap and aggregation parts used in the Random Forest technique.

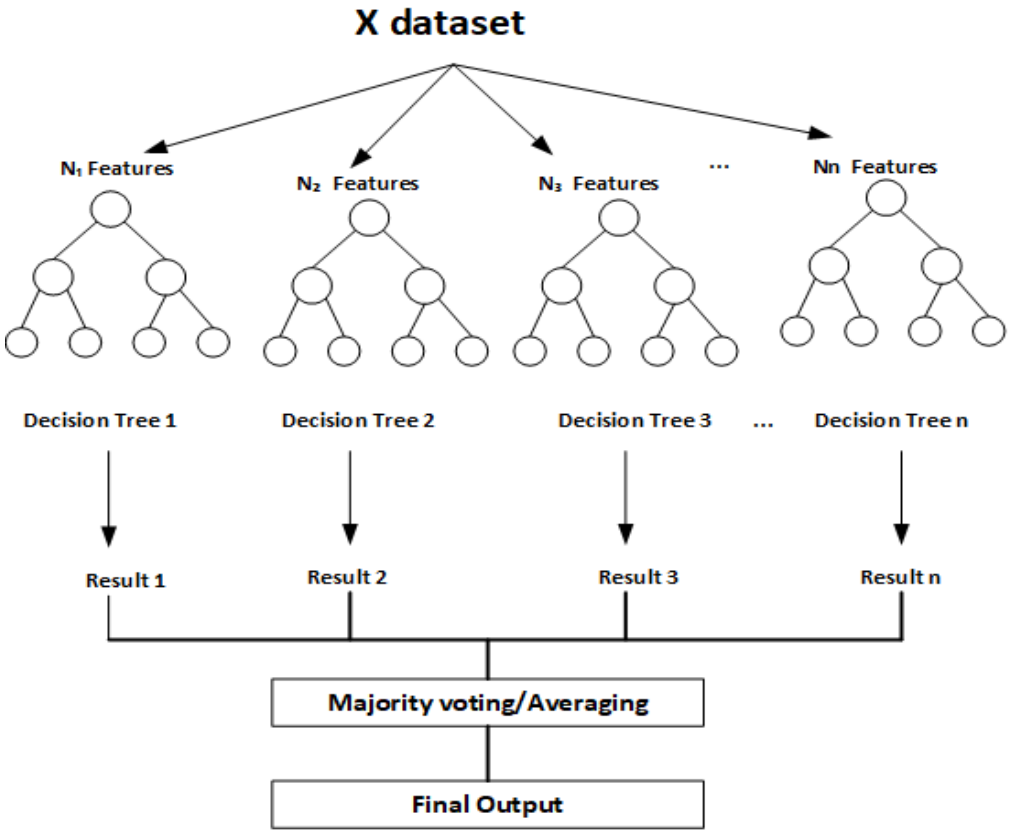


Figure 5.3 : Random Forest classifier/ regressor.

Algorithm Random Forest

Input: training set $D = (x_1, y_i)$, *...*(x_n, y_n)*, features F, and number of trees M. Output: prediction of the Random Forest at y.*

for *k=1 to F* ***do***

1. *Randomly select* k *features with (or without) replacement from total F features.*
2. *Among the "k" features, calculate the node "d" using the best split point.*
3. *Split the node into children nodes using the best split.*
4. *Repeat 1 to 3 steps until "l" number of nodes has been reached.*
5. *Build forest by repeating steps 1 to 4 for "n" number times to create "n" number of trees.*

6. ***end for***

7. *For prediction: take the test features and use the rules of each randomly created decision tree to predict the outcome and stores the predicted outcome y.*
8. *Compute the average (MAE) for each predicted target, which is considered as the final prediction from the Random Forest algorithm.*

5.3 Experiment and Results

5.3.1 Data Set

The EMR at the Royal Victoria Infirmary were interrogated for cases that received anti-VEGF intravitreal injections over 1 year or more for MO secondary to RVO. 428 eligible eyes were identified (222 left eyes, 200 central RVO, 213 males, mean age 72.6, mean interval between diagnosis and treatment 118 days). Mean baseline VA was 50.6 ETDRS letters improving to 59.0 letters following 1 year of anti-VEGF treatment.

5.3.2 Feature Importance

The RF algorithm has a built-in feature importance function which is computed using several approaches, one of which is called Gini importance (mean decrease in impurity), which is calculated from the RF structure. The RF is a set of decision trees, and each decision tree has a set of internal nodes and leaves. The features for internal nodes are selected with a

specific criterion which may be Gini impurity or information gain for classification and variance reduction for regression. Therefore, it can be determined how each feature, on average, decreases impurity. The average over all trees in the forest is the measure of feature importance, in which a higher value signifies the greater importance of the feature. This technique is available for both the classifier and regressor in the scikit-learn implementation in Python. Scikit-learn is an open-source data analysis library and the key library for the Python programming language that is employed in ML projects. It is a free software focuses on ML tools including statistical, mathematical, and many various types of algorithms which developed for ML and related technologies[137].

The most significant advantage of this method is the speed of computation, where all values needed are computed during RF training. When training a tree, a calculation is required of how much each feature diminishes the impurity. The more a feature reduces impurity, the more influential that feature is. In random forests, the decline in impurity for each feature is averaged across trees to determine the final importance of the variable. A feature's importance defines its contribution towards the output of the Random Forest regression model. This helps in giving the developer knowledge of the features and gaining a better understanding of the model's logic, so that the improvement of the model focuses only on the important variables and avoids the 'black box' effect. In this work, which involves a regression problem, scikit-learn in Python is utilised to calculate the reduction variance using the mean square error (MSE) as follows[138]:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \mu)^2 \qquad (5.9)$$

Where y_i is the label (target variable) for an instance, *N* is the number of instances, and μ is the mean given by $\frac{1}{N}\sum_{i=1}^{N} y_i$.

From this [139], the mean absolute error (MAE) is:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|(y_i - \mu)| \qquad (5.10)$$

Where y_i is a target variable for an instance, *N* is the number of instances, and μ is the mean given by $\frac{1}{N}\sum_{i=1}^{N} y_i$.

The features used in this work are the following:

1. **Baseline visual acuity**: VA was assessed on the first visit to a clinic for RVO patients.
2. **Age**: the age of RVO patients at the first clinic visit.
3. **Delay between diagnosis and treatment**: the delay between the diagnosis of RVO patients and taking the first anti-VEGF injection.
4. **Central and branch RVO diagnosis**: the type of RVO disease.
5. **Left and right eye**: the eye affected by RVO disease, either left eye or right eye**.**
6. **Male/female:** the RVO patient's sex.

The response variable (target variable) is the visual acuity of RVO patients after 12 months of VEGF treatment. Table 5.1 explains the importance of each feature used in the method. The Random Forest regressor and linear regression have been utilised, as contained within the scikit-learn package ("sklearn.linear model. Linear Regression", "sklearn.ensemble. Random Forest Regressor", version 1.1.1, jupyter-notebook version 6.0.3) to predict VA for RVO patients after 12 months of anti-VEGF treatment. The regression analysis was carried out using the dataset of EMR collected from the Royal Victoria Infirmary in Newcastle upon Tyne. The dataset consists of 428 instances and in turn contains 6 features. The dataset is automatically separated into 80% (342 eyes) for the training and 20% (86 eyes) for the testing datasets. In implementing Random Forest regressor, the number of trees in the model was 300, the number of features were the 6 described above. To calculate the correctness of predications made by the model, R-squared (R^2) coefficient used to measure the accuracy of the Linear regression and Random Forest regression. This metric quantifies the difference between the predicted values and the actual values of the target variable. Values of R-squared (R2) were used to assess the method's performance. R2 is calculated as shown in equation 5.11 below:

$$R^2 = 1 - \frac{\sum(Y(\text{actual}) - Y(predicted))^2}{\sum(\text{Y(actual)} - Y(mean))^2} \qquad (5.11)$$

Linear regression produced a model accounting for 57% of the accuracy seen in 1-year visual acuity (R^2=0.57). Using the same data, Random Forest regression surpassed this with the model accounting for 62% of accuracy (R^2=0.62). Other measures of the Random Forest model's prognostic ability include a mean absolute error of 9.8 (MAE) and mean squared

error (MSE) of 163.02. MAE represents the number of ETDRS letters of disagreement that can be expected between the predicted and observed 1-year VA levels for an individual, whereas the MSE highlights outliers where VA was predicted particularly poorly. This description helps in understanding the viability of such prognostic models in the clinic, but also facilitates external comparison. Notably, this application of machine learning has already demonstrated comparable prognostic capabilities to similar methods applied to age-related macular degeneration when imaging also contributed data [85]. Machine learning can provide clinicians with individualised visual prognosis at baseline. Given the significant minority of individuals found to have little visual gain from serial IVIs, a dependable prognostic tool could be of real value to patients in deciding whether or not to commit to treatment.

5.3.3 Adding Extra Features to the Experiment

It is worth mentioning that the fundamental goal of this research is to help RVO patients and ophthalmologists in deciding whether or not to continue having the VEGF injections. This decision should be taken in relation to the baseline EMR representing the six features investigated above. Two experiments have been investigated. Firstly, by joining VA after three months of treatment to the initial six features (gender, age, RVO's type, left or right eye, delay between diagnosis and treatment, and baseline visual acuity). In the second experiment, VA after six months has been combined to the initial six features.

5.3.3.1 Adding Visual Acuity after Three Months of VEGF Treatment

In addition to these six features, VA after three months of treatment is added to the algorithm. After implementing the Random Forest and linear regression, the accuracy of the model accounts for 74% (R^2=0.74) and 67% (R^2=0.67) respectively.

5.3.3.2 Adding Visual Acuity after Six Months of VEGF Treatment

Adding the feature of VA after six months to the main six features improves the model with an accuracy of 87% (R^2=0.87) for Random Forest regression and 84% (R^2=0.84) for linear regression. The results showed a noticeable improvement in both regression techniques by adding these features. Meanwhile, the Random Forest (RF) still outperforms linear Regression in all cases. Although the complexity of RF is higher than Linear Regression as it requires building multiple decision trees and combining their results, the strength of RF comes from the following reasons: in terms of accuracy, RF can handle non-linear relationships and noisy data better than Linear Regression, making it more accurate in many cases. In contrast, Linear

Regression can perform well when a linear relationship exists between the input features and the target variable. However, it may need to capture non-linear relationships better [140]. Random Forest Regression can help avoid overfitting by combining the results of multiple decision trees and using ensemble learning techniques. Linear Regression is prone to overfitting when there are many input features, and the number of training samples is low. Linear Regression is a parametric approach that assumes a linear relationship between the input features and the target variable. Its complexity increases linearly with the number of input features [141, 142]. Table 5.2 describes the feature importance of all eight features. Figure 5.4 illustrates distribution of the feature variables with respect to the target variable

Feature	**Feature importance**
Baseline visual acuity	0.6303
Visual acuity after three months of treatment	0.8537
Visual acuity after six months of treatment	0.9289
Age	0.1881
Delay between diagnosis and treatment	0.1186
Central and branch RVO diagnosis	0.0267
Left and right eye	0.0197
Male/female	0.0166

Table 5.2: Feature importance for the eight features.

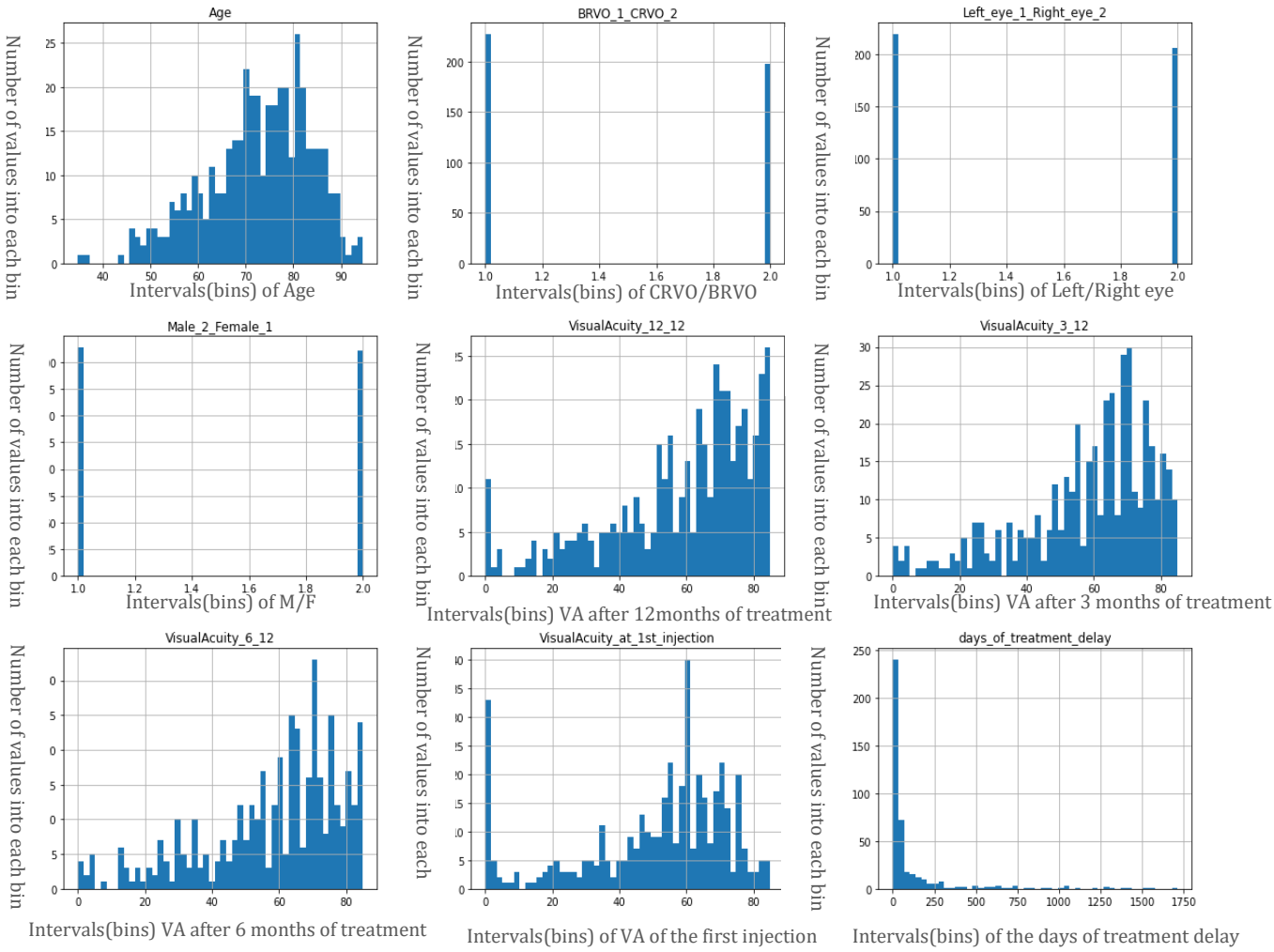


Figure *5.4:Distribution of the feature variables with respect to the target variable: visual acuity of RVO patients after 12 months of treatment.*

5.4 Summary

In this study, Random Forest and linear regression were compared to predict VA after 12 months of anti-VEGF treatment for RVO patients using EMR only as a dataset. The study firstly analysed EMR features at the baseline visit. Then, VA after three months of VEGF therapy was included in the analysis and following this, another experiment was investigated by adding VA after six months of treatment. In all cases, the performance of Random Forest regression outperformed linear regression. To conclude, when using primary demographic data, machine learning techniques demonstrate a superior ability to predict visual prognosis for patients with RVO. Unlike traditional statistical techniques, this approach also has the potential to draw on retinal imaging to improve its performance even further. Such an enhanced tool could help clinicians to personalise patient information and commissioners to

tailor cost-utility analyses for each case. Analysing OCT images, combing them with EMR and using the combination as input into machine learning techniques could improve the system's performance. The next chapter aims to employ feature extraction methods to analyse OCT images and evaluate several machine learning techniques by comparing their performance with linear regression for predicting VA after one year of treatment.

Chapter 6 PREDICTING VISUAL ACUITY OF RVO PATIENTS AFTER 1-YEAR USING SEVERAL MACHINE LEARNING TECHNIQUES AND HOG AND LBP FEATURE EXTRACTION METHODS

This study aimed to apply machine learning techniques to improve prognostication in RVO complicated by MO to better inform treatment discussions with patients. Five machine learning techniques and linear regression were implemented for the prediction of VA after one year of anti-VEGF treatment for RVO patients. Two common feature extraction methods were used to extract important information from OCT images, and these feature vectors were combined with the EMR data structure which represents the input data for machine learning techniques and linear regression. The rest of this chapter is organised as follows; the next section presents the materials and proposed method; the experiments and results are introduced in Section 6.3 and Section 6.4 concludes the chapter.

6.1 Materials and Method

The proposed method is outlined in Figure 6.1. The methodology is formulated as a regression procedure. First, the EMR data structure described in detail in Section 6.2.1, is prepared and any cases with missing data are eliminated. Then, the OCT images as presented in detail in Section 6.2.2 are cropped and their useful information is extracted using HOG and LBP methods. Next, the feature vectors generated from the OCT images by the feature extraction methods and EMR are combined to represent the input data for the machine learning techniques and linear regression. Several machine learning techniques are described in section 6.2.3. Finally, a regression step is implemented using linear regression and a number of machine learning methods are performed to predict VA after 1 year after treatment, as discussed in detail in section 6.3. The dataset is split into 80% for training and 20% for testing the system.

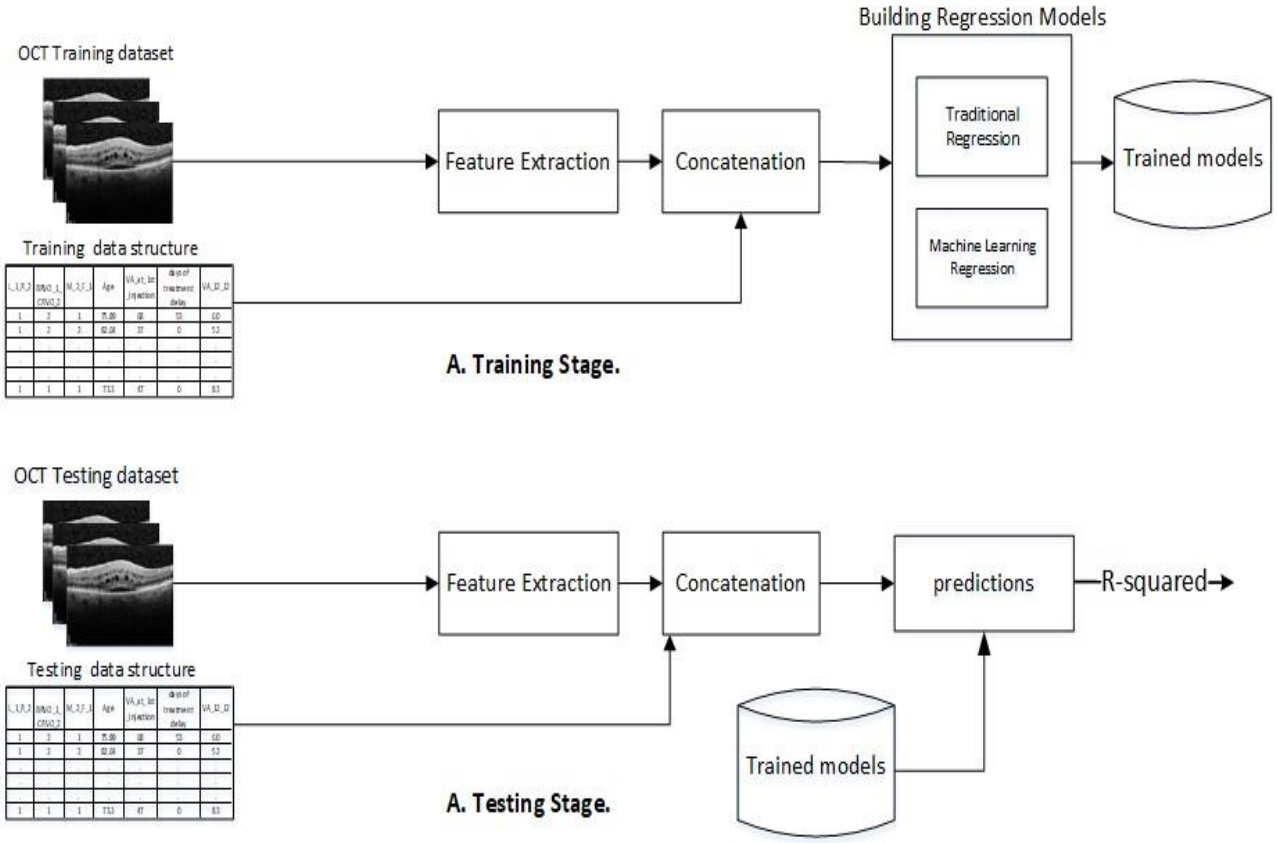


Figure 6.1:Proposed technique for predicting the variability of individual prognosis from the anti-VEGF treatment of RVO.

Left eye =1, Right eye=2	**BRVO=1, CRVO=2**	**female=1, meal =2**	**Age**	**Visual acuity at 1st_injection**	**Days of treatment delay**	**Visual acuity 1 year**
1	2	1	75.89	68	53	60
1	2	2	82.64	37	0	52
2	1	2	69.51	35	0	36
2	1	1	79.75	63	20	79
1	1	2	71.17	0	16	39
1	1	1	73.30	47	0	83
2	1	2	71.06	54	140	71
2	2	1	46.19	69	405	85
2	2	1	70.85	0	0	5
2	1	1	68.13	67	0	74
2	1	1	62.51	59	1713	78

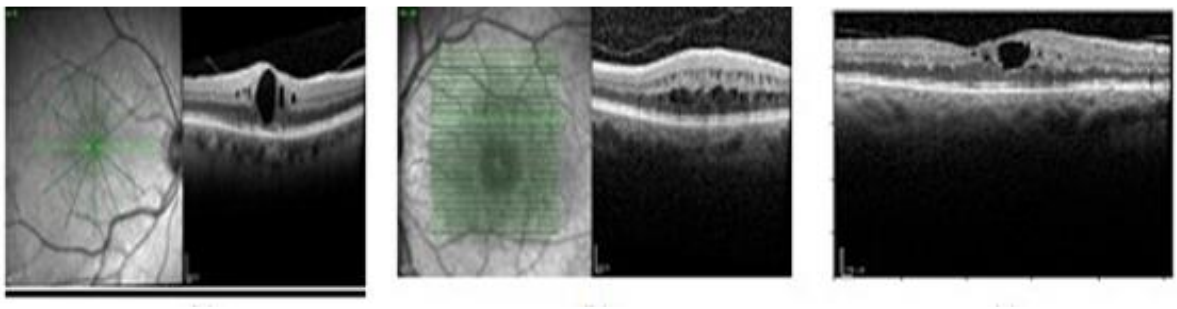
Table 6.1:Sample of the electronic medical record at the Royal Victoria Infirmary hospital.

6.1.1 Data Structure (DS) of Electronic Medical Records.

This section described the EMR at the Royal Victoria Infirmary which were interrogated for cases that received anti-VEGF intravitreal injections over 1 year. 244 eligible eyes were identified (129 left eye, 115 right eye, 110 central, 134 branch, 133 male, 111 female, mean age 71.94, mean delay between diagnosis and treatment 117.49) with a mean visual acuity at treatment initiation of 67.5 ETDRS letters and a mean VA at 1 year of 58.20 letters. Table 6.2 shows a sample of the electronic medical records at the Royal Victoria Infirmary.

6.1.2 OCT Image preparation

Each row in the structural data has an ID number which represents the identification of a specific patient's information. The role of this ID is to connect the patient's information with their images. To prepare the images, a comparison of the available data structure and image dataset has been conducted. Cases of missing data and unmatched imaging data are the omitted in order to make sure that each row in the data structure is complete. Additionally, each image has two parts, the cross-sectional OCT retina image, and the infra-red en-face scout image of the retina, where the position of each associated cross-sectional OCT image is denoted by a green line and given a number. During the time period over which the study images were taken, two different OCT scanning protocols were in use at the RVI and so there are two configurations of OCT cross-sectional image placements on the en-face scout images. OCT images with the numbers 3 and 15 for each scale have been chosen as shown in Figure. 6.2 (a) and (b) respectively, while Figure. 6.2 (c) illustrates a separate OCT image after cropping.



(a) (b) (c)

Figure 6.2: Sample of OCT images used to predict a particular value based on a set of a previous data: (a) scale 3 image; (b) scale 15 image; (c) cropped image.

6.1.3 Machine Learning Techniques

For the prediction of 1-year VA of cases that received anti-VEGF intravitreal injection, learning regression and different machine learning techniques were applied, including Ridge Regression, LASSO Regression, Multi-layer Perceptron (MLP) Regression Neural Network, PLS Regression, and Random Forest Regression. These regression methods fall within the category of supervised machine learning. These ML techniques will be briefly explained below, except Random Forest because it has been illustrated in chapter 5.

1. Ridge Regression

Ridge regression is a common and widely used parameter estimation technique employed to address the collinearity issue repeatedly arising in multiple linear regression. When this happens, least squares estimates are unbiased. However, their variances can be significant, and could be far from the actual value. Ridge regression performs L2 regularisation to reduce overfitting. This decreases the error by adding a degree of bias to the regression estimates. Ridge regression addresses some of the problems of ordinary least squares (OLS) by composing a penalty on the size of the coefficients[6] . The term ridge regression refers to a linear regression model whose coefficients are estimated by a ridge estimator instead of OLS, as explained in the previous chapter. Even though the ridge estimator is biased, it still gives lower variance than the OLS estimator. Ridge regression utilises a slightly different formula as described in equation 6.1:

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}\beta_j^2=RSS+\lambda\sum_{j=1}^{p}\beta_j^2 \quad (6.1)$$

where a set of training data (x_1, y_1)...(x_n, y_n) is used from which the parameters β are estimated and $x_i=\left(x_{i1}, x_{i2}, \ldots, x_{ip}\right)^T$ is a vector of features of the *i*th case, and the coefficients $\beta=(\beta_0,\beta_1,\ldots,\beta_p)^T$. $\lambda\ \sum_{j=1}^{p}\beta_j^2$ represent the L2 penalty (or regularisation) as a penalty term for the loss function. Here, if $\lambda=0$ then the formula reverts to OLS, whereas if λ is very large this will add too much weight and leads to underfitting. Therefore, the choice of λ is very important in obtaining accurate results [143].

2. LASSO Regression

LASSO (least absolute shrinkage and selection operator) regression is a regularisation method. It is a type of linear regression that uses shrinkage and a variable selection method

for regression models. Shrinkage is where the values of the data are shrunk towards a central point. The L1 regularisation technique is performed to avoid the overfitting of data. The goal of regularisation is to reduce the magnitude of the coefficients instead of reducing the number of features, and the aim of LASSO is to identify the variables and corresponding regression coefficients that lead to a model which minimises prediction errors. The LASSO technique achieves this by imposing a constant on the parameters which shrinks the regression coefficients towards zero; that is, by imposing the sum of the absolute value of the regression coefficients to be less than λ. The selection of λ uses an automated k-fold cross-validation method. The idea behind this approach is that the dataset is partitioned randomly into *k* sub-samples of equal dataset size, whereas the *(k−1)* sub-samples are employed for the development of a prediction model and the remaining sub-sample is utilised for validation. This process is executed *k* times, with each of the *k* sub-samples in turn being used for model validation and the rest for the development. Then, the separate *k* validation results are combined for a range of λ values and the preferred λ is chosen and applied to determine the final model [7]. This process is utilised to perform feature selection automatically, adding a penalty equal to the absolute value of the magnitude of coefficients. For a dataset $(x_{ij}, y_j), i = 1,2, \ldots, n$, and $j = 1,2, \ldots, p$, x_{ij} are the independent variables, and y_i represents the responses. The mathematical equation for LASSO regression that solves the l_1 penalized regression problems in finding $\lambda = \lambda_j$ to minimise the residual sum of squares (RSS)+λ (sum of the absolute value of the magnitude of coefficients) is:

$$\sum_{i=1}^{n}\left(y_i - \sum_j x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}\left|\beta_j\right| \qquad (6.2)$$

where λ represents the amount of shrinkage (the coefficient used to penalize the weights). When λ= 0, all features are considered which means that it is equivalent to linear regression, while if λ = ∞ this denotes that no features are considered, which means more features are eliminated. With increases in λ the bias also increases, whereas with decreases in λ, the variance increases [144]. It is worth mentioning that LASSO and ridge regression are part of the linear regression family where input variables and output are assumed to have a linear relationship. Linear regression is considered as a basic form where a model is not penalised for its choice of weights. This means that, in the training stage of a model investigating one particular feature as most important, the model often ascribes a large weight to the feature. This may lead to overfitting problems. However, the ridge regression technique penalizes the model for the sum of squared value of the weights. So, not only are the weights likely to

have smaller absolute values, but instead, the extremes of the weights will tend to be penalised. In the LASSO technique, the model is penalised for the sum of absolute values of the weights, which means that the absolute weight values will be reduced and many of them will tend towards zero during the training stage [145].

3. Multi-layer Perceptron (MLP) Regression Neural Network

The MLP is a supervised learning technique which learns a function by training on a dataset, $f(.): \mathcal{R}^m \rightarrow \mathcal{R}^0$ where m is the number of dimensions for input and o is the number of dimensions for output. Significant benefits over other statistical models are asserted for these networks in that they can 'learn' response relationships of arbitrary complexity between input and output variables without human intervention and can be employed for predicting and controlling complex nonlinear systems [8]. For a set of features $X = x_1, x_2, \dots, x_m$ and a target y, MLP can learn a non-linear function approximator for either regression or classification[146]. In MLP Figure 6.3, there can be one or more non-linear layers between the input and the output layer called hidden layers. The input layer contains a set of neurons $x_1, x_2, \dots, x_m$called the input features. Each node computes a weighted sum of its inputs and passes the sum through a soft non-linearity process and passes this activation level through a transfer function to generate output. Each neuron in the hidden layer transforms a biased weighted summation $w_1x_1 + w_2x_2 + \cdots + w_mx_m$, and passes this activation level to a transfer function to create an output. Hyperbolic tangent sigmoid and logistic activation functions are considered to be most commonly used in MLP[147].

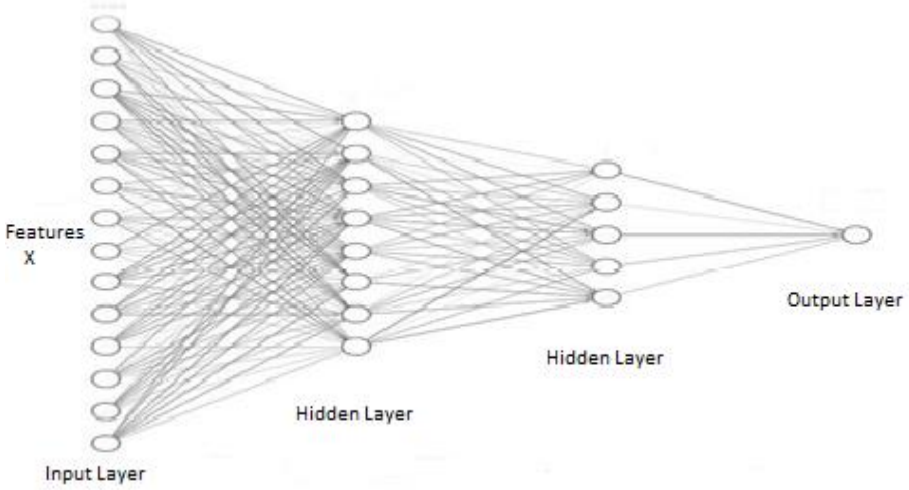


Figure 6.3: MLP with two hidden layers.

4. PLS Regression

PLS (partial least squares) regression is an algorithm which decreases the number of predictors to a small set of uncorrelated components and performs least squares regression on these components instead of the original input. A set of components is extracted that describes the closest relationship between the predictors and response variable. Cross-validation is usually utilised to recognise the smaller set of components which offer the greatest predictive ability. This technique is useful when the predictors are very collinear [9]. PLS is a hybrid combination of multiple regression and principal component analysis (PCA) [148].

The advantages of PLS are its ability to deal with multicollinearity, allowing it to take into account the data structure for both input variables and the target. It provides visual results that help in interpretation and is also called projection to latent structures which is a technique that reduces the number of explanatory variables to a smaller set of uncorrelated variables. The basic idea behind this is to find a low-dimensional set of input space variables that is most correlated with a given set of output data [149].

PLS regression searches for components from *X* that are also relevant for *Y*. Specifically, PLS try to find a set of components called latent vectors and performs a simultaneous decomposition of *X* and *Y* with the restriction that these components should explain as much as possible of the covariance between *X* and *Y*. This step generalises PCA and is followed by a regression where the decomposition of *X* is employed to predict *Y*. PLS regression first extracts a set of A latent variables $\mathrm{T} = [t_1, \ldots, t_A]$ from an $N \times K$ measurement matrix *X* of *N* samples and *K* variables demonstrating significant covariance with the response. Then, y is regressed against the $N \times A$ matrix T to establish the functional relationship between measurements and responses. PLS regression starts by using the response variable to estimate the weights *w* for the *X* matrix such that the covariance between X_w and *y* is maximised. The weight vector is further normalised to the unit norm and the X-scores are then estimated as t = X_w and *y* is regressed against *t*. Eventually, *X* and *y* are reduced to withdraw the variation explained by the current latent variable. The process is repeated until a cross-validation statistic indicates that no increase in model performance occurs when extracting additional latent variables [150].

6.1.4 Feature Extraction

6.1.4.1 Histogram of Oriented Gradients

The histogram of oriented gradients (HOG) is a feature descriptor used for the detection of objects in computer vision and image processing. The idea behind this technique is to calculate occurrences of gradient orientation in localised portions of an image. The HOG descriptor concentrates on the shape of the object or its structure. To compute features, it uses the magnitude as well as angle of the gradient, and for an image's region it creates histograms using the magnitude and orientations of the gradient. The first presentation of the HOG by Dalal and Triggs was in the context of detecting the human presence in images [151]. The descriptor of the image is computed by dividing it into small cells and computing histograms of oriented gradients for each cell and then integrating all of the resulting histograms. This approach was adopted in the present study to extract new features from local patches. The calculation of a HOG features is summarized to the following steps:

1. Resizing of OCT image Figure 6.4 into an image of 128x64 pixels Figure 6.5. This dimension was suggested by the authors in the original paper since they obtained better results in their task of detection[151]. However, it is optional and depends on the problem that needs to be solved.
2. Calculation of the gradient of the image. The gradient is obtained by joining magnitude and angle from the image. First, horizontal gradients G_x and vertical gradients G_y are calculated for each pixel. This can be accomplished by filtering the image using the kernels Figure 6.6. Then, the magnitude and the direction of the gradients can be calculated using the following formula:

$$\text{Magnitude}(\mu) = \sqrt{G_x^2 + G_y^2} \qquad (6.3)$$

$$\text{Angle}(\theta) = \left|\tan^{-}1\left(\frac{G_y}{G_x}\right)\right| \qquad (6.4)$$

3. The magnitude and angle matrices that are calculated in step 2 are divided into 8x8 cells to generate a block. For each block, a histogram is computed using nine bins, where each bin has an angle of 20 degrees, in order to obtain the intensity of each bin.
4. Once the histogram was calculated for all blocks, four blocks from the nine-bins histogram matrix were combined to form a new block (2x2). This grouping is obtained based on an overlapping manner with a stride of 8 pixels. Finally, all nine-bin histograms for each constituent cell are concatenated to form a 36-feature vector for all four cells in a block.
5. All histograms are normalised within a block of cells using the L2 norm to reduce the effect of changes in contrast among images of the same object. Based on the size of the image, there are 7 blocks in the horizontal direction and 15 blocks in the vertical direction, and each block 36-point feature vector is collected, so the total length of HOG features is 7 x 15 x 36 = 3780 HOG features in an image. Figure 6.7 shows the HOG features descriptor on OCT image.

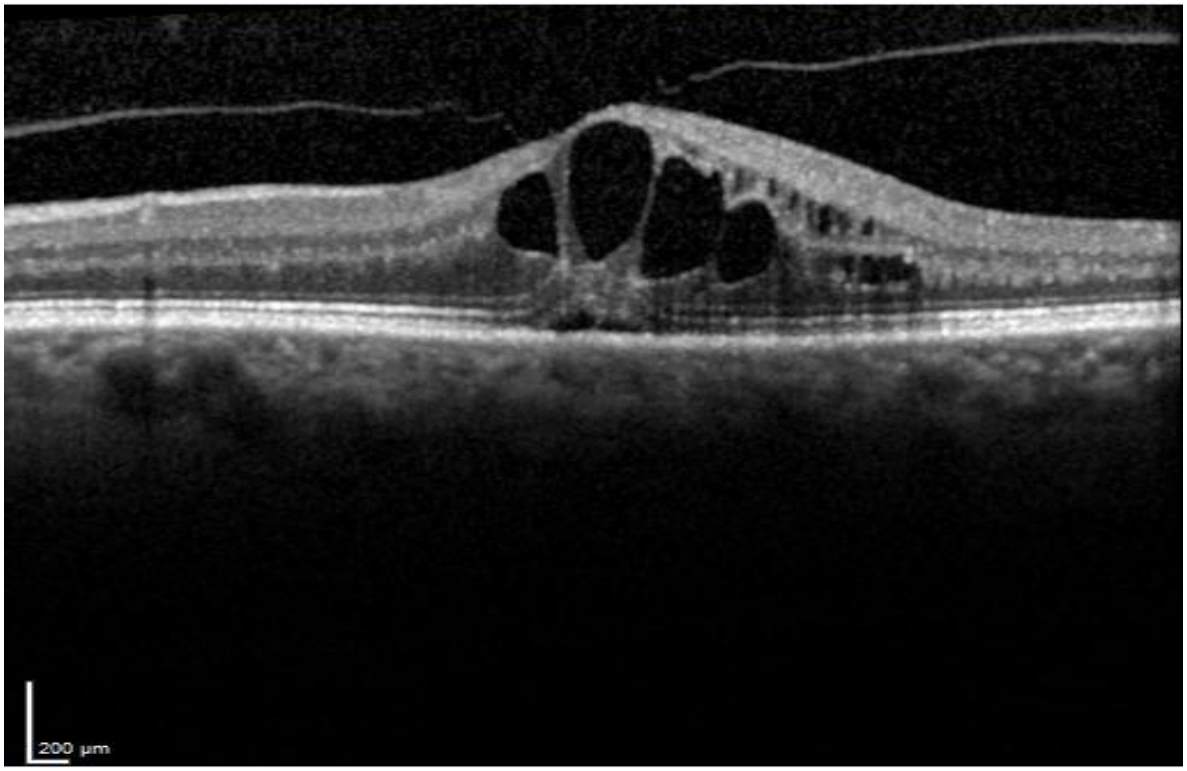


Figure 6.4: OCT image.

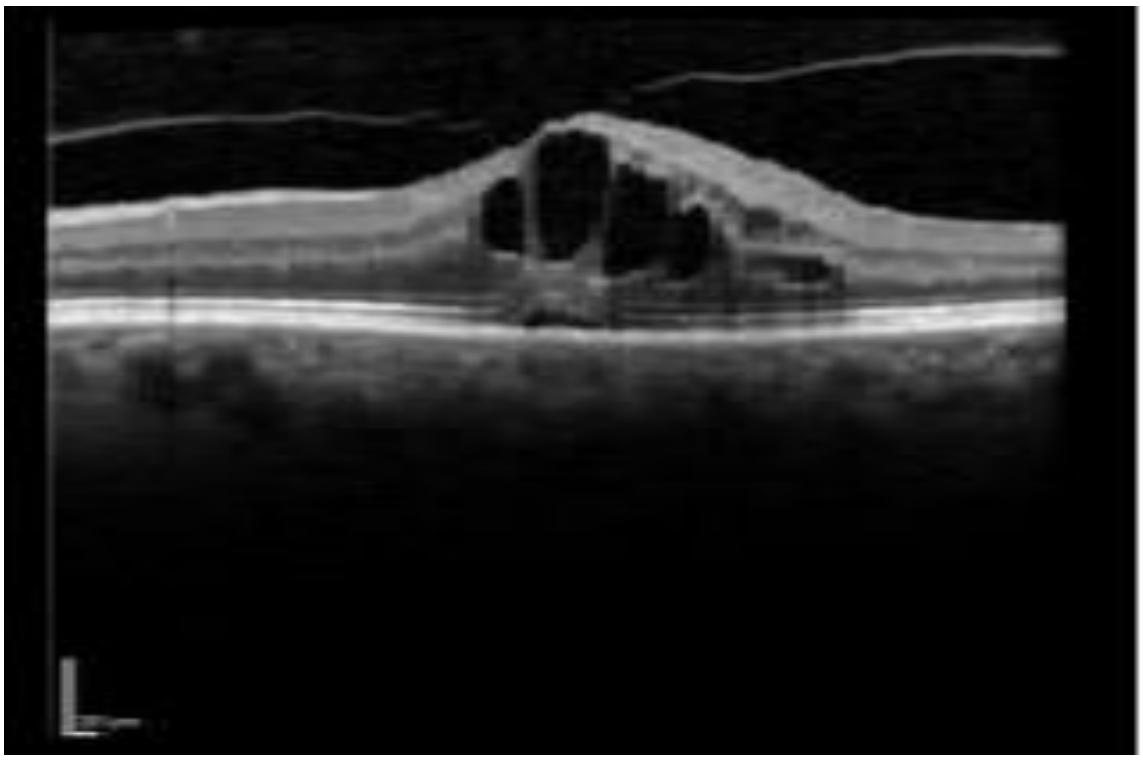


Figure 6.5 : Resized OCT image (128x64 pixels).

-1	0	1

-1
0
1

Figure 6.6: Kernel for filtering an image.

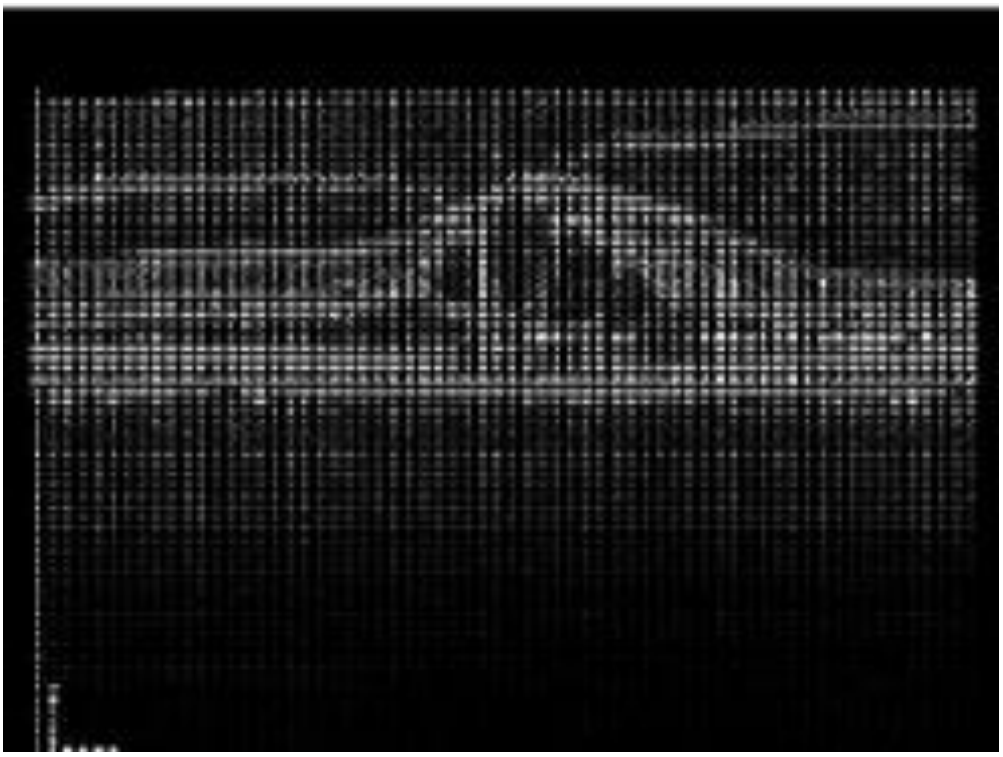


Figure 6.7: HOG features descriptor on an image.

6.1.4.2 Local Binary Patterns (LBPs)

Another type of feature extraction is called the Local Binary Pattern (LBP) method, which is type of visual descriptor used in computer vision. It was originally designed for the description of texture. The fundamental concept in developing the LBP operator is that two-dimensional surface textures could be represented by two complementary measures: local spatial patterns and greyscale contrast. The original LBP operator [152] forms labels for the image pixels by thresholding the 3 x 3 neighbourhood of each pixel with the centre value and considering the result as a binary number. The histogram of these 28 = 256 different labels can then be employed as a texture descriptor. This operator when utilised jointly with a simple local contrast measure delivered good performance in unsupervised texture segmentation [153]. In its simplest form, the LBP vector feature vector is created as follows. The image is divided into patches (such as with 16x16 pixels for each patch). For each pixel in a patch, the pixel is compared to each of its 8 neighbours and if the centre pixel's value is smaller than or equal to the neighbour's value, it is written as '1' or if not as '0'. This gives

a result which is an 8-digit binary number. Then, the histogram over the cell is computed, and this histogram becomes a 256- dimensional feature vector, normalizing the histogram. The final process is to concatenate the histograms of all patches so that the occurrences of the LBP codes in an image are collected into a single histogram, as shown in Figure 6.8 for an illustration of the computation of the LBP descriptor. An example of applying LBP on OCT image is shown in Figure 6.9. This introduces a feature vector for the entire image which can then be utilised in regression and classification. The LBP operator was later extended to employ neighbourhoods of different sizes [154]. Here a local neighbourhood is described as a group of sampling points spaced on a circle centred at the pixel to be labelled, which permits any radius and number of sampling points. Bilinear interpolation is utilised when a sampling point does not fall in the centre of a pixel. The notation (P, R) is used for pixel neighbourhoods, which means P sampling points on a circle of radius of R. To obtain the LBP code using equation 6.5.

$$LBP = \sum_{P=0}^{P-1} s(g_p - g_c)^{2p} \qquad (6.5)$$

Where

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where g_c is the intensities of the central pixel, g_p is a given neighbouring pixel, and P is the number of sampling points in the circle of radius R. Figure 6.10 shows several sizes of LBP resolutions with corresponding P and R.

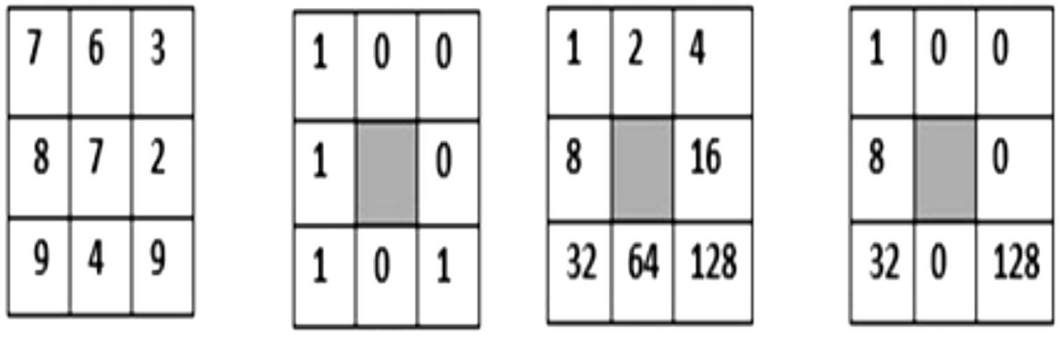


Figure 6.8: Computation of the local binary pattern (LBP).

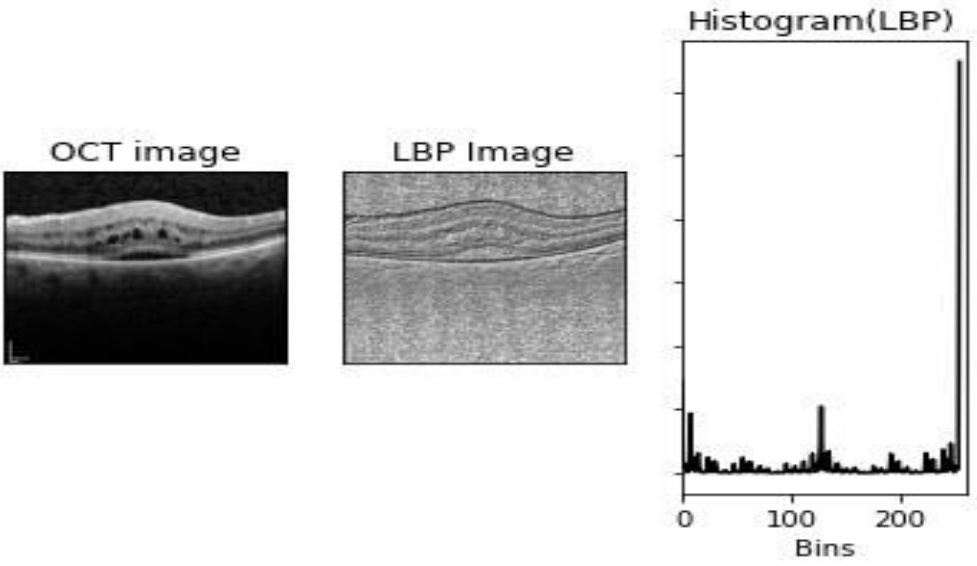


Figure 6.9:Local Binary Pattern Histogram.

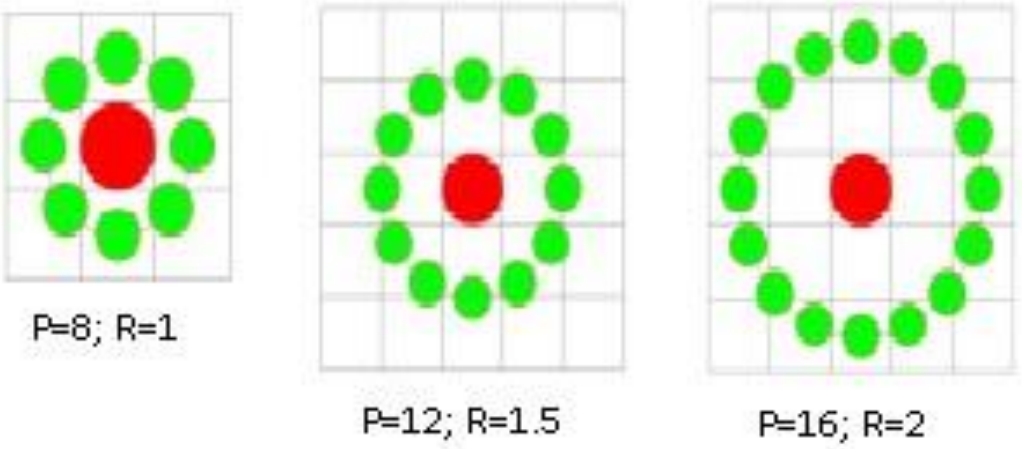


Figure 6.10: Radius R with corresponding neighbourhood pixels P.

6.2 Experiments and Results

The five machine learning approaches were evaluated against the linear regression algorithm for their performance in predicting VA after one year. The two-feature extraction methods LBP and HOG are employed in the analysis of OCT images and fused to improve the performance of the system. The evaluation is conducted by combining the EMRs and LBP and HOG features. 194 samples were used for training and 50 samples for testing. Values of R-squared (R2) were used to assess the method's performance. R2 is calculated as shown in equation 6.6 below:

$$R^2 = 1 - \frac{\sum(Y(\text{actual}) - Y(predicted))^2}{\sum(\text{Y(actual)} - Y(mean))^2} \quad (6.6)$$

6.2.1 Analysis of Machine Learning Techniques for Prediction using the LBP Method

To obtain the optimal parameter setting for LBP, various samples of the neighbourhood (P) and radius of the circle R are considered. After the pre-processing stage, values of P equivalent to 4,6, and 8 and of radius R equivalent to 2 and 4 have been investigated in this experiment. Table 6.2 illustrates the machine learning techniques considered in this study, and that merging EMR and LPB (P=4, R=2) gained higher accuracy where R^2 = 74% (0.74) was achieved using the Random Forest algorithm. It can be observed from the whole table that machine learning techniques outperformed traditional linear regression. This is true for all different samples of neighbourhood and radius and, in particular, the performance of the Random Forest method was superior among the four machine learning techniques used.

Model	EMR + LPB P = 8, R = 2	EMR + LPB P = 8, R = 4	EMR + LPB P = 6, R = 2	EMR + LPB P = 6, R = 4	EMR + LPB P = 4, R = 2	EMR + LPB P = 4, R = 4
Linear Regression	0.6883	0.7187	0.6614	0.6923	0.6947	0.6976
Ridge Regression	0.7206	0.7208	0.7206	0.7207	0.7205	0.7207
Neural network (MLP Regressor)	0.7218	0.7226	0.7217	0.7222	0.7217	0.7219
PLS Regression	0.6967	0.7189	0.6794	0.6997	0.6982	0.7006
LASSO Regression	0.7068	0.7068	0.7068	0.7068	0.7068	0.7068
Random Forest Regressor	0.7344	0.7349	0.7375	0.7378	0.7414	0.7374

Table 6.2: Illustration of the accuracy of linear regression and 5 ML techniques fusing EMR and LBP with different samples of various values of neighbourhood (P) and radius of the circle (R).

6.2.2 Analysis of Machine Learning Techniques for Prediction using the HOG Method

As shown in Table 6.3, several HOG descriptors have been considered in examining the performance of linear regression and machine learning techniques. As can be seen, the Random Forest regressor provided the best performance in both methods and its highest score was R2 = 0.7514 when using the HOG descriptor with 8 features. Again, machine learning techniques showed superior performance compared to traditional regression, and the Random Forest regression demonstrated higher accuracy than the other machine learning algorithms.

Model	EMR +HOG descriptor with 8 features	EMR +HOG descriptor with 12 features	EMR +HOG descriptor with 16 features	EMR +HOG descriptor with 20 features
Linear Regression	0.6996	0.7078	0.6174	0.6829
Ridge Regression	0.7193	0.7193	0.7193	0.7190
Neural network (MLP Regressor)	0.7200	0.7205	0.7122	0.7095
PLS Regression	0.6996	0.7088	0.6336	0.6865
LASSO Regression	0.7068	0.7068	0.7068	0.7068
Random Forest Regressor	0.7514	0.7495	0.7228	0.7057

Table 6.3 : Performance of linear regression and 5 ML techniques using various HOG descriptors fused with EMR.

6.3 Summary

The work described in this chapter investigated several machine learning algorithms and linear regression to predict VA after one year of treatment using baseline EMR and OCT images. Two feature extraction methods have been utilised in the analysis of OCT images, fused with the EMR dataset to improve the performance of the method. A number of experiments have been conducted to evaluate the performance of learning regression for the prediction of the variability seen in visual acuity 12 months after treatment. The machine learning techniques studied demonstrated superior ability than multivariate linear regression to predict visual prognoses for patients with RVO. The Random Forest algorithm achieved the best performance using EMR fused with the HOG descriptor. Such algorithm could allow a consulting physician to offer a personalised prediction of treatment outcomes in order to support patient-centred decision-making in routine practice. Additionally, the need to benchmark against clinical performance to understand the threshold of performance could offer clinical value. The next chapter aims to increase the system's performance by pre-processing OCT images, adding a feature extraction method called GLCM to previous methods (LBP and HOG) to analyse the images, fusing with EMR and feeding them to the CNN model. The final step is evaluating the outcomes against ophthalmologists' performance.

Chapter 7 PREDICTING THE VISUAL ACUITY OF RVO PATIENTS AFTER 12 MONTHS OF TREATMENT USING AN LBP-HOG-GLCM HYBRID AND CNN

In previous chapters machine learning models and linear regression have been addressed to predict VA after 12 months of ant-VEGF injection. EMR data has been used and two feature extraction methods for OCT image analysis have been investigated. From the results, it is proved that the machine learning approaches examined presented a superior ability than linear regression to forecast visual prognoses for RVO patients. For improving the predicting system's performance this chapter introduce an approach used CNN model and compare the model performance against clinical performance.

In the present research, a novel solution is proposed to provide RVO case-specific VA forecasts to support ophthalmologists in making better treatment decisions. Firstly, relevant features are extracted from OCT scans in two different ways to represent the disease region more efficiently. Then, the extracted features and corresponding parts of EMR are combined to feed into CNN model to train the predictor. Finally, the VA of RVO patients after one year of treatment is predicted by the trained model. The major contributions of this experimental study are as follows:

1. Based on HOG, LBP and the grey level co-occurrence matrix (GLCM), a hybrid vision feature extraction method is proposed which shows better feature representation ability in capturing VA information.

2. The vision features extracted from OCTs are fused with their corresponding EMR data to generate the new hybrid features. A CNN model is trained based on the hybrid features to predict the VA of RVO patients after 12 months of anti-VEGF treatment.

3. The prediction performance of the proposed method is then compared with VA forecasts from 11 different ophthalmology specialty registrars, and 1 specialist retina consultant.

Figure 7.1 shows the pipeline of the proposed method and the entire procedure of treatment prediction. The pipeline comprises of four stages of pre-processing, feature extraction and fusion, regression model training, and model evaluation. Next sections will explain these stages in mor details.

The rest of this chapter is structured as follows. Section 7.2 illustrates in details image processing methods, extraction method, and CNN used for this experiment. Section 7.3

describes the proposed novel system in detail. Evaluations and results are then provided in Section 7.4 and conclusions are drawn in Section 7.5.

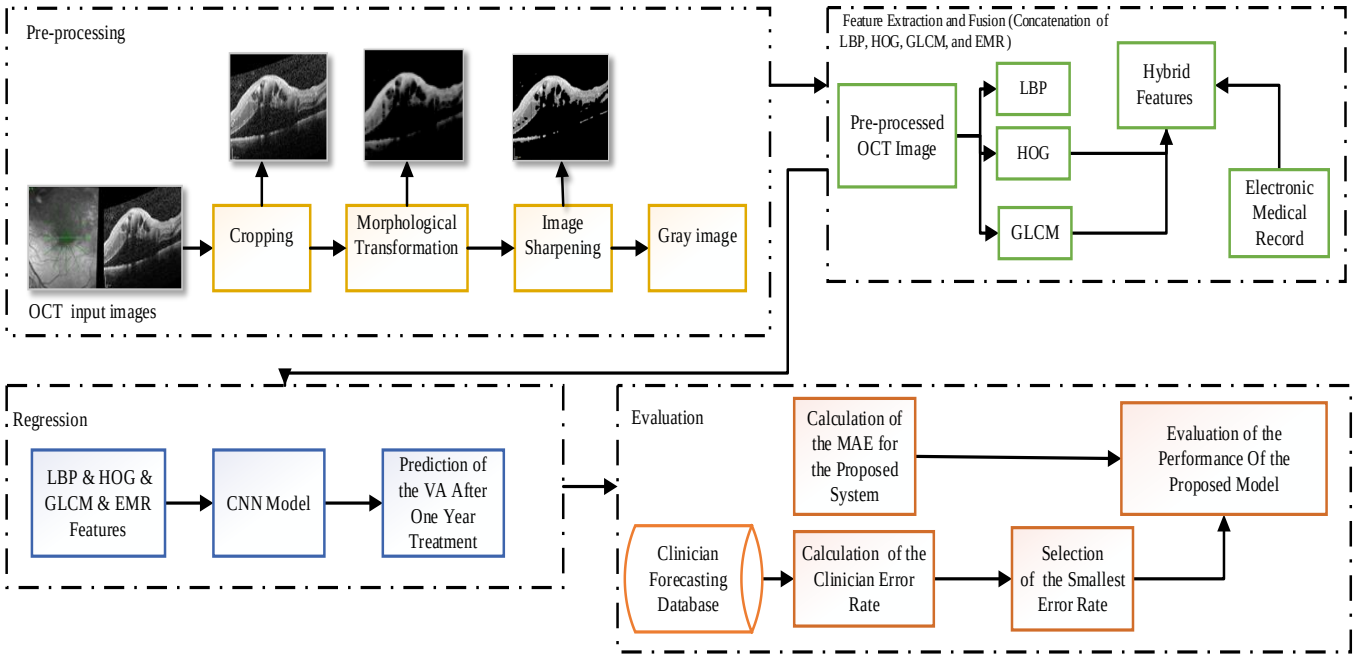


Figure 7.1:Pipeline of the proposed method and the procedure of treatment prediction, comprising four stages of pre-processing shown in different colours: feature extraction and fusion, regression model training, and model evaluation.

7.1 Image processing Methods

In order to enhance OCT images and to obtain extra some useful information from them, image processing operations has been performed to improve the system's performance. Morphological Transformation Technique, sharpening technique and Gaussian Blurring have been investigated for OCT images analysis.

7.1.1 Morphological Transformation Technique

Morphology transformation is a technique of image processing based on the form and shape of objects. The technique applies a structuring element to an input image. Each pixel's value is based on a comparison of the corresponding pixel in the input image with those of its neighbours. The construction of a morphological operation that is sensitive to specific shapes

in the input image can be applied by choosing the shape and the size of the neighbour. Morphological transformation operations such as opening, dilation, erosion, and closing are based on the image shape. The concepts of the morphological profile are then used to create a feature vector from a single image. Combinations of these operations are usually utilised in performing morphological image analysis [155, 156]. The morphological profile and the derivative of the morphological profile (DMP) are used to generate a feature vector from a given image. These concepts are based on the repeated use of the closing and opening operators which are very commonly utilised in mathematical morphology [157].

1. Dilation

The main idea behind dilation is that it adds pixels to the boundaries of images' objects. The number of pixels added relies on the shape and the size of the structuring element. As the structuring element is scanned over the image, the maximal pixels value overlapped by the structuring element is computed and replaced with the image pixel in the centre of that structuring element. This operation causes bright regions within the image to grow. This transformation produces an image with the same shape as the original but of a different size. The dilation of *A* by the structuring element *B* is defined in equation (7.1)

$$A \oplus B = \left\{ z \mid \left[(\hat{B})_Z \cap A \right] \in A \right\} \tag{7.1}$$

The equation illustrates that when the A dilated by B, the outcome element z could be that there will be at least one element in B that intersects with an element in A.

The structuring element B is positioned with its origin (x, y), the following rule is used to determine the new pixel:

$$g(x,y) = \begin{cases} 1 & \text{if } B \text{ hits } A \\ 0 & \text{otherwise} \end{cases}$$

Figure 7.2 demonstrates the dilation on the grayscale image. The dilation function uses the maximum value of all pixels in the neighbourhood as the value of the corresponding pixel in the output image. In the case of a binary image, if any pixels are set to 1, the output pixel is set to 1[158]. Figure 7.3 shows applying morphological dilation method to an OCT image.

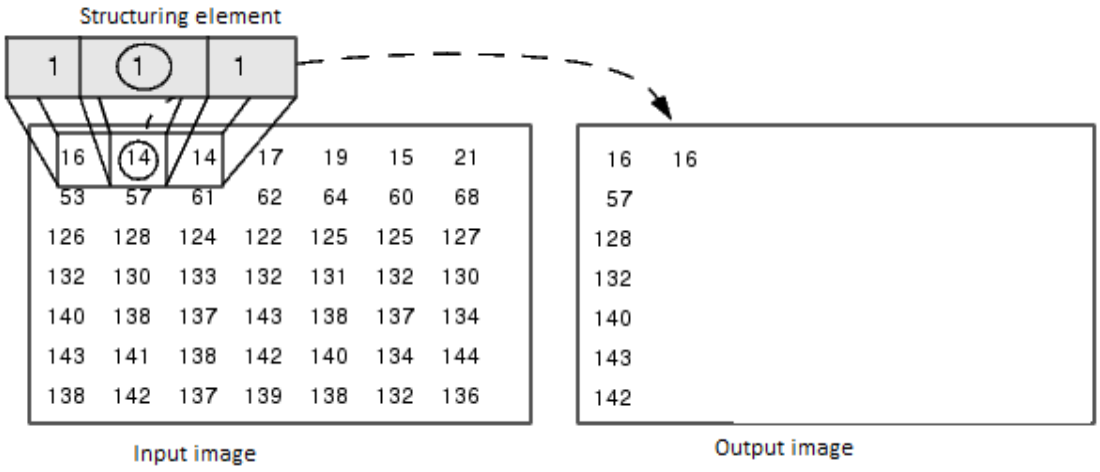


Figure 7.2: Dilation process of Grayscale Images[158].

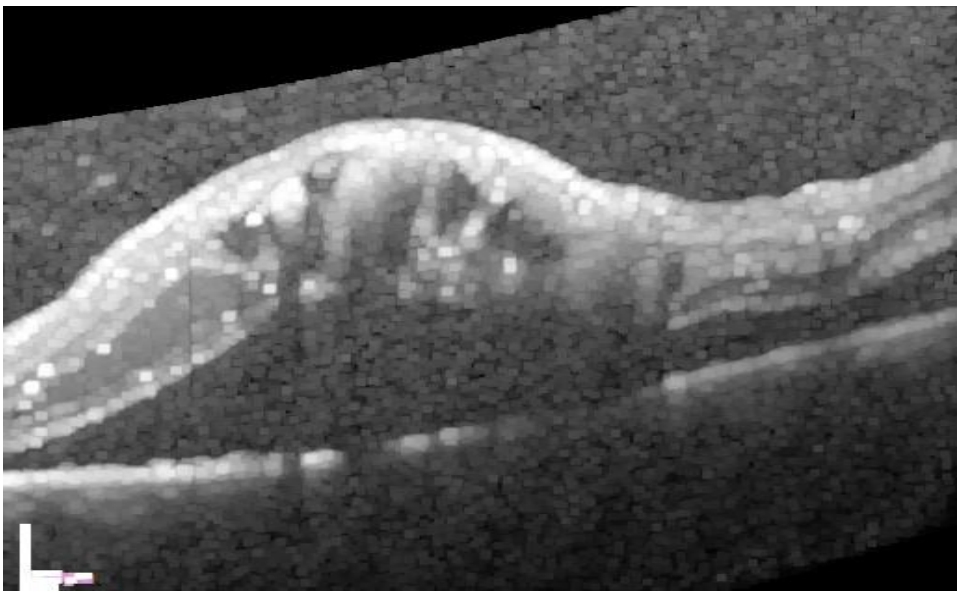


Figure 7.3:Application of the dilation method to an OCT image (5,5) kernel.

2. Erosion

This operation is the exact opposite of dilation. It erodes the boundaries of a foreground object. This means that it reduces the sizes of objects in the image. In this operation, in a grayscale image, the value of the corresponding pixel in the output image is the minimum value in the input pixel's neighbourhood as illustrated in Figure 7.4. In a binary image, the output pixel is set to 0 if any of the pixels are set to 0. Erosion is defined by the equation (7.2)[158]:

$$A \ominus B = \{z \mid (\hat{B})_Z \in A\} \tag{7.2}$$

The equation indicates that the erosion of image A by the structuring element B, shifted by z, represents a subset of A that is B is entirely contained within A. This process decreases the number of pixels from object boundary. Figure 7.5 shows applying erosion method to an OCT image.

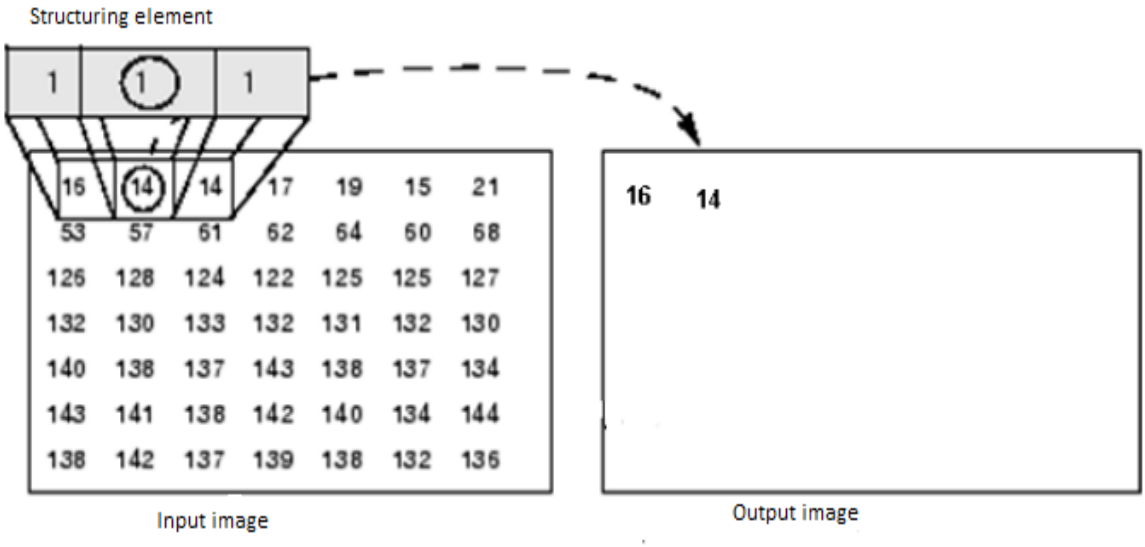


Figure 7.4: Erosion process of a Grayscale Image[158].

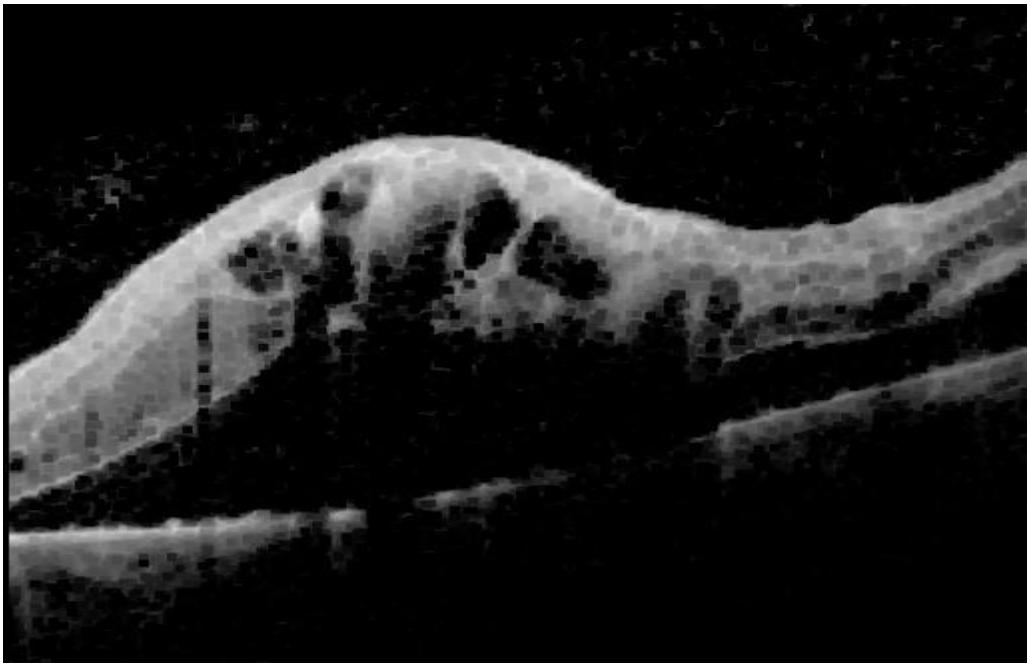


Figure 7.5:Application of the erosion method to an OCT image (5,5) kernel.

3. Opening

This operation is simply another name for erosion followed by dilation as explained above. This operation has been applied with the OCT images in the present study as the system achieved better performance. Figure 7.6 shows applying opening method to an OCT image. Opening process of image, A by B can be obtained by the erosion of A by B, followed by the dilation process as illustrated in equation 7.3[158]:

$$A \circ B = (A \ominus B) \oplus B \qquad (7.3)$$

Figure 7.6:Application of the opening method to an OCT image (5,5) kernel.

4.Closing

This transformation is simply the opposite of the opening operation in that it is dilation followed by erosion. Applying closing function on an OCT image is showed in figure 7.7. It is useful to close a small hole in the object. followed by erosion of the resulting structure by B as shown in equation 7.4 [158]:

$$A \cdot B = (A \oplus B) \ominus B \qquad (7.4)$$

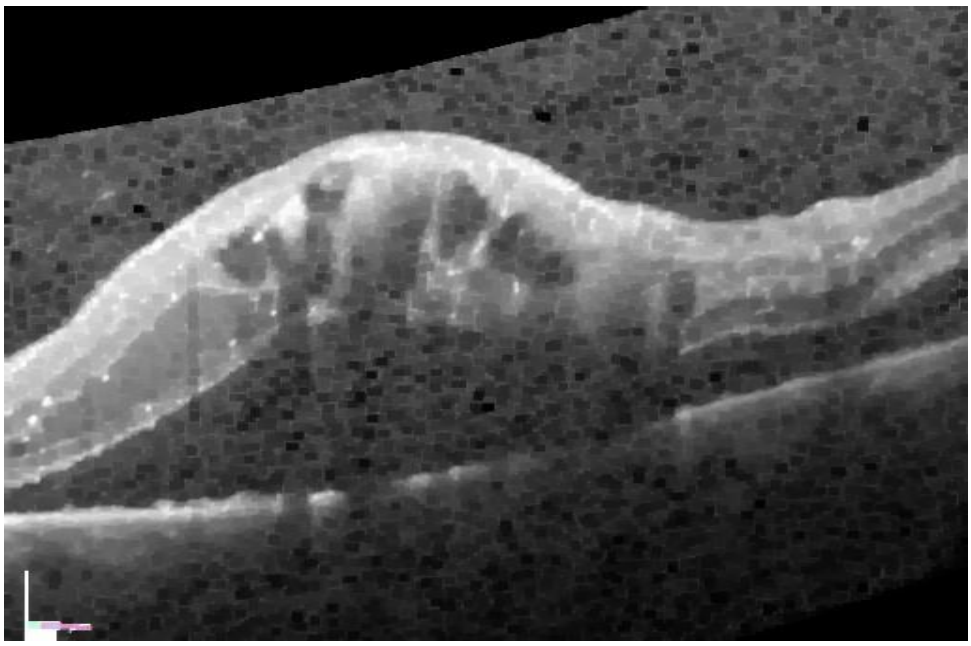


Figure 7.7:Application of the closing method to an OCT image (5,5) kernel.

7.1.2 Sharpening Technique and Gaussian Blurring

Prominent techniques for image enhancement are commonly separated into pixel domain and frequency domain methods. In the former, the grey value of every pixel is altered based on statistical information about the digital image while, in the latter, high-frequency components of the image are divided from low-frequency content in two distinct signals. Each signal is processed independently and then combined to reconstruct the sharpened signal. The basic idea is to add a fraction of the high pass filtered version of the input image to the original image itself. The unsharp masking technique belongs to the second group and is one of the most widely utilised image enhancement processing tools improving the sharpness of images. The operation of this tool is divided into the following steps. Firstly, a low-pass filter such as a Gaussian blur filter is applied, and the resulting filtered image is called the base layer that contains the main structure of the original image. Alternatively, a high-pass filter such as the Laplacian filter is employed to extract the detail layer. The subtraction of the base layer from the original image results in the detail layer that contains the fine details of the original image. Next, the detail layer is amplified by multiplication with a scaling factor λ and then the result is added back to the original image to generate an enhanced version [159], and this process can be defined by the following equation 7.5:

$$F = I + \lambda(I - B) \tag{7.5}$$

where λ > 0, I is the original image, B is the filtered image, and F is the final image which represents the sharpened or enhanced version of the image. Figure 7.8 illustrates the sharpening process implemented for an image of 'Lena'.

The low-pass filter that has been used for sharpening OCT images in the present study is the Gaussian blur (or Gaussian smoothing). This is a linear shift-invariant low-pass filter which assigns various weights to the pixels inside each kernel area and replaces the central element with the Gaussian weighted sum. In each kernel, the central pixels tend to have higher weighting, which is reduced in both the x and y directions by keeping close to the kernel's edge. This technique is used as a pre-processing phase to enhance image structures and reduce the image noise and details. It is a type of image blurring filter which uses a Gaussian function, the formula of which in one dimension is:

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \tag{7.6}$$

and in two dimensions is:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{\frac{x^2+y^2}{2\sigma^2}} \tag{7.7}$$

where *x* is the distance on the horizontal axis from the origin and *y* is the distance in the vertical axis from the origin, where the origins of these axes are at the centre (0,0); and σ represents the standard deviation of the Gaussian distribution assuming that the distribution mean = 0. Figures 7.9 and 7.10 show one- and two-dimensional Gaussian distributions with a mean of 0 and σ =1 and a mean of (0,0) and σ=1 respectively. Figure 7.11 illustrates applying the gaussian smoothing filters to an image by specifying a scalar value for sigma.

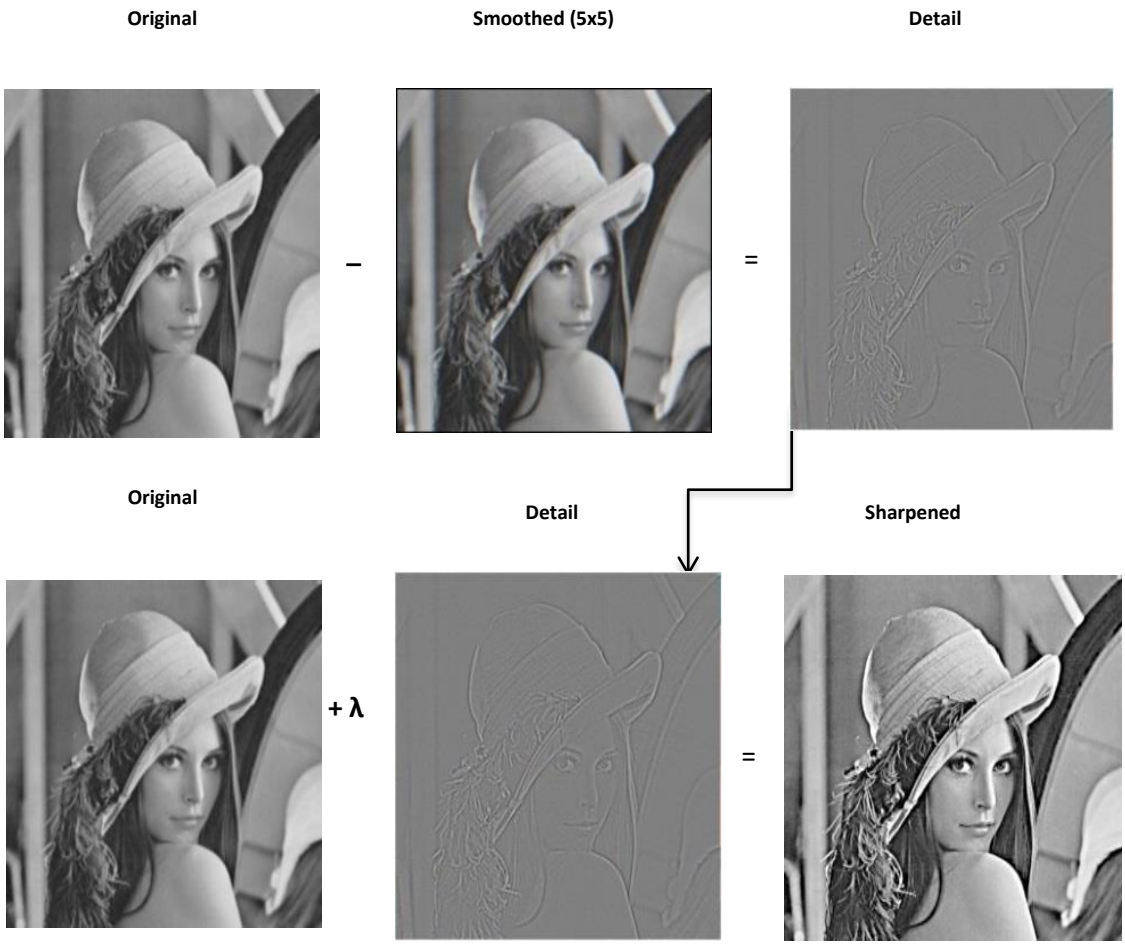


Figure 7.8:Sharpening process implemented for image of 'Lena'.

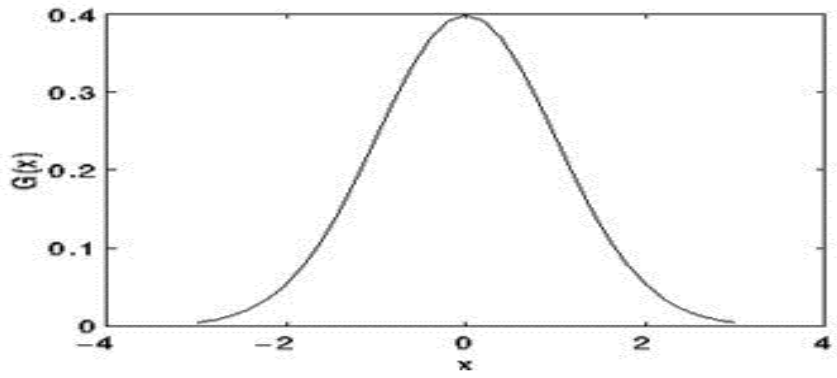


Figure 7.9:One-dimensional Gaussian distribution with mean 0 and σ=1.

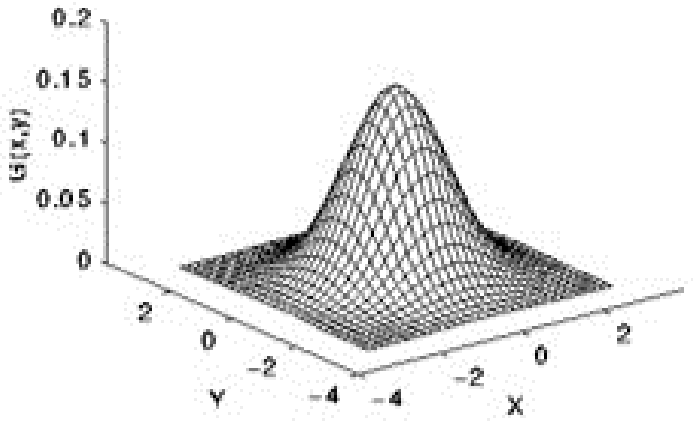


Figure 7.10:Two-dimensional Gaussian distribution with mean (0,0) and σ=1.

Original image

Smoothed image, $\sigma = 2$

Smoothed image, $\sigma = 4$

Smoothed image, $\sigma = 8$

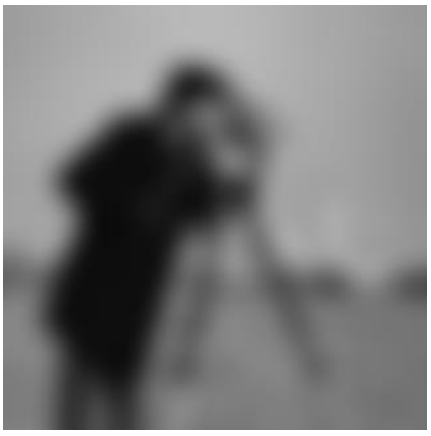


Figure 7.11:Application of Gaussian smoothing filters to an image by specifying a scalar value for sigma.

7.2 Grey Level Co-Occurrence Matrix (GLCM) Feature Extraction Method

The GLCM is considered to be one of the most coherent techniques used for textural feature analysis and was proposed by Haralick et al. [160]. It is a statistical method used to examine texture that investigates the spatial relationships among pixels. A corresponding GLCM for a given image is calculated by:

$$P(i,j,d,\theta) = \sum_{x=0}^{n}\sum_{y=0}^{m} \begin{cases} 1, \text{if } I(x,y) = i \text{ and } I(x + d\cos\ \theta, y + d\sin\ \theta) = j \\ 0, \text{otherwise} \end{cases} \quad (7.8)$$

Where $P(i,j,d,\theta)$ illustrates the relative frequencies with which two pixels are separated by a particular distance d and a specified angle θ occurring on the image, one with grey level i and the other with grey level j. This approach involves the two major parameters of direction and distance. The GLCM is applied by the selection of pairs of distance and angle. It computes how often pairs of pixels occur with a specific value and offset in the image in order to find its GLCM matrix. Figure 7.12 explains how the GLCM is calculated from a grey comatrix of a 4-by-5 image for d = 1 and θ=0°. In the method proposed in the present study, various combinations of directional angles and distances were used for the calculation of the GLCM by recognizing particular patterns based on their textural features. The GLCM is used with distance d, which is equal to 1, 2,3,5,6, while θ is measured in four directions of 0°, 45°, 90° and 135°.This process will give seven GLCM matrices. Later, to estimate the similarity between the various GLCM matrices, five of the statistical features (constant, homogeneity, energy, correlation, and dissimilarity) were selected and features and formulae are briefly described in Table 7.1 [161]. Here, G represents the number of the grey levels in the frame, and $P_d^{(i,j)}$ indicates the (i,j)th entry in the GLCM which symbolises the probability of specific pixel pairs at specific distances and angles.

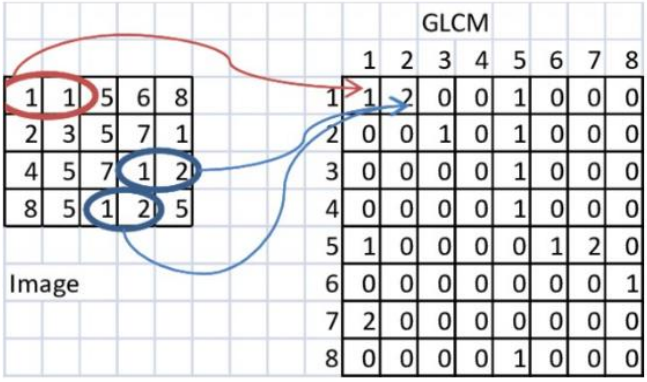


Figure 7.12:GLCM calculation from a 4-by-5 image [162].

Feature description	Formula
Contrast: To compute the intensity contrast between a pixel and its neighbour over the image. A high amount of variation, the contrast will be high.	$\sum_{i=0}^{G-1}\sum_{i=0}^{G-1}\lvert i-j\rvert^2 p_d^{\theta}(i,j)$
Homogeneity: Calculate the similarity in the image. A value of 0 indicates a high similarity in the image.	$\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\frac{1}{1+(i-j)^2}p_d^0(i,j)$
Dissimilarity: Is similar to Contrast and will be high if the local region has high contrast.	$\sum_{i,j-0}^{N-1} p(i,i) * \lvert(i-j)$
Energy: Can be used to measure the textural uniformity of the image. It provides the sum of squared elements in the GLCM.	$\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\left[p_d^{\theta}(i,j)\right]^2$
Correlation: Measure the combined probability incidence of the certain pixel pairs.	$\sum_{i=0}^{S-1}\sum_{i=0}^{G-1}\frac{ijp_d^{\theta}(i,j)-\mu_x\mu_y}{\sigma_x\sigma_y}$

Table 7.1: GLCM Feature descriptions and formulae

7.3 Convolutional Neural Network (CNN)

Fundamentally, a CNN is a computational model inspired by the characteristics of the human brain's neural system, which comprises numerous interconnected nodes organised in stacked layers to achieve simple operations on input data. Like the traditional neural network, the CNN consists of an input layer, middle layer, and output layer. However, in a traditional neural network, middle layers are described as hidden since their inputs and outputs are masked by the activation function and final convolution (Figure 7.13). In the CNN, the hidden layers include layers that perform convolutions and are the main building block of CNNs which computes the convolution of a set of learnable filters and 3D input data. The vital difference between a convolutional layer and a densely connected layer is that dense layers learn global patterns in their input feature space, whereas convolution layers learn local patterns [163]. The term convolutional neural network indicates that the neural network utilises a mathematical operation called convolution. The convolution kernel corresponds to a neuron, and its size is named the neuron's receptive field. The convolution kernel is a two-dimensional matrix with a size $n \times n$ in which each point has a corresponding weight. In a

convolutional layer, each neuron receives input only from its receptive field in the previous layer. This considerably diminishes the number of free parameters and allows convolution layers to be applied to extensive data such as images [164]. The convolution operation generates a feature map which contributes to the next layer's input. Following the convolution operation, a pooling layer is conducted to reduce computation and avoid overfitting. This layer is independently performed on every depth slice of the input volume. It is a pooling operation that computes the maximum value for patches of a feature map and utilises it to produce a pooled feature map (Figure 7.14) and is often used after a convolutional layer. Next, the generated feature maps are sent to activation functions that perform non-linear transformations. After several convolution and pooling layers, fully connected layers are performed. The final layer is the loss layer to determine how training penalises the error between the actual and predicted values in regression or classification problems. The input is a tensor with a shape: a number of inputs × input height × input width × input channels. In the first calculation process, convolution kernels in the network perform convolution calculations on the image. Here, the pixel values at the corresponding positions of the image and the weight values in the convolutional kernel are added to the receptive field. Next, the convolutional kernels slide to the following location of the image according to the step size and the process is repeated until all of the image's pixels have been counted. This calculation results from a feature map of the original image [165].

In order to improve the fitting ability of the CNN, the non-linear function is applied to map the output characteristic graph of the CNN. This is a ReLU function because this function can simulate the brain environment. It is expressed as in equation 7.9:

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \tag{7.9}$$

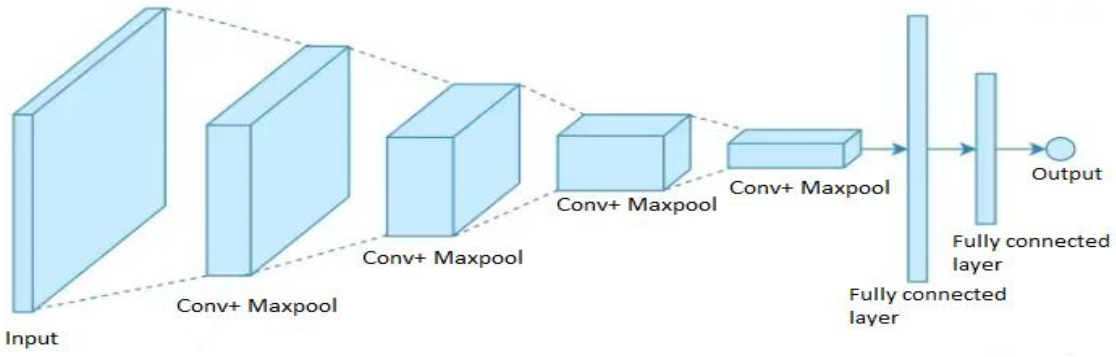


Figure 7.13:Convolutional Neural Network (CNN) Architecture.

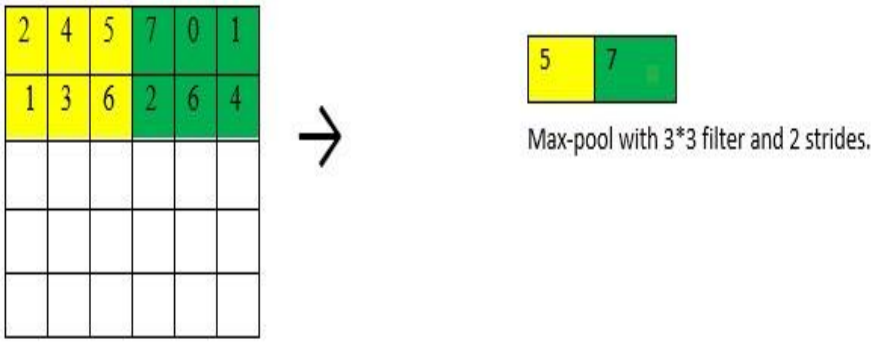


Figure 7.14:Computation of max pooling.

7.4 Proposed Method

A CNN-based method has been designed to predict VA after 12-months of anti-VEGF treatment, and the proposed method is explained in this section. The pipeline comprises of three major steps. Firstly, the pre-processing of OCT scans is required. Then, based on the pre-processed scans, feature extraction and fusion operations are applied, after which the CNN model is trained using the training data and target. Details of each step are provided in the following sections.

7.4.1 Pre-processing

7.4.1.1 Cropped Images

In the collected OCT scans, each image has two parts: an en-face and a cross-sectional image. Each OCT scan set contains either 6 or 25 different cross-sectional images arranged radially or orthogonally and centred on the fovea. The images are labelled from 0-5 or 0-24 as shown in Figures 7.15 and 7.16 respectively. These two styles represent a transition in OCT imaging protocol which occurred over the period of time in which data were recorded at the RVI. We selected images with the numbers 3 or 15 on each scale to include a fovea-centred cross-section and cropped these images to obtain the region of interest where only images with retina parts remained, as shown in Figure 7.17.

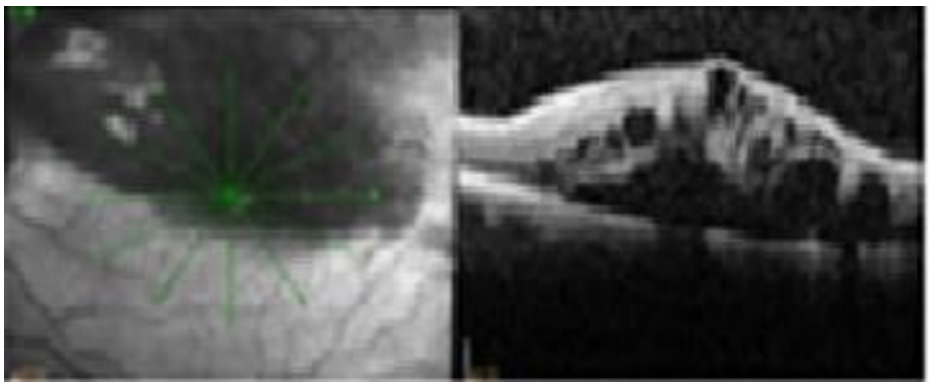


Figure 7.15: OCT scans (1008×496 pixels) including: (a) the scale of 6 different cross-sectional images; and (b) one example of an en-face OCT image. The OCT image with scale number 3 has been selected as this slice maps onto the patient's central vision and so is likely to hold much of the visual information of prognostic value represented by the OCT imaging as a whole.

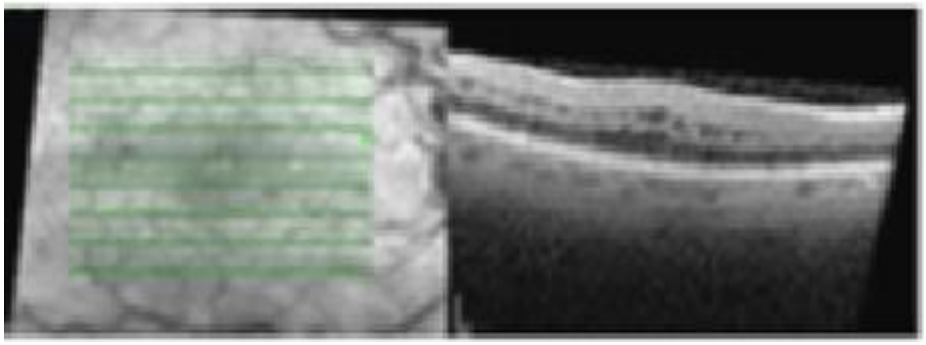


Figure 7.16: OCT image: (a) scale of 25 different cross-sectional images (1008×496 pixels); and (b) an example of an en-face OCT image. The OCT image with scale number 15 has been selected as this slice maps onto the patient's central vision and so is likely to hold much of the visual information of prognostic value represented by the OCT imaging as a whole.

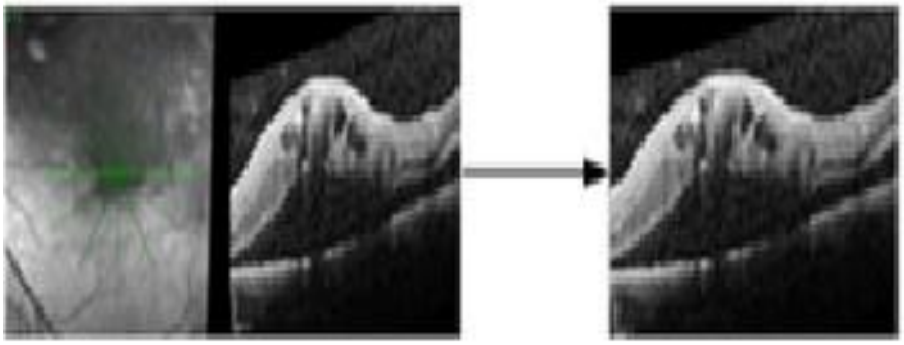
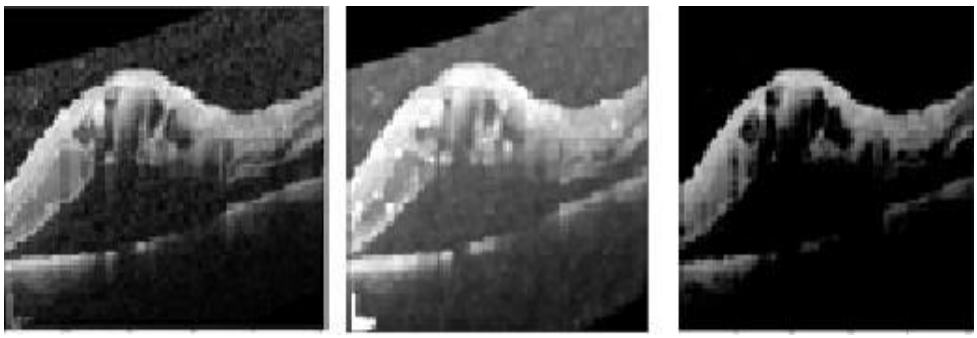


Figure 7.17:Original OCT scan 1008×496 pixels in size and cropped to select the region of interest resized to 513×513 pixels.

7.4.1.2 Morphological Transformation

The enhancement of medical images is considered to be a significant step in achieving better object recognition. Enhancement improves the visual perception of unclear images, enabling better diagnosis and treatment[166]. Enhancement techniques utilise mathematical principles to extract the features of an image that are useful in describing the shapes of zones, applying a structuring element to an input image and creating an output image at the same size [167]. Each pixel's value is based on an evaluation of the corresponding pixel in the input image with its neighbours. Morphological operators have two sets of input. The first set represents the original image which is binary or in grey level, while the second input contains the structural element (mask) which is a matrix containing binary values. The main morphological transformations in the pre-processing steps are erosion, dilation, opening and closing. The concepts of the morphological profile are used to create a feature vector from a single image. Combinations of these operations are usually utilised to perform morphological image analysis [155, 156]. In the method proposed here, the opening and closing operations are used to achieve the morphological transformation required. The opening and closing operations comprise both erosion and dilation steps. Firstly, erosion is applied to thin the boundaries of the foreground object; hence, it reduces the object's size in the image by reducing the peaks and enlarging the widths of minimum regions. Positive noise is removed but impulsive noise is only slightly affected. However, the erosion operation removes noise but at the same time shrinks the object. Therefore, a dilation operation is utilised to resize the shape to ensure that the sizes of the raw OCT and pre-processed scans are identical. This transformation is beneficial to join up broken parts in an object. A closing operation is applied to the OCT images, which involves simple dilation followed by erosion, to remove minor holes, fill cracks, and fully preserve the shape (Figure 7.18). By examining the results of all of the above operations, opening and closing operations contribute to improving the results of the method.

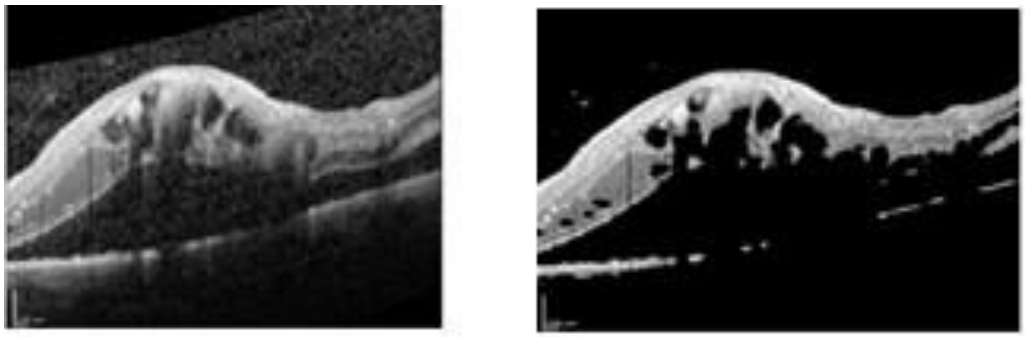


(a) (b) (c)

Figure 7.18: Application of the closing operation, involving dilation followed by erosion operations with the structuring element using an elliptically shaped kernel: (a) original OCT image; (b) image after dilation, with increases in both the white region and the size of the foregrounded object; (c) image after the subsequent erosion process which removes the white noise and shrinks the interesting part of the image.

7.4.1.3 Sharpened Image and Grey Image

To enhance the quality of the images and to reduce the noise in the OCT scans, a sharpening operation is applied. This process involves firstly applying a Gaussian filter to smooth any sharp edges in the image while minimising blurring. Here, most of the high frequency components are subtracted and the smoothed image from the original image is retained. The output of this process called 'mask'. This mask is then enhanced and added to the original image as shown in Figure 7.19. If Figure 7.19 (a) and (b) are compared, the image is seen to be enhanced by sharpening its edges, thus overcoming some of the blurring in the original image. To speed up and reduce the complexity of the code used, the OCT images were converted to grey scale.



(a) (b)

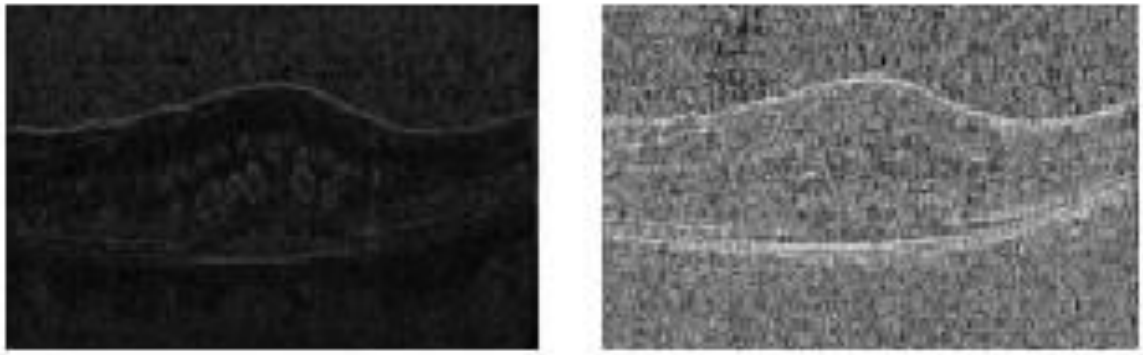
Figure 7.19: Sharpening: (a) original OCT image; (b) the corresponding sharpened image.

7.4.2 Feature Extraction and Fusion

Several methods exist for the extraction of the most useful features from OCT images. In the proposed method, three techniques are utilised to extract pattern feature information from the OCT scans. The first one is called the local binary pattern (LBP) method. This approach was introduced by Ojala et al [168] to describe the local texture patterns in an image. In its simplest form, the LBP feature vector is created by dividing the image into patches (such as 8x8 pixels for each patch), and each pixel in a patch is compared with its neighbours, as described in detail in chapter 6. The second method is the histogram of oriented gradients (HOG), which is a feature descriptor used for object detection in computer vision and image processing. It is a shape descriptor which chooses the local object's appearance and shape. The idea behind this technique is to calculate occurrences of gradient orientation in localised portions of an image. HOG was first presented by Dalal and Triggs in the context of the detection of the presence of a human in images [151] (also described in detail in chapter 6). The third technique is GLCM (described above). In order to better describe the local and overall features of the image, a fusion of LBP, HOG, and GLCM method is investigated. To fuse the HOG, LBP, and GLCM features, the pre-processed image which is 513×513 in size is fed to the HOG algorithm, and Figure 7.20 shows the results of the magnitude and angle calculations. Their matrices are then obtained and divided into 8×8 cells to form a block. Nine points of the histogram are calculated for each block and, after computing histograms for all blocks, four blocks from the nine-point histogram were joined together to form a new block. The 9-point histograms for each cell have been concatenated to form a 9-feature vector. For each block, a 9-point feature vector was collected, and in both the vertical and horizontal directions there are 64 blocks. So, the total length of the HOG features are 64×64×9= 36,864 dimensions. Secondly, the uniform pattern LBP is applied to calculate the histogram of each block and, taking into consideration the 20 neighbours on a circle with a radius of 8, the resulting histograms have 22 dimensions. Thirdly, by applying five statistical features of the GLCM (correlation, constant, homogeneity, energy, and dissimilarity) on the 7 GLCM matrices obtained using the various distances and angles described in Table 7.2, the table illustrates different parameters of angles and distances used in this work. This process generates 35 GLCM features. Later, the block HOG's detailed features, LBP texture features, and GLCM features are joined together to enhance the performance of feature extraction. Finally, these features are fused with EMR data, which contributes 6 features, and hence the total is 36864+22+35+6=36927 features. Fusion in the order of (EMR, LBP,

HOG, and GLCM) gives better performance. This step generates hybrid features for the 330 patients which are then fed to the CNN for the regression of VA after 12 months of treatment. Figures 7.21, 7.22 show the visualisation of the HOG, LBP. Table 7.2 describes the parameter settings for the three techniques.

A key feature of CNNs is their ability to learn and extract relevant features automatically from row data by applying convolutional filters to the input image. These filters detect patterns and features at different spatial locations in the input image and transform them into a new representation, which is then used as input to the next layer. However, there are some scenarios where we may want to use pre-extracted features as input to CNNs. One common reason to use pre-extracted features is when we have a limited amount of training data available. Training a CNN from scratch requires a large amount of labelled data to learn the relevant features and weights of the network. In some cases, we may not have enough data to train a CNN from scratch. Another reason to use pre-extracted features is to improve the efficiency of the training process. Training a CNN from scratch can be computationally expensive and time-consuming, especially for large datasets or complex architectures. By using pre-extracted features as input, we can reduce the number of parameters that need to be learned during the training process, making the training faster and more efficient[169, 170].In this method as the amount of medical data that has been provided is limited, pre-extracted features are used to improve the efficiency of the training process.

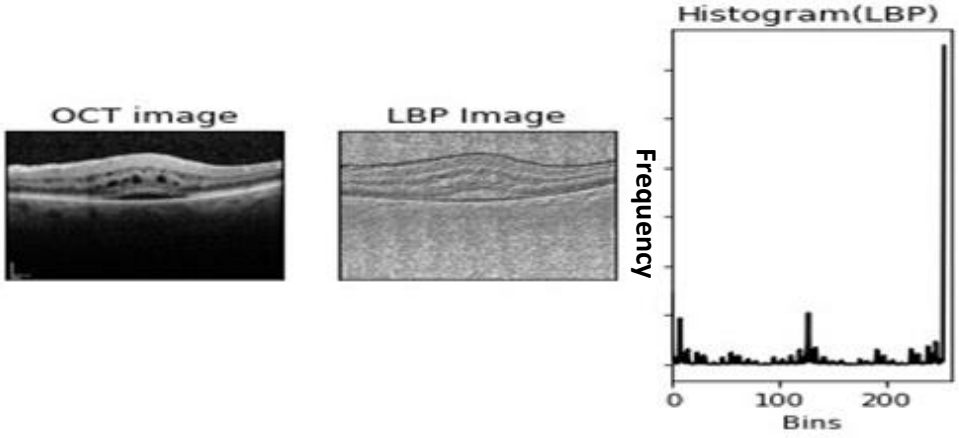


(a) (b)

Figure 7.20: Visualisation for one example of an OCT image: (a) magnitude of the image; (b) angle of the image.

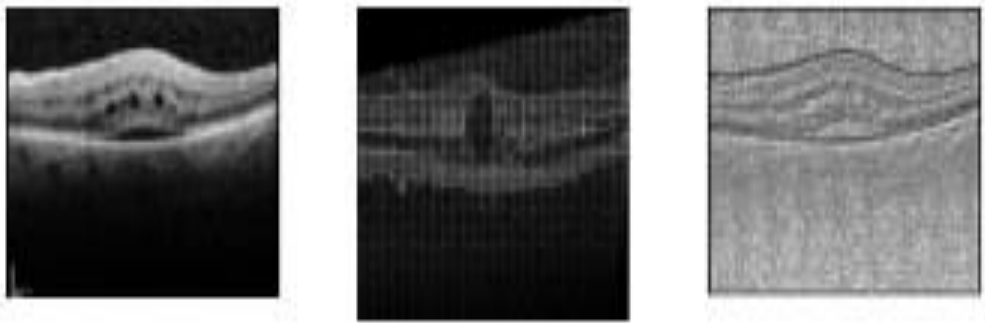
GLCM		LBP		HOG	
Distance	Angle	Parameter	Value	Parameter	Value
1	90°	Number of Points	20	Number of orientations	9
1	135°				
2	0°	Radius	8	Pixels per cell	(8,8)
3	45°				
4	0°	Method	Uniform	Cells per block	(1,1)
5	90°				
6	0°				

Table 7.2: GLCM, LBP and HOG parameter settings.



(a) (b) (C)

Figure 7.21: (a) Input OCT image; (b) visualization of LBP image; (c) LBP histogram.

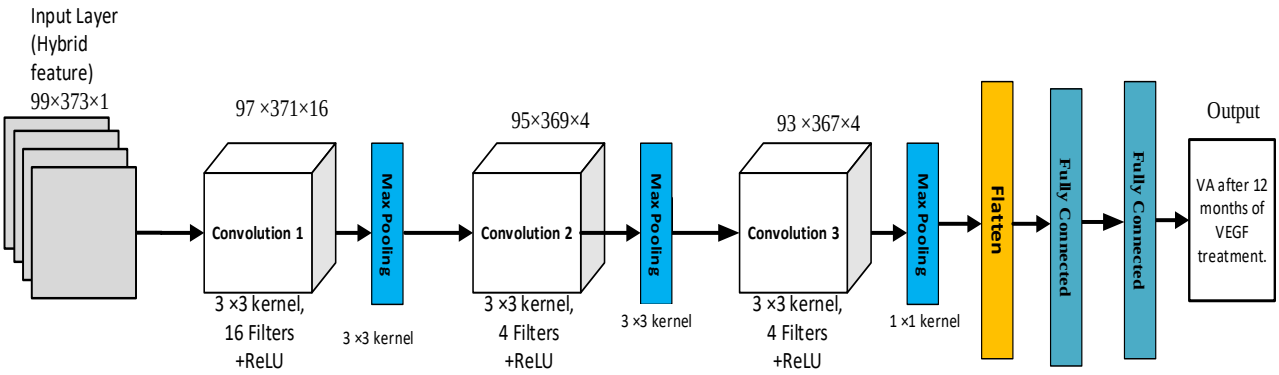


(a) (b) (c)

Figure 7.22: (a) OCT image; (b) visualization of HOG features applied on OCT image; (c) visualization of LBP representation for OCT.

7.4.3 Convolutional neural network (CNN) model

Once the hybrid features have been obtained from the OCT scans, they are fed to the CNN regressor to predict VA for RVO patients after 12 months of anti-VEGF treatment. The CNN model in this study is shown in Figure 7.23. The first layer of a 2-dimensional CNN is a convolutional layer which has a filter size of 16, a stride size of 1, and the padding = 'same', followed by a max-pooling layer with a size of 3 × 3 and a stride size of 2. The next layer is a convolutional layer with a filter size of 4 and a stride size of 1, which is followed by a max-pooling layer with size of 2×2 and a stride of 1. The third layer is a convolutional layer with a filter size of 4 and the stride is 1. It is followed by a max-pooling layer with a size 1×1 and a stride of 1. Each of the convolutional layers is followed by an activation layer that uses a rectified linear unit (ReLU). The subsequent layers are flattened, with each followed by a fully connected layer. The final layer is a linear function layer. Momentum term is 0.9 and learning rate 0.001. In the dataset, the input shape fed into the CNN model is 330 ×36927 pixels, and the output is a value that represents the VA of RVO patients after 12 months of treatment. Table 7.3 illustrated the layers' structure and their parameter values used in the CNN model. As the problem to be solved was set as a regression model, the evaluation metric was used to predict VA after 12 months of treatment using the mean absolute error (MAE) loss function, and the MAE formula is given in Equation 7.10:



$$MAE = \frac{1}{N}\sum_{j-1}^{N}\left|Y_j - \hat{Y}_j\right| \qquad (7.10)$$

Figure 7.23: A CNN sequence to predict VA after 12 months of Anti-VEGF treatment for RVO patients.

Layer array	Parameters and values
Input Layer	Two-dimension size (330 ×36927)
First Convolutional Layer	Kernel size (3 × 3), Filters (16) with stride 1 and padding =’same’
Output after of the first Convolution	99×371×16
ReLU	-
MaxPooling	Kernel size 3× 3 with stride 1
Second Convolutional Layer	Kernel size (3 × 3), Filters (4) with stride 1 and padding =’same’
Output after of the second Convolution	95 ×369×4
ReLU	-
MaxPooling	Kernel size 3× 3 with stride 1
Third Convolutional Layer	Kernel size (3 × 3), Filters (4) with stride 1 and padding =’same’
Output after of the third Convolution	93×367×4
ReLU	-
MaxPooling	Kernel size 1× 1 with stride 1
Flatten	136524
Fully Connected	16
Fully Connected	8
Linear activation Function	1
Regression	Output values

Table 7.3: CNN structure used in the regression model.

7.5 Evaluations and Results

This section evaluates the proposed method using the real clinical dataset to demonstrate its efficiency and compares the prediction results with the forecasts made by the ophthalmologists.

7.5.1 Experimental Settings

In the experiments, all of the OCT scans and accompanying EMR data were collected from the RVI for cases that received anti-VEGF intravitreal injections, over 1 year for MO secondary to RVO. In this experiment, 412 eligible eyes were identified (212 left eye, 194 CRVO, 205 males, mean age 72.6, mean delay between diagnosis and treatment 111.78 days) with a mean VA at treatment initiation of 50.3 early treatment diabetic retinopathy study (ETDRS) letters and a mean VA of 58.8 letters following 1 year of treatment. Figures

7.24 and 7.25 show the distribution of the target values of VA after 12 months of the treatment for the training and testing datasets. Both plots visualize histograms with 30 bins with kernel density plot overlaid on top of the target in training and testing dataset.

The data sets were split into two subsets with a ratio of 80/20, giving a training dataset of 330 subjects and a test dataset of 82 subjects. To evaluate the performance of the proposed method, 11 different ophthalmology medical staff were provided with the same test data for 82 subjects and exactly the same patient details. These data were divided in two subsets. The first subset for subjects 1 to 41 were given forecasts by 7 ophthalmology doctors, while the second subset for subjects 42 to 82 were given forecasts by another 4 ophthalmology doctors, as shown in Table 7.4.

Table 7.5 shows the descriptive statistics of the electronic medical records details of subjects in the training dataset. It includes information on the left/right eye, branch/central RVO diagnosis, male/female, age, baseline visual acuity, the Delay between diagnosis and treatment, and visual acuity after 12 months. In the first column, the Left/Right eye represents whether the patient's left or right eye was affected by the condition being studied. The value of 1 represents the left eye, while the value of 2 represents the right eye. Secondly, the Branch/central RVO diagnosis column represents whether the patient was diagnosed with branch retinal vein occlusion (BRVO) or central retinal vein occlusion (CRVO). The value of 1 represents BRVO, while the value of 2 represents CRVO. There are 330 patients in the sample, so the counts of 1s and 2s should add up to 330. The third column, Male/female, represents the gender of the patient. The value of 1 represents male, while the value of 2 represents female. There are 330 patients in the sample, so the counts of 1s and 2s should add up to 330. The column Age represents the age of the patients being studied. The mean age is 72.57 years, with a standard deviation of 10.86 years. The youngest patient in the sample is 36.57 years old, and the oldest is 94.61 years old. The fifth column is Baseline visual acuity, which represents the patients' visual acuity at the beginning of the study before any treatment was administered. The mean visual acuity is 50.33, with a standard deviation of 23.13. The lowest recorded visual acuity is 0, which could indicate that some patients had no visual acuity in one or both eyes. The sixtieth column, Delay between diagnosis and treatment, represents the amount of time that passed between the diagnosis of the condition being studied and the start of treatment. The mean Delay is 111.78 days, with a standard deviation of 263.18 days. The most extended Delay recorded is 1713 days, which is over 4.5 years. The last column, Visual acuity after 12 months (the target), represents the patients' visual acuity after 12 months of treatment. The mean visual acuity after 12 months

is 58.95, with a standard deviation of 21.07. The highest recorded visual acuity after 12 months is 85, which could indicate that some patients had a full or near-full recovery of their vision. Table 7.6 shows the dispersion and shape of subjects' electronic medical records details in the testing dataset. The table shows that the testing dataset includes 82 subjects. The mean values for the variables are similar to those in the training dataset. However, the standard deviation values for Age, Baseline visual acuity, Delay between diagnosis and treatment, and Visual acuity after 12 months are lower in the testing dataset, indicating less variability in these variables compared to the training dataset.

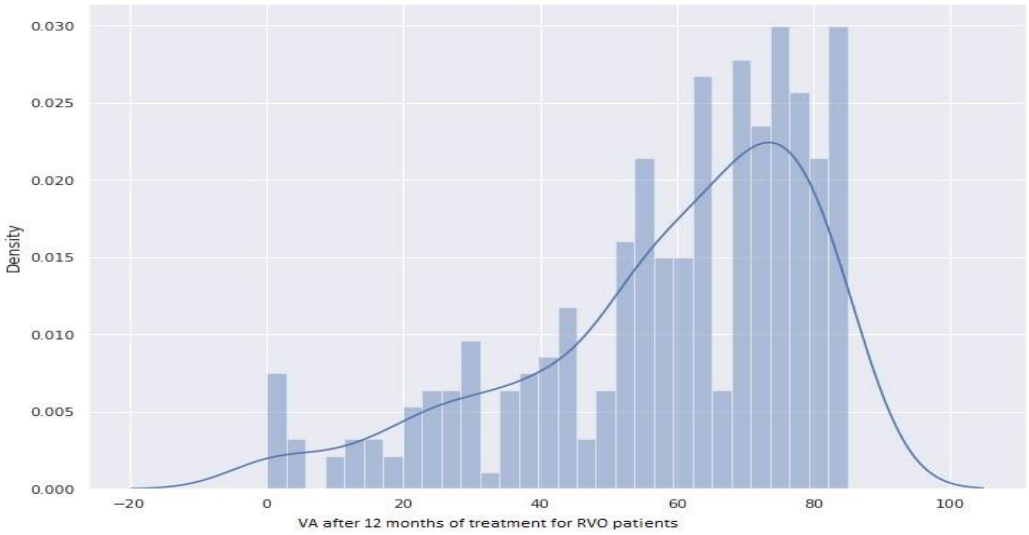


Figure 7.24: Shows a histogram of the dataset with 30 bins with a kernel density plot overlaid on top and distribution of target values of VA after 12 months of the treatment in the training dataset.

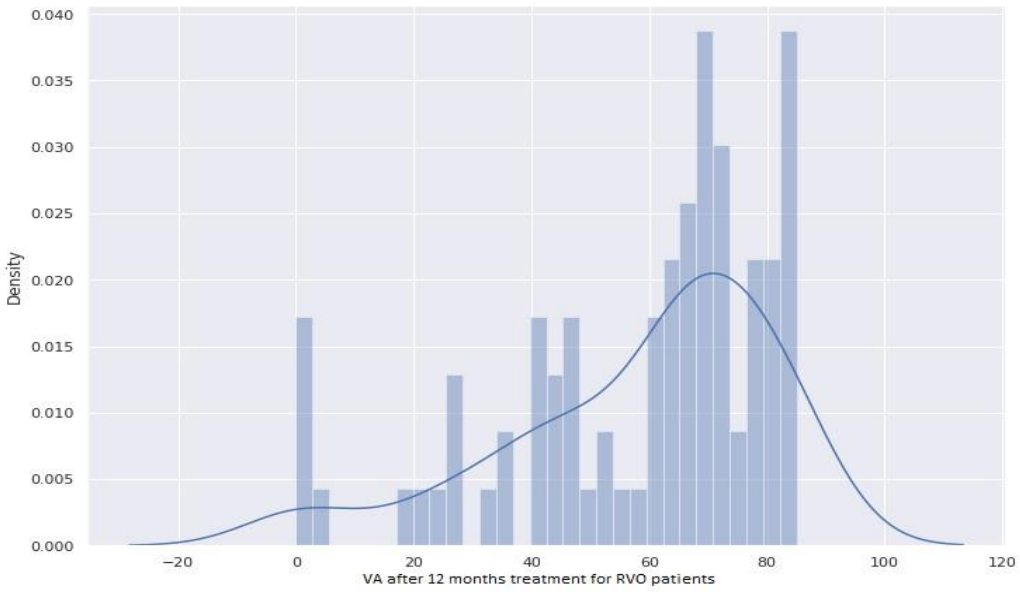


Figure 7.25: Shows a histogram of the dataset with 30 bins with a kernel density plot overlaid on top and distribution of the target values of VA after 12 months of treatment in the testing dataset.

Forecasting ophthalmology specialty clinician	Mean Absolute Error	Data set
ORS 1	16.07	Top subset (from 1 to 41)
ORS 2	16.32	Top subset (from 1 to 41)
ORS 3	15.49	Top subset (from 1 to 41)
ORS 4	15.95	Top subset (from 1 to 41)
ORS 5	17.98	Top subset (from 1 to 41)
ORS 6	14.76	Top subset (from 1 to 41)
ORS 7	17.27	Top subset (from 1 to 41)
ORS 8	16.34	Bottom subset (from 42 to 82)
ORS 9	16.80	Bottom subset (from 42 to 82)
ORS 10	18.49	Bottom subset (from 42 to 82)
ORS 11	18.51	Bottom subset (from 42 to 82)

Table 7.4: Clinician Forecasting of VA after 12 months of VEGF treatment for RVO patients, with 82 cases divided into subsets. The top subset is forecast by 7 ophthalmologists for patients 1 to 41; The bottom subset is forecast by another 4 ophthalmologists for patients 1 to 41; The

bottom subset is forecast by another 4 ophthalmologists for patients 42 to 82 (OSR = ophthalmology speciality registrar.)

	Left/Right eye	Branch/central RVO diagnosis	Male/female	Age	Baseline visual acuity	Delay between diagnosis and treatment	Visual acuity after 12 months
Count	330	330	330	330	330	330	330
Mean	1.49	1.47	1.49	72.57	50.33	111.78	58.95
Std	0.50	0.49	0.57	10.86	23.13	263.18	21.07
Min	1	1	1	36.57	0	0	0
Max	2	2	2	94.61	85	1713	85

Table 7.5: Dispersion and shape the electronic medical records details of subjects in training dataset.

	Left/Right eye	Branch/central RVO diagnosis	Male/female	Age	Baseline visual acuity	Delay between diagnosis and treatment	Visual acuity after 12 months
Count	82	82	82	82	82	82	82
Mean	1.46	1.49	1.5	72.76	50.34	107.61	58.51
Std	0.50	0.50	0.50	11.81	21.07	222.53	22.79
Min	1	1	1	34.76	0	0	0
Max	2	2	2	94.25	84	1379	85

Table 7.6: Dispersion and shape of the electronic medical records details of subjects in testing dataset.

7.5.2 Training Details

The pre-processed images were fed to the LBP, HOG, and GLCM to extract useful information. The feature vectors obtained from the LBP are captured in a histogram, with the number of points set to 20 and the radius parameter used was set to 8 pixels. The resulting histograms have 22 dimensions (bins) for each image. In the HOG, an 8 × 8 cell size and 1 × 1 block size were used in computing the features. The output of the HOG features extraction is a histogram with 36864 dimensions (bins) for each image. The result of the GLCM calculation selecting various pairs of distances and angles produced 7 GLCM matrices over 5 selected statistical features, thus generating 35 features. The LBP, HOG

descriptors and GLCM features were then concatenated and combined with the EMR data with 6 features. The resulting vectors were set to 36927, which represents the hybrid features fed to the CNN regressor. To train and evaluate the proposed method, a Python programming environment and supporting libraries provided by Anaconda Distributions were utilised. By monitoring the performance of the model, 'Adam' optimizer with a learning rate of 0.001, the batch size was kept at 10, and training was undertaken using 100 epochs were selected based on trial-and-error approach, where different learning rates were tested during training to find the optimal one that leads to the best performance, and MAE was used for the loss function.

7.5.3 Experimental Results and Analysis

To further confirm the efficiency of the proposed hybrid feature combination and the usage of EMR, this section presents the results of the experimental analysis conducted in three parts.

7.5.3.1 Comparison of different feature inputs.

Here, a CNN model was built as described in previous section, but different features were used as input to train the regressor and the results are shown in Table 7.7. Firstly, only the HOG features were fed into the CNN. It can be seen from Table 7.7 that when the HOG features are used, the value of its corresponding MAE is 16.67. Then, using only the LBP as the input to train the CNN regressor, the MAE value in this case is equal to 17.18, whereas using only the GLCM features the MAE was 16.86. Meanwhile, the combination of HOG and GLCM features gave an MAE of 17.35, while that of LBP and GLCM yielded an MAE of 16.93. Merging the three sets of features from GLCM, LBP, and HOG produced an MAE value of 16.61. Table 7.7 shows that the hybrid HOG/LBP feature combination produced the best MAE value of 16.57. However, compared with using only the HOG or LBP features, this performance improvement is not statistically significant. With the HOG features, the MAE improvement is around 0.61 and with the LBP feature, this value is around 0.1 while fusing all three features did not improve the results. The LBP mechanism works for each neighbouring pixel, and if its value is larger than that of the centre pixel then a value of 0 is entered; otherwise, a value of 1 is given. However, this method of discarding gradient magnitudes using only values of 1 and 0 may lead to the loss of a substantial amount of information. The HOG, on the other hand, takes account of the gradient magnitudes for the image. These are useful in identifying corners and edges, which represent substantial

information about the image. Thus, it appears that the LBP may lose such information. Generally, shape information is extracted by the HOG and textural information is provided by the LBP. The strength of the HOG is that it can detect edges and corners in images, whereas LBP captures local patterns. Both techniques attempt to use a similar type of information concerning gradients around a pixel. Thus, they are complementary in nature, and both may exhibit effective performance in recognition and object detection [152, 171].

7.5.3.2 Inclusion of EMR data.

As described previously, corresponding EMR data for all subjects were also available. Therefore, it was predicted that the addition of EMR data to the analysis would further improve performance, and the results are shown in Table 7.7 From Table 7.7 it can be seen that, after adding the EMR information, the performance in prediction is enhanced. For example, when only LBP was used as the visual feature of the input, an MAE of 17.18 resulted. However, when both LBP and EMR data were used as input features for the trained model, the prediction performance was further improved with an MAE value of 15.51. According to all cases listed, the results in Table 7.7 indicate that the addition of EMR information in both training and prediction stages boosts performance. Therefore, it can be concluded that the use of EMRs is a key factor in boosting performance. If EMR data are added, the performance of using an individual vision feature could outperform the hybrid vision feature combination. For example, if only the vision hybrid feature fusion (GLCM+LBP+HOG) is used, it can give an MAE of 16.61 for prediction performance. However, when using an individual vision feature and EMR in HOG+EMR, LBP+EMR or GLCM+EMR, the corresponding MAE values are 16.36, 15.51 and 14.38 respectively. The final step was the fusion of EMR with the LBP, HOG, and GLCM as inputs to train the CNN regressor. As Table 7.7 shows, our proposed method gives the best result with an MAE of 13.65. Hence, it is clear that by using hybrid feature fusion, more feature information can be extracted from the OCTs and the MAE and hence prediction performance can be further improved. Meanwhile, the experimental results also show the efficiency of adding EMR data, where its use with even a single individual feature to train the regressor gives a level of performance similar to that of hybrid feature fusion.

Feature Extraction Technique	Number of features	MAE
HOG only	36864	16.67
LBP only	22	17.18
GLCM only	35	16.86
HOG+LBP	36886	16.57
GLCM+HOG	36899	17.35
GLCM+LBP	57	16.93
GLCM+LBP+HOG	36921	16.61
HOG +EMR	36870	16.36
LBP+EMR	28	15.51
GLCM+EMR	41	14.38
LBP+HOG+ EMR	36892	14.69
GLCM+HOG+EMR	36905	14.95
GLCM+LBP+EMR	63	14.21
GLCM+LBP+HOG+EMR	36927	13.65

Table 7.7 : Mean Absolute Error for Different Inputs in CNN and Number of Features are used in the system.

7.5.3.3 Comparison with predictions by ophthalmologists.

To evaluate the performance of our model, its results were compared with the predictions made by 11 specialist ophthalmology registrars. Firstly, an evaluation dataset was selected which comprised 82 eyes from the cohort, for which there were no statistically significant differences in EMR data distribution compared to the dataset as a whole. Then, the selected dataset was divided into two non-overlapping subsets, named the top and bottom subsets, each containing 41 patients. Seven of the ophthalmologists examined the results for the top subset and another 4 ophthalmologists examined those for the bottom subset. Table 7.4 details the forecasting results reported by the ophthalmologists. The best performance among the forecasts for the first subset was achieved by specialist registrar No. 6 with an error of 14.76, whereas the worst performance was by specialty registrar 5 with an error of 17.98. Meanwhile, the best performance for the second subset was specialist registrar No. 8

(error=16.34), while the worst performance was by specialty registrar 11 (error=18.51). Further evaluation took place by comparing these results with the predictions of a specialist retina consultant who examined all 82 cases and achieved an MAE of 14.27. Although their performance bettered that of the 11 specialist registrars, our proposed method outperformed all of the doctors, with an MAE of 13.65. However, the fact that the MAE does not change significantly when going from 36,905 features to just 63 features suggests that many of the features extracted may not be providing much additional information to the model. It is possible that the model is already able to capture the most relevant information from a smaller set of features, and adding more features does not significantly improve its performance.

7.6 Summary

In this chapter, a deep learning-based system is proposed to predict VA among retinal vein occlusion (RVO) patients after 12 months of anti-VEGF treatment for macular oedema. In order to improve forecasting performance, a method using a combination of hybrid feature information and EMR data were used as training data for the proposed model. The results were compared with predictions made by 11 different ophthalmology specialty registrars, and 1 specialist retina consultant, and the proposed method achieved competitive or better performance. All the results shown above confirm the efficiency of the proposed method. For future research, models trained with greater volumes and quality of data could be explored. The present study drew on a hand-curated dataset, which necessitated the clinician-led prioritisation of 6 different data categories. With greater data curation resources, cleaner data recording, and a more efficient data curation pipeline, a greater breadth of data points could be collected for each case. Similarly, the heterogeneity of the real-world imaging protocols which produced the OCT data necessitated the analysis of single 2-dimensional images for each case. Large datasets using consistent imaging protocols are likely to support the training of models to give higher performance. These opportunities for further progress will be important as whilst achieving parity with expert ophthalmologists is a significant achievement, ophthalmologists do not yet routinely apply this level of ability to support their clinical decision making.

Chapter 8 MACHINE LEARNING FOR PREDICTING ANTI-VEGF TREATMENT DEMAND FOR RVO PATIENTS

Intravitreal injections of anti-vascular endothelial growth factor (VEGF) agents represent one of the most substantial advances in contemporary medicine. It can be considered as the current treatment for many diseases, including RVO disease. Knowing at the start of the treatment whether or not a patient will respond well, partially respond, or not respond would help to plan, set patient expectations, and improved healthcare efficiency Machine learning techniques have been applied to predict patients' long-term treatment needs from their initial presentation. These models aim to predict whether or not a patient will require low or high demand for anti-VEGF injections [90, 172, 173]. In this work, three machine learning models, which are the Random Forest and CatBoost classifiers and logistic regression, were trained to predict high and low anti-VEGF demand at the early stage of treatment and after three months of therapy, using the LBP, HOG, and GLCM feature extraction methods for the analysis of OCT images and fusing these features to improve the system's performance. In addition, regression of number of anti-VEGF injections given after 12 months was predicted using the Random Forest regressor. This chapter is organised as follows: The next section presents a brief overview of the machine learning models used in this work for the prediction of low and high demand in patients with RVO. Following this, the proposed method is introduced in Section 8.2. Section 8.3 discusses the evaluation and results, and finally Section 8.4 summarises this chapter.

8.1 Machine Learning Techniques for Classification

8.1.1 Logistic Regression

Logistic regression is a robust supervised machine learning algorithm used for binary classification. It was originally formulated by David Cox in 1958 [174] in building a logistic model. It is a method of modelling the possibility of a discrete output given input variables to study effects of predictor variables on categorical outcomes, for example, the absence or presence of a disease. Logistic regression can be considered to be linear regression but intended for classification problems. A logistic function is defined for logistic regression to model a dichotomous output variable. The main difference between linear regression and

logistic regression is that the latter's range is bounded between 0 and 1. Furthermore, unlike linear regression, logistic regression does not require linear relationship between inputs and output variables because of the use of a non-linear logarithmic transformation to the odds ratio. It is simply the ratio of two odds and is commonly used in medical studies as a measure of effect for categorical data. Odds are often expressed in terms of the probability of an event. This means that, if the probability of an event is p, then the odds are defined as *p/(1−p)* [175]:

$$\text{Logistic function} = \frac{1}{1+e^{-}x} \qquad (8.1)$$

where *x* is the input variable. Figure 8.1 illustrates an example of feeding an input *x* into a logistic function with values between -30 to 30. The logit function is defined as a natural log of the odds. The logistic function ranges between 0 and 1, whereas a logit function can be any real number from minus infinity to positive infinity *($P \in [-\infty, \infty]$):*

$$Odds = \frac{P}{1-P} \rightarrow \text{Logit}\,(P) = \ln\frac{P}{1-P} \qquad (8.2)$$

Replacing the logit of *P* to be equal to *mx* + *b* leads to:

$$\text{logit}\,(P) = mx + b \rightarrow mx + b = \ln\left(\frac{P}{1-P}\right)$$

$$\left(\frac{P}{1-P}\right) = e^{(mx+b)} \rightarrow P = \frac{e^{(mx+b)}}{1+e^{(mx+b)}} \rightarrow P(x) = \frac{1}{1+e^{-(mx+b)}} \qquad (8.3)$$

Unlike a linear regression where RMSE and MSE are utilized as loss function, logistic regression employs a loss function referred to as maximum likelihood estimation (MLE) which is a technique for estimating the values of a model's parameters (*m* and *b*) using observed data. These parameter values maximise the likelihood of finding the distribution that is more accurate to explain the observed data.

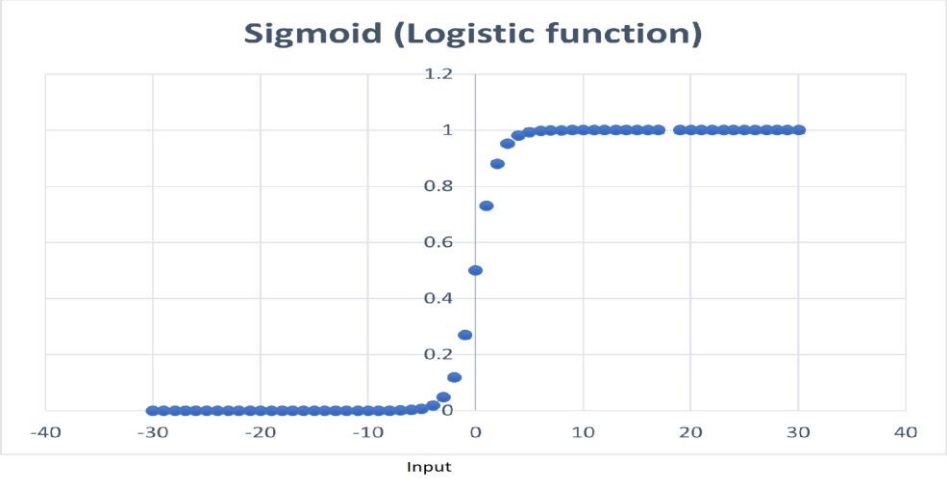


Figure 8.1: Logistic regression applied to a range of -30 to 30.

8.1.2 Random Forest Classifier

The Random Forest classifier is a support vector machine, tree-based classifier developed by Breiman [176]. Classification in random forests uses an ensemble methodology to gain the output. The Random Forest classifier utilises an extension to bagging techniques, where each classifier is created individually. This technique is based on various decision trees. Each decision tree contains decision nodes, root nodes, and leaf nodes, where the latter for each tree is the final output generated by that specific decision tree. To train different decision trees, training data is fed to the classifier. The features of a dataset will be selected randomly during the splitting of nodes. At each node, only selected features are examined in order to find the best split. A Random Forest classifier contains N trees, where N represents the number of trees to be grown and can be any value defined by the user. Finally, the majority-voting system is employed to select the final output of the trained forest system. The error rate of the forest classification is decided by the number of random features selected per decision node in a specific tree. This error rates relies on the correlation between any two trees and the strength of classification of each individual tree. Reducing and increasing the selected number of random features will affect both the correlation and strength of each individual tree. Figure 8.2 clarifies a simple Random Forest classifier.

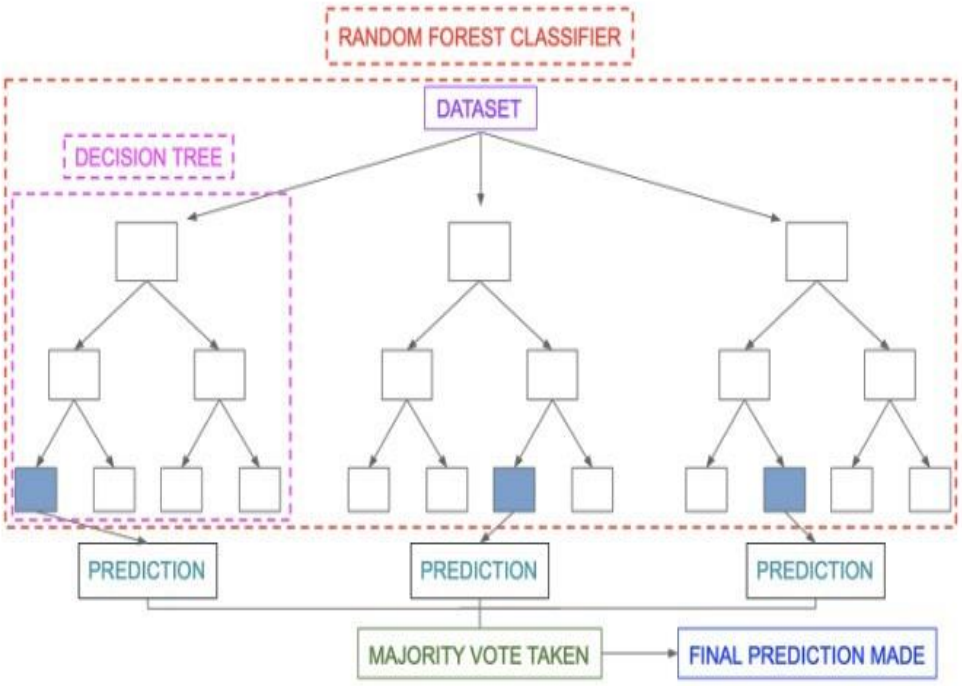


Figure 8.2: A simple Random Forest classifier[177].

8.1.3 Gradient Categorical Boosting Classifier (CatBoost)

Gradient boosting is a robust machine learning technique which gives state-of-the-art results in various applications. For many years, it has remained the essential technique for learning problems such as noisy data, web search, recommendation systems, weather forecasting, and many others [178-180]. Most gradient boosting algorithms utilise decision trees as base predictors, which are convenient to apply for numerical features. However, numerous datasets contain categorical features, which are significant for prediction. A categorical feature has a discrete set of values that do not necessarily match each other, such as user IDs, the names of patients, and addresses. The CatBoost algorithm is a type of gradient boosting that handles categorical features. Another benefit of this algorithm is that it applies a new schema for the calculation of leaf values when determining the tree structure, which reduces overfitting when building models with a novel gradient boosting schema. Additionally, for many applications, CatBoost provides results with default parameters, which can help in reducing the time spent on parameter tuning. A straightforward and common approach to deal with categorical features is to convert them into numerical values during pre-processing. An alternative and widely used strategy for handling categorical features is one-hot

encoding, which is implemented in CatBoost. This method can be applied during the pre-processing phase or during training.

Calculation model values for gradient estimation[181]

Input $\{(\mathbf{X_k}, Y_k)\}_{k=1}^{n}$, the number of trees I, ordered according to σ;

Mi ← 0 for i = $1..n$;

for iter ← 1 to I do

for i ← 1 to n do

for j ← 1 to i 1 do $g_j \leftarrow \frac{a}{da}\text{Loss}\left(y_j, a\right)\Big|_{a=M_i(\mathbf{X}_j)}$;

$M \leftarrow \text{LearnOneTree}\left(\left(\mathbf{X}_j, g_j\right) \text{ for } j = 1.i-1\right)$;

$M_i \leftarrow M_i + M$;

return $M_1 \ldots M_n; M_1(\mathbf{X}_1), M_2(\mathbf{X}_2) M_n(\mathbf{X}_n)$

In most gradient boosting decision trees, generating the next trees requires two steps: selecting the tree structure and setting values in leaves after the tree structure is fixed. In order to select the best tree structure, the algorithm recounts through various splits, constructs trees with these splits, allocates values in the leaves obtained, scores the trees and chooses the best split. In both phases, leaf values are calculated using gradient approximations [182]. Suppose that F^i is the model constructed after building the first i trees, and $g^i(X_k, Y_k)$ is the gradient value for the $k-th$ training sample. It is thought to be highly preferable to utilise unbiased estimates of the gradient step[181]. Then, with each separate model M_K, the gradient is estimated on X_K and used to score the resulting tree. The above pseudo-code [181] explains how to calculate model values for gradient estimation. In the case of CatBoost, its implementation provides the following idea: all M_i share the same tree structures, and random permutations are generated which are the same ones employed to calculate statistics for categorical features so as to improve the robustness of the algorithm, leading to reduced overfitting.

8.2 Proposed Method

Pipeline of the proposed method illustrates in Figure 8.3, showing the entire procedure of classification and regression. The pipeline comprises of three stages of pre-processing stage, feature extraction stage using HOG, LBP, and GLCM, regression model training using the Random Forest regressor, HvO, and LvO classifiers models, and the model evaluation stage.

8.2.1 Dataset

The RVO dataset was collected from the Royal Victoria Infirmary (RVI), Newcastle upon Tyne. The dataset represents the EMR and was interrogated for cases that received anti-VEGF intravitreal injections over 1 year along with non-invasive optical coherence tomography (OCT) images. 412 eligible eyes were identified (212 left eye, 194 CRVO, 205 males, mean age 72.6, mean delay between diagnosis and treatment 111.78 days) with a mean VA(VA) at treatment initiation of 50.3 early treatment diabetic retinopathy study (ETDRS) letters. In addition to EMR, OCT scans of the enrolled eyes acquired at each visit were investigated. Each image has two parts: an en-face and cross-sectional image. An OCT scan set contains either 6 or 25 different cross-sectional images arranged radially or orthogonally and centred on the fovea. These two styles represent a transition in the OCT protocol which occurred during the time in which the patients presented to the RVI. Images with the numbers 3 and 15 in each scale were selected and cropped to obtain only the retina part. Operations of morphological transformation and changing images to gray scale were then applied. The dataset was divided into 330 patients for training and 82 for testing.

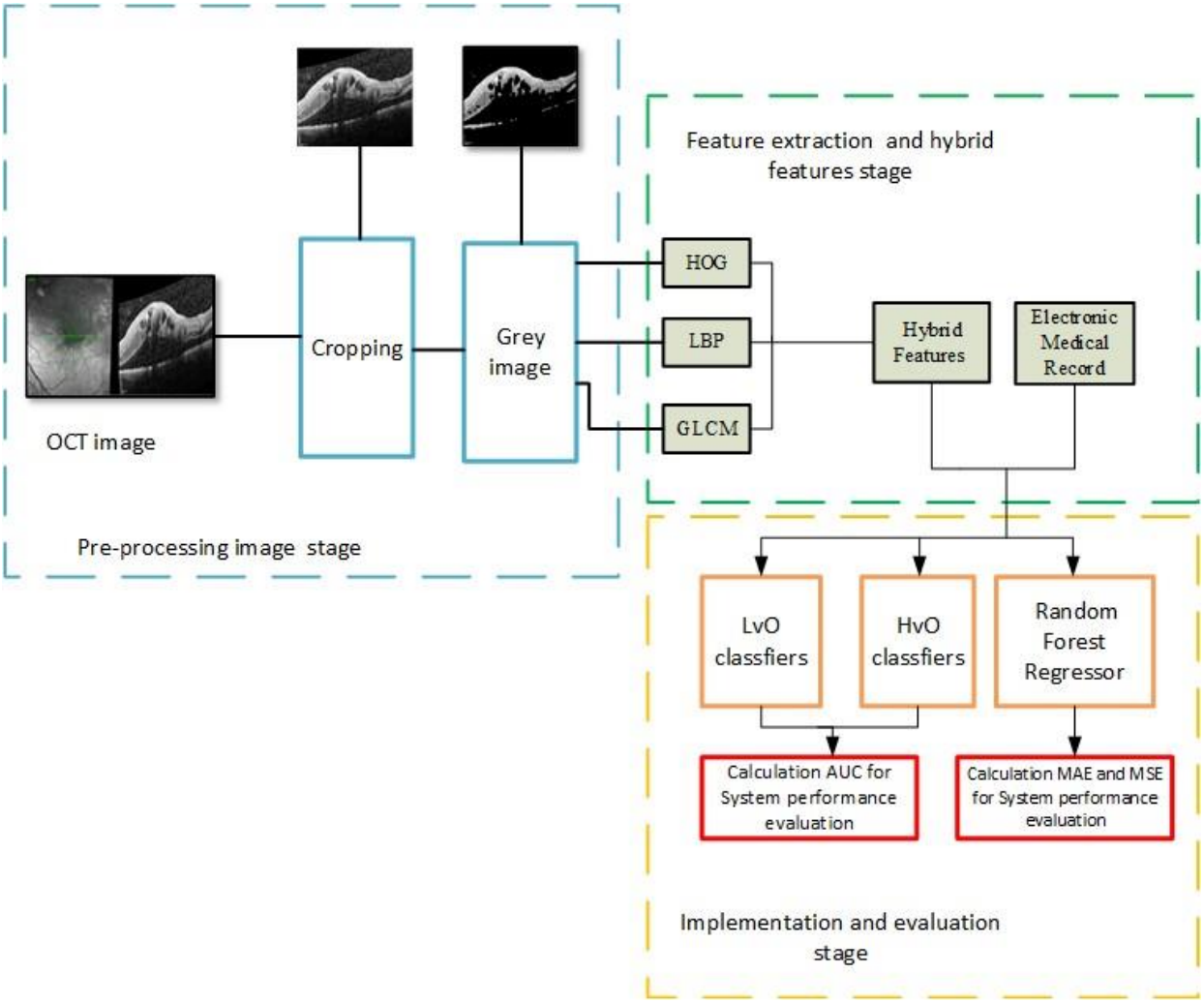


Figure 8.3: Pipeline of the proposed method, showing the entire procedure of classification and regression, including three stages of pre-processing, feature extraction and fusion, and regression model training using the Random Forest regressor, HvO, and LvO classifiers models, and the model evaluation stage.

8.2.2 Feature Extraction and Fusion

To extract the pattern information from the OCT scans, the three techniques of LBP, HOG and GLCM were utilised. These methods have been described in detail in previous chapters. In order to better describe the features of the image, a fusion of these methods was investigated. The uniform pattern LBPs were performed to calculate the histogram of each block where the 20 neighbours on a circle with a radius of 8 were considered. The histograms produced have a 22-dimensional feature vector. The total length of the HOG features is 64×64×9=36,864 pixels. Five statistical features of the GLCM were applied (correlation, constant, homogeneity, energy, and dissimilarity) on the seven GLCM matrices acquired using different angles and distances, and this process resulted in 35 GLCM features. These

features were combined to enhancing the performance of feature extraction. Furthermore, six features of EMR (age, sex, BRVO or CRVO status, VA at the first visit, left or right eye, number of days of delay between diagnosis and treatment) were added to the previous features to generate a total of 36864+22+35+6=36927 features.

8.2.3 Regression for anti-VEGF Demand after 12 Months

It is worth mentioning that the primary aim of this dissertation is to predict the VA of RVO patients after 12 months of anti-VEGF treatment. Levels of improvement in VA will relate to anti-VEGF response. This, in turn, can help patients and ophthalmologists to decide between continuing or stopping the treatment. However, in this chapter, anti-VEGF injection after 12 months of therapy is the target to predict low and high anti-VEGF demand. The three extraction feature methods of GLCM, HOG and LBP were combined, and the results fed to the Random Forest regressor in order to predict the anti-VEGF injections after one year of treatment. This type of regressor was selected because it is more appropriate when trained on relatively small datasets and provides interpretability in the prediction process[90]. The Random Forest regressor used 1000 trees, with a maximum tree depth of 100. To evaluate the effectiveness of the modelling technique in predicting the anti-VEGF injections after 12 months of treatment, the MAE was employed which reached 1.66, with an MSE of 2.08.

8.2.4 Prediction of Low and High Anti-VEGF Demand

This study investigated a number of machine learning techniques to predict low and high anti-VEGF demand utilising information from initial visits and after three months of therapy. Two binary classifiers were investigated to predict high and low anti-VEGF demand: high versus others (HvO) predicts the probability of an eye showing high demand; and low versus others (LvO) predicts the likelihood of an eye showing low demand. Treatment demand was grouped into three groups of low (0-5), moderate (6-9), and high (10+), same categories in this[90] . Three machine learning classifiers were implemented: logistic regression [183], the Random Forest classifier [184], and the CatBoost classifier [182]. LvO and HvO classifiers were computed at the three months visit using EMR only and using EMR merged with the features obtained from OCT scans using the three feature extraction methods. For training and testing the classifiers, the open-source package Scikit-learn version 0.24.2 of Python 3.8.10 was used.

8.3 Evaluation and Experimental Results

This section evaluates the system using the real clinical dataset to demonstrate its efficiency. The performance of the classifiers was measured using the area under the receiver operating characteristic curve (AUC). The motivation for using AUC is that classifier models provide a possibility of belonging to one class and not the estimated class itself of a sample. This is an effective measure of the diagnostic accuracy of a binary classifier system, describing how much the model is capable of distinguishing among classes. It gives a range of values between 0 to 1, where 0 reflects a perfectly inaccurate test, while 1 indicates a perfectly accurate test. Generally, an AUC = 0.5 represents a model with no class discrimination power. A value of AUC between 0.7 to 0.8 is considered to be acceptable, and higher AUC between 0.8 to 0.9 is excellent. Also, the reason why we do not depend on the accuracy metric is because the clinical use case requires more holistic assessments of sensitivity, specificity and other diagnostic accuracy statistics that are outside the scope of this work. In this study, predictions were made from the VA level at the initial visit as well as the visit after three months of treatment. The three machine learning classifiers utilised for both HvO and LvO were the Random Forest classifier, CatBoost classifier, and logistic regression, and in all cases, features extracted from the OCT scans and EMR were used. The following sections present the results of the experimental analysis conducted in two parts achieved by the hybrid feature combination and the usage of EMR.

8.3.1 High Versus Others

Tables 8.1-8.3 show the results of the application of the three classifiers for HvO with the initial visit and after three months of treatment using Random Forest, CatBoost, and Logistic regression. From Table 8.1, when using the Random Forest classifier, fusing GLCM, HOG, and EMR using 36905 features reached an AUC of 0.78 only in the initial visit. Similarly, Table 8.22 shows that the CatBoost classifier achieved the same result for AUC value (0.78) by adding VA after three months of treatment and using 36870 features combining HOG and EMR only. The same result obtained from logistic regression by fusing only GLCM and EMR in three months information described in table 8.3. Figures (8.4,8.5,8.6) show HvO classifiers using Random Forest, CatBoost, and Logistic regression with information from initial visit and visit after three months.

Feature extraction technique	Number of features	AUC (initial visit) 6 features	AUC (initial visit + three months) 7 features
HOG only	36864	0.61	0.61
LBP only	22	0.56	0.56
GLCM only	35	0.57	0.57
HOG+LBP	36886	0.63	0.63
GLCM+HOG	36899	0.70	0.70
GLCM+LBP	57	0.47	0.47
GLCM+LBP+HOG	36921	0.63	0.62
HOG +EMR	36870	0.69	0.67
LBP+EMR	28	0.54	0.53
GLCM+EMR	41	0.49	0.46
LBP+HOG+ EMR	36892	0.70	0.72
GLCM+HOG+EMR	36905	0.78	0.76
GLCM+LBP+EMR	63	0.52	0.50
GLCM+LBP+HOG+EMR	36927	0.64	0.69

Table 8.1 : HvO using Random Forest with initial visit and after three months of treatment.

Feature extraction technique	Number of features	AUC (initial visit) 6 features	AUC (initial visit + three months) 7 features
HOG only	36864	0.38	0.38
LBP only	22	0.25	0.24
GLCM only	35	0.38	0.38
HOG+LBP	36886	0.72	0.72
GLCM+HOG	36899	0.65	0.66
GLCM+LBP	57	0.69	0.69
GLCM+LBP+HOG	36921	0.56	0.56
HOG +EMR	36870	0.55	0.78
LBP+EMR	28	0.66	0.66
GLCM+EMR	41	0.53	0.58
LBP+HOG+ EMR	36892	0.65	0.68
GLCM+HOG+EMR	36905	0.57	0.64
GLCM+LBP+EMR	63	0.61	0.66
GLCM+LBP+HOG+EMR	36927	0.50	0.63

Table 8.2: HvO using CatBoost with initial visit and after three months of treatment.

Feature extraction technique	Number of features	AUC (initial visit) 6 features	AUC (initial visit + three months) 7 features
HOG only	36864	0.60	0.60
LBP only	22	0.35	0.35
GLCM only	35	0.61	0.61
HOG+LBP	36886	0.60	0.60
GLCM+HOG	36899	0.63	0.62
GLCM+LBP	57	0.61	0.61
GLCM+LBP+HOG	36921	0.62	0.62
HOG +EMR	36870	0.61	0.66
LBP+EMR	28	0.70	0.74
GLCM+EMR	41	0.72	0.78
LBP+HOG+ EMR	36892	0.62	0.66
GLCM+HOG+EMR	36905	0.61	0.62
GLCM+LBP+EMR	63	0.72	0.77
GLCM+LBP+HOG+EMR	36927	0.60	0.60

Table 8.3 : HvO using logistic regression with the initial visit and after three months of treatment.

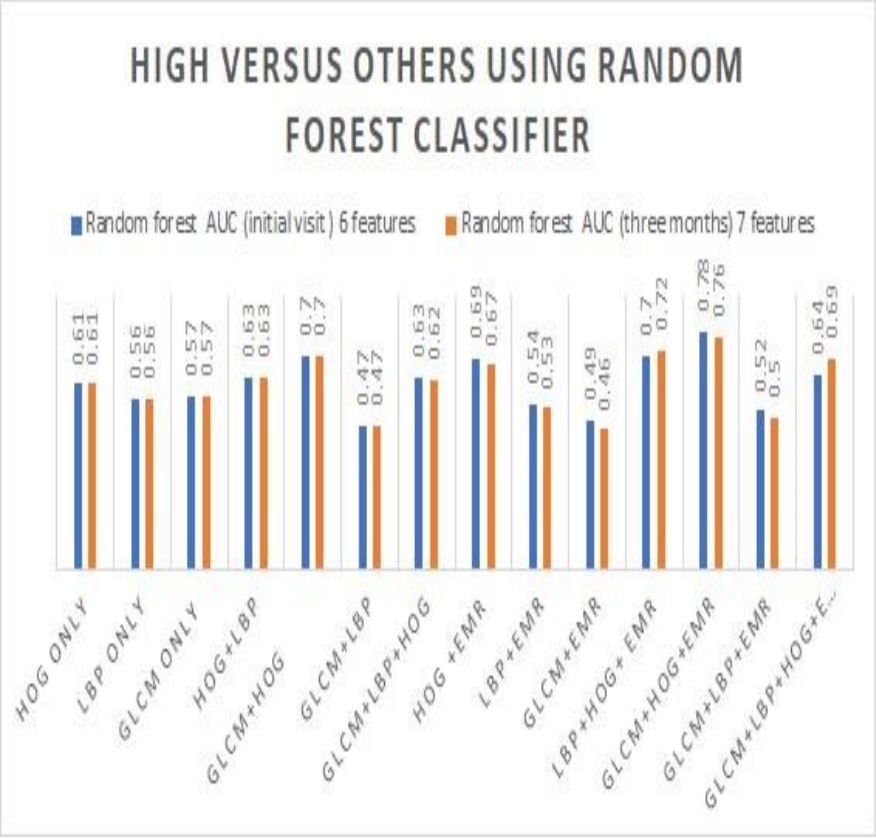


Figure 8.4 :HvO classifier using Random Forest with information from initial visit and visit after three months.

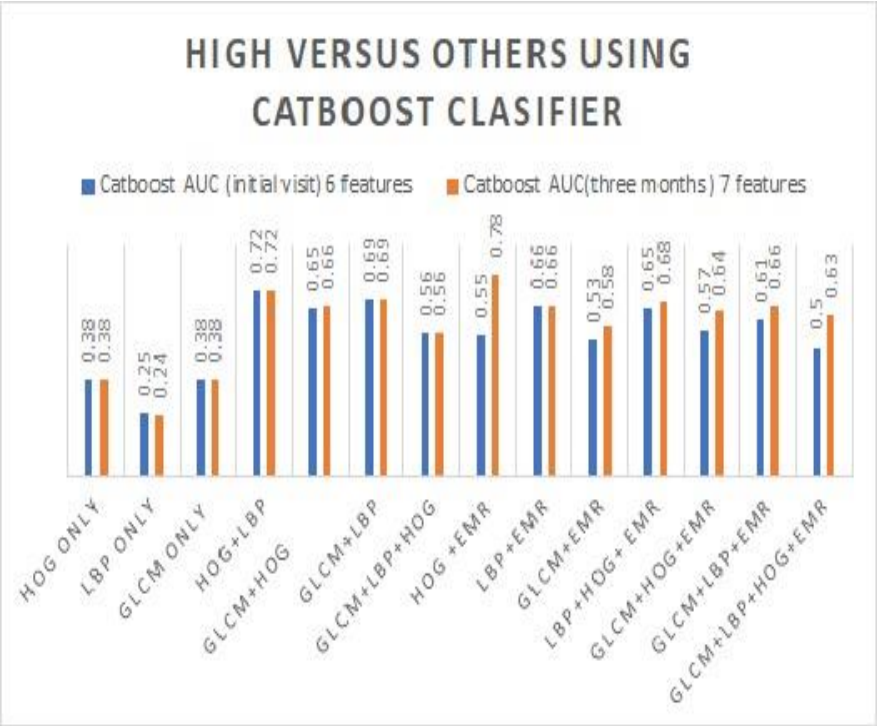


Figure 8.5 :HvO classifier using CatBoost classifier with information from initial visit and visit after three months.

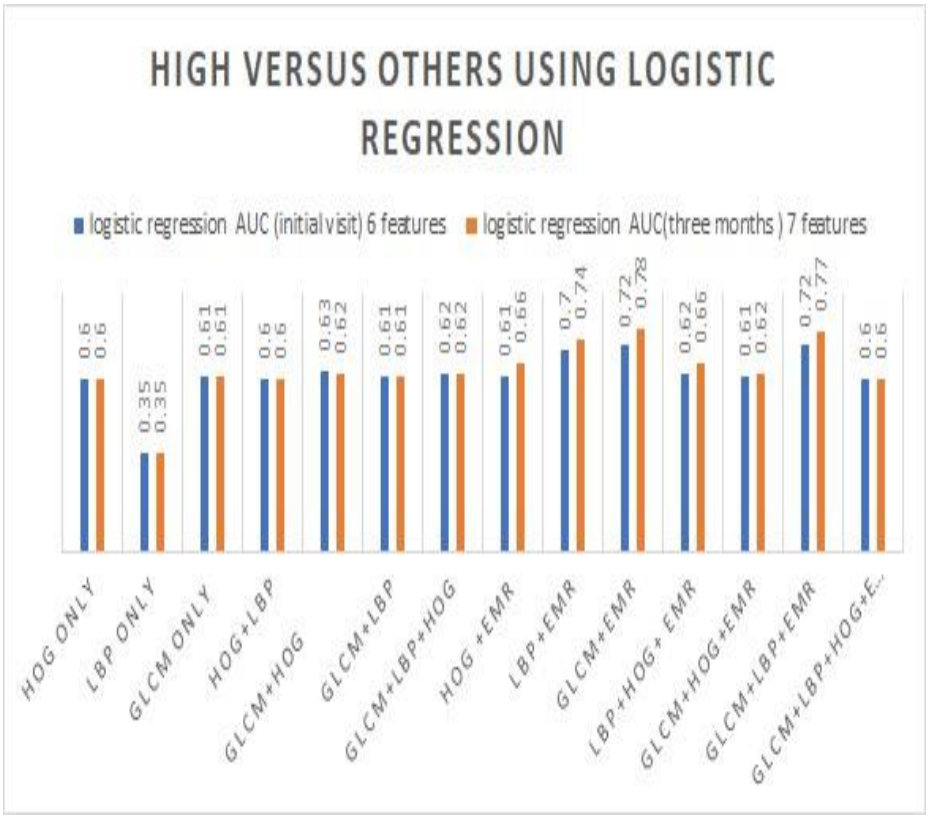


Figure 8.6 :HvO classifier using logistic regression classifier with information from initial visit and visit after three months.

8.3.2 Low Versus Others

Tables 8.4 - 8.6 illustrate the results using the three classifiers for LvO. Table 8.4 shows that using hybrid features and EMR based on VA at baseline, the Random Forest classifier provides an AUC = 0.64 using only GLCM features (35), whereas logistic regression reached an AUC of 0.60 using LBP+EMR and the CatBoost classifier reached an AUC of 0.59 using GLCM+LBP+EMR. There is no change in the AUC (0.64) using the Random Forest classifier even when adding VA information after the three months' VEGF injection. However, a slight improvement is observed using the CatBoost classifier with an AUC of 0.62 using GLCM+EMR features. In the case of logistic regression, there is no effect of adding the extra VA information after three months, as the best result reached was AUC = 0.59 when combining GLCM and HOG features.

This study shows that predicting individual demand for anti-VEGF treatment is possible. High demand for treatment was determined with reasonable accuracy already at baseline before the start of treatment with an AUC=0.78 using the Random Forest classifier. Also, using random forest, low demand can be identified at the initial visit with an AUC=0.64. For both HvO and LvO classifiers, Random Forest demonstrated higher AUC for prediction using baseline information. Figures (8.7,8.8, and 8.9) show Lv0 classifiers using Random Forest, CatBoost, and logistic regression with information from initial visit and visit after three months.

Feature extraction technique	Number of features	AUC (initial visit) 6 features	AUC (initial visit + three months) 7 features
HOG only	36864	0.53	0.53
LBP only	22	0.57	0.57
GLCM only	35	0.64	0.64
HOG+LBP	36886	0.50	0.50
GLCM+HOG	36899	0.57	0.57
GLCM+LBP	57	0.62	0.62
GLCM+LBP+HOG	36921	0.57	0.57
HOG +EMR	36870	0.51	0.49
LBP+EMR	28	0.60	0.60
GLCM+EMR	41	0.58	0.60
LBP+HOG+ EMR	36892	0.60	0.55
GLCM+HOG+EMR	36905	0.55	0.51
GLCM+LBP+EMR	63	0.51	0.53
GLCM+LBP+HOG+EMR	36927	0.52	0.54

Table 8.4 : Lv0 using Random Forest with initial visit and after three months of treatment.

Feature extraction technique	Number of features	AUC (initial visit) 6 features	AUC (initial visit + three months) 7 features
HOG only	36864	0.47	0.47
LBP only	22	0.54	0.54
GLCM only	35	0.51	0.50
HOG+LBP	36886	0.47	0.43
GLCM+HOG	36899	0.43	0.43
GLCM+LBP	57	0.49	0.49
GLCM+LBP+HOG	36921	0.59	0.59
HOG +EMR	36870	0.53	0.53
LBP+EMR	28	0.46	0.57
GLCM+EMR	41	0.54	0.62
LBP+HOG+ EMR	36892	0.56	0.44
GLCM+HOG+EMR	36905	0.46	0.45
GLCM+LBP+EMR	63	0.59	0.52
GLCM+LBP+HOG+EMR	36927	0.50	0.50

Table 8.5 : LvO using CatBoost with initial visit and after three months of treatment.

Feature extraction technique	Number of features	AUC (initial visit) 6 features	AUC (initial visit + three months) 7 features
HOG only	36864	0.39	0.39
LBP only	22	0.41	0.41
GLCM only	35	0.56	0.56
HOG+LBP	36886	0.40	0.40
GLCM+HOG	36899	0.59	0.59
GLCM+LBP	57	0.54	0.54
GLCM+LBP+HOG	36921	0.46	0.46
HOG +EMR	36870	0.43	0.41
LBP+EMR	28	0.60	0.51
GLCM+EMR	41	0.53	0.50
LBP+HOG+ EMR	36892	0.41	0.39
GLCM+HOG+EMR	36905	0.53	0.52
GLCM+LBP+EMR	63	0.53	0.49
GLCM+LBP+HOG+EMR	36927	0.52	0.49

Table 8.6: LvO using logistic regression with the initial visit and after three months of treatment.

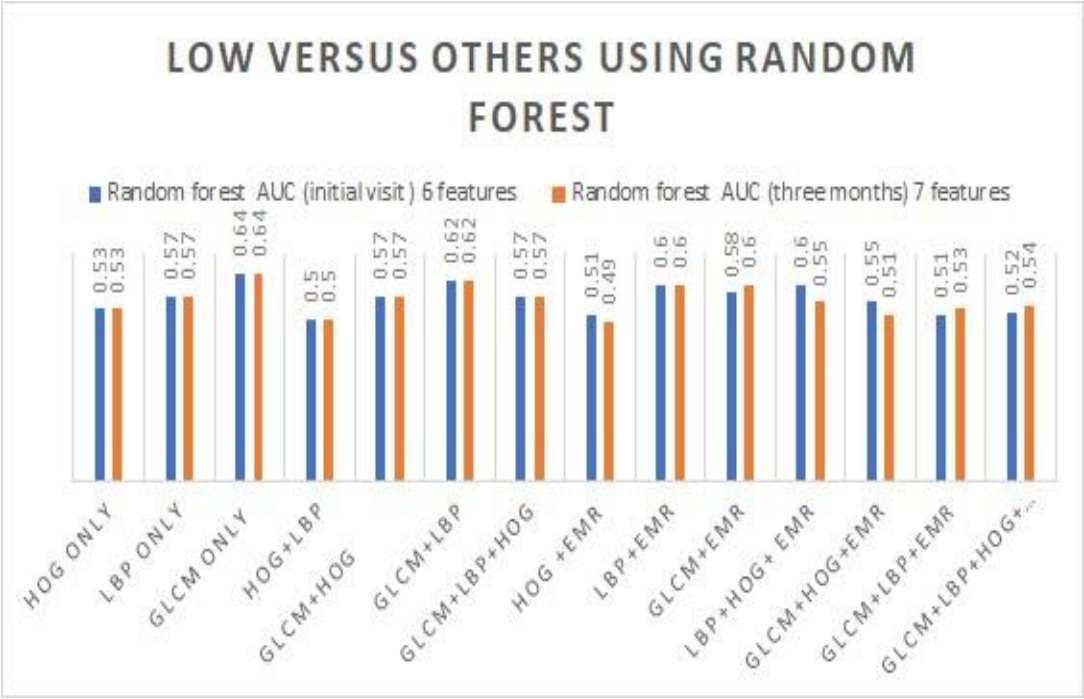


Figure 8.7: LvO classifier using Random Forest with information from initial visit and visit after three months.

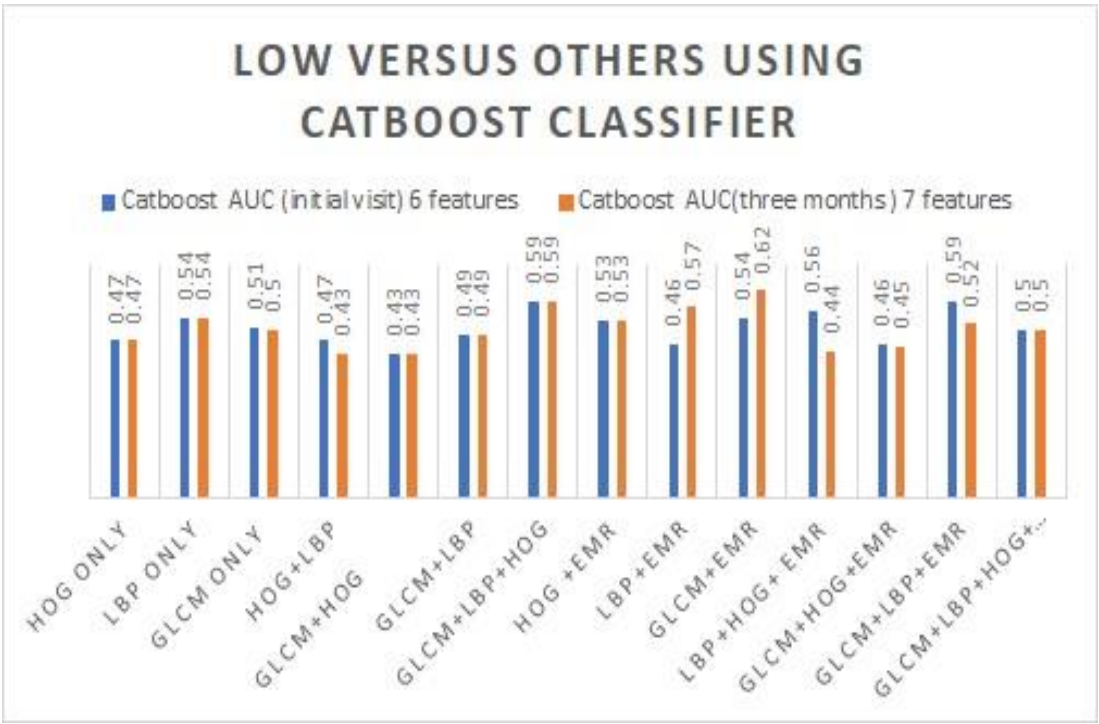


Figure 8.8 :LvO classifier using CatBoost with information from initial visit and visit after three months.

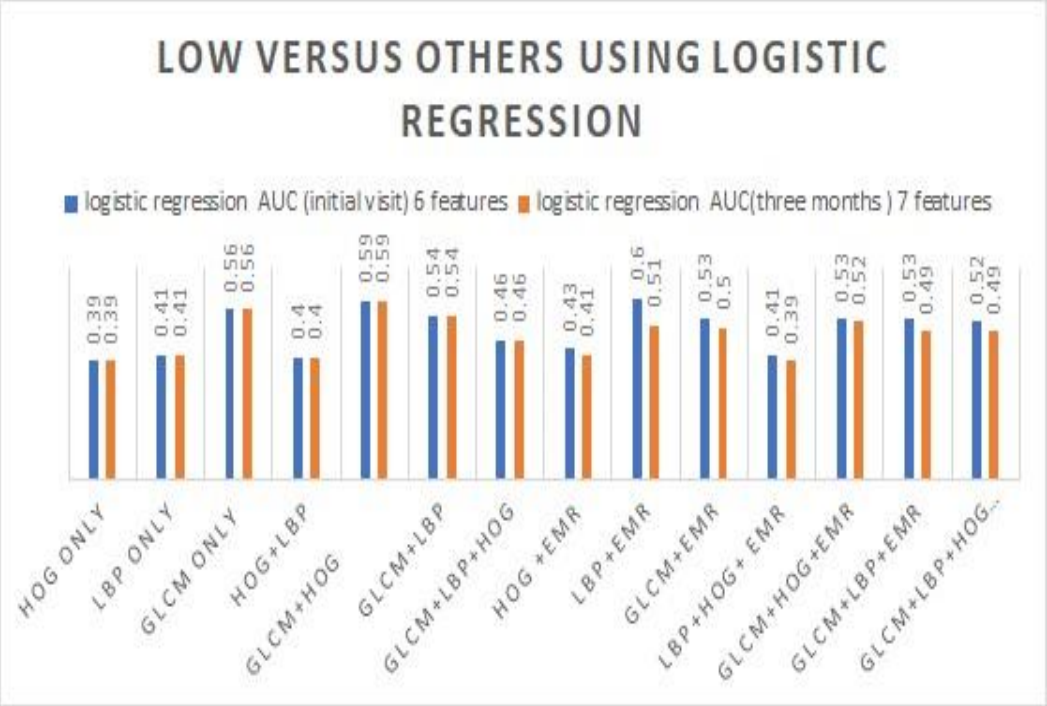


Figure 8.9: LvO classifier using logistic regression with information from initial visit and visit after three months.

8.4 Summary

In this chapter, RVO EMR and OCT imaging data from the RVI in Newcastle upon Tyne were analysed and their features extracted using various techniques. The system firstly used the Random Forest regressor in order to predict anti-VEGF demand after 12 months of treatment. Then, LvO and HvO classifiers at the baseline visit as well as the visit after three months were predicted using the three machine learning techniques of the Random Forest classifier, CatBoost classifier, and logistic regression. It is proven that the Random Forest classifier outperformed other techniques in both LvO and HvO classifiers with only baseline visit information, and it is demonstrated to work well in solving both classification and regression problems.

Chapter 9 CONCLUSION AND FUTURE WORK

9.1 Conclusion

RVO is the most common retinal vascular occlusive disorder and carries a significant burden of vision loss. It represents the second leading cause of retinal vascular blindness after diabetic retinopathy. Although the current treatment for this disease is based on intravitreal injection of anti-VEGF agents, patients show variable responses to the treatment. Knowledge at treatment initiation of whether RVO patients will respond well or poorly would help to plan and support patients' expectations. It would also enable more efficient and precise treatment solutions. The aim of this work was to examine machine and deep learning techniques to predict VA after one year of anti-VEGF treatment for RVO patients. The prediction's result could help RVO patients and their clinicians to determine the injection response and in turn to better inform decision making. Linear regression, several machine learning methods, and CNN were investigated, and novel methods were proposed to address this problem.

A new method using Linear regression and Random Forest regression is performed to predict 1-year VA from baseline data. Only EMR data were utilized in this work. Linear regression produced a model accounting for 57% of the accuracy seen in 1-year visual acuity. Using the same data, Random Forest regression surpassed this with the model accounting for 62% of accuracy.

Another new method is developed to predict VA 12 months after treatment initiation using several machine linear regression and machine learning techniques: Random Forest Regression, Ridge Regression, Multi-layer Perceptron (MLP) regression neural network, Partial Least Squares regression (PLS), and Lasso Regression. OCT images were collected at baseline and identified for analysis using feature extraction methods which are HOG and LBP and combined with the EMR dataset. Without access to vectorised imaging data, linear regression produced a model accounting for 57% of the accuracy seen in visual acuity 1 year after treatment initiation. With addition of OCT data this rose to 69%. Machine learning techniques including Ridge Regression, MLP, PLS Regression, Lasso Regression, and Random Forest Regression accounted for 71%, 72%, 70%, 71%, 75% of accuracy respectively. Random Forest Regressor outperformed the linear regression model and other techniques to provide the highest score where R^2= 75%.

A new method is addressed, a CNN technique to predict VA after 12 months of treatment. OCT images are pre-processed. A hybrid method using HOG, LBP, and GLCM is examined. These features are combined to improve the prognostic performance. The hybrid feature provides a good representation of OCT images to complement EMR data and is input to our CNN regressor which forecasts VA after one year of treatment. The performance of the system was evaluated by comparison with 11 different specialist ophthalmology registrars, and 1 specialist retina consultant. All the results proven confirm the efficiency of the proposed method.

Another experiment using machine learning methods is investigated for predicting high and low anti-VEGF treatment demand for RVO patients. GLCM, HOG, and LBP methods are utilized to extract useful features and merged with EMRs to be used as the input for training and testing the model for predicting anti-VEGF demand using Random Forest Regressor. Also, two binary classifications are trained to predict the probability of an eye showing High demand and Low demand Vs others using Logistic Regression, Random Forest Classifier, and CatBoost Classifier. In both Low vs others and High vs others classifiers using only baseline visit information, Random Forest Classifier outperformed other techniques. It is demonstrated to work well in solving both classification and regression problems.

This thesis showed that it is possible to predict one year of anti-VEGF treatment for individual with RVO. With basic demographic data and imaging data, machine learning techniques and a CNN model demonstrated an ability to predict visual prognosis for patients with RVO superior to clinicians. Such work could help clinicians personalise patient information and commissioners to analyse the cost-utility of treatment in more detail. To advance the model towards clinical implementation, the subsequent phase would entail external validation, which necessitates evaluating the model's performance on a comparable dataset from another clinical site. If the external validation yields favourable outcomes, a prospective clinical study could be conducted, where patients or different eye units would be randomized to receive consultations either assisted by the model or not. Subsequently, the outcomes between the two groups could be compared to determine differences in patient experience/satisfaction, cost, resultant visual outcomes, or adverse events/safety. These data could then substantiate the regulatory approval of

the model, enabling its clinical use. However, currently, the results indicate a promising achievement.

9.2 Future Work

Although the thesis presents novel methods for predicting VA after a year of anti-VEGF treatment for RVO patients, there are further opportunities. For example, the methods proposed in this work could be repeated with large numbers of OCT images or with more diverse clinical datasets such as OCT angiography and Fluorescein angiography (FA)images in the training and validation stages to ensure the reliability of a prognosis system. Furthermore, the model could be validated by large number of ophthalmologists in than who engaged during the research (11 ophthalmologists). Additionally, the model could be validated on an external dataset to assess the viability of applying this model across multiple clinical centres. Distinguishing between Hemiretinal Vein Occlusion (HRVO) and BRVO could deliver greater prognostic value from EMR data. This type occurs at the optic disc involving the inferior or superior hemifield. It is considered a mild form of CRVO with a similar natural history and the exact underlying aetiology but was categorised as BRVO within the dataset available here. Beyond improving and validating the model it would also be valuable to understand how a clinical intervention containing the tool could be designed for clinical deployment. This could involve human-computer-interaction methods and qualitative research methods to explore the perspectives of stakeholders who would influence the implementation of this tool.

References

[1] S. Yeh *et al.*, "Therapies for macular edema associated with central retinal vein occlusion: a report by the American Academy of Ophthalmology," *Ophthalmology,* vol. 122, no. 4, pp. 769-778, 2015.

[2] J. P. Ehlers *et al.*, "Therapies for macular edema associated with branch retinal vein occlusion: a report by the American Academy of Ophthalmology," *Ophthalmology,* vol. 124, no. 9, pp. 1412-1423, 2017.

[3] D. Shroff, A. Kothari, G. Bhatia, and C. Gupta, "Clinical Diagnosis of retinal vein occlusion," *International Journal of Ophthalmic Research,* vol. 2, no. 2, pp. 137-142, 2016.

[4] R. Katz *et al.*, "Active Learning of the Contrast Sensitivity Function as a New Clinical Endpoint for Retina Vein Occlusion," *Investigative Ophthalmology & Visual Science,* vol. 60, no. 9, pp. 2579-2579, 2019.

[5] A. Bora *et al.*, "Deep Learning for Identifying Retinal Vein Occlusion Features in Fundus Images," *Investigative Ophthalmology & Visual Science,* vol. 60, no. 9, pp. 1448-1448, 2019.

[6] A. E. Hoerl, R. W. Kannard, and K. F. Baldwin, "Ridge regression: some simulations," *Communications in Statistics-Theory and Methods,* vol. 4, no. 2, pp. 105-123, 1975.

[7] J. Ranstam and J. Cook, "LASSO regression," *Journal of British Surgery,* vol. 105, no. 10, pp. 1348-1348, 2018.

[8] M. Aitkin and R. Foxall, "Statistical modelling of artificial neural networks using the multi-layer perceptron," *Statistics and Computing,* vol. 13, no. 3, pp. 227-239, 2003.

[9] A. Höskuldsson, "PLS regression methods," *Journal of chemometrics,* vol. 2, no. 3, pp. 211-228, 1988.

[10] G. Biau and E. Scornet, "A random forest guided tour," *Test,* vol. 25, no. 2, pp. 197-227, 2016.

[11] B. Choudhury, P. H. Then, and V. Raman, "Automated Detection of Central Retinal Vein Occlusion Using Convolutional Neural Network," in *Big Data and Visual Analytics*: Springer, 2017, pp. 1-21.

[12] H. Kolb, E. Fernandez, and R. Nelson, "Webvision: the organization of the retina and visual system [Internet]," 1995.

[13] K. M. Meek and C. Knupp, "Corneal structure and transparency," *Progress in retinal and eye research,* vol. 49, pp. 1-16, 2015.

[14] H. Kolb, "Simple anatomy of the retina," 2012.

[15] R. S. Molday and O. L. Moritz, "Photoreceptors at a glance," *Journal of cell science,* vol. 128, no. 22, pp. 4039-4045, 2015.

[16] Y. Sun and L. E. Smith, "Retinal vasculature in development and diseases," *Annual review of vision science,* vol. 4, p. 101, 2018.

[17] T. Bek, "Regional morphology and pathophysiology of retinal vascular disease," *Progress in retinal and eye research,* vol. 36, pp. 247-259, 2013.

[18] S. S. Hayreh, "Acute retinal arterial occlusive disorders," *Progress in retinal and eye research,* vol. 30, no. 5, pp. 359-394, 2011.

[19] M. Ros, L. Magargal, and M. Uram, "Branch retinal-artery obstruction: a review of 201 eyes," *Annals of ophthalmology,* vol. 21, no. 3, pp. 103-107, 1989.

[20] S. Rogers *et al.*, "The prevalence of retinal vein occlusion: pooled data from population studies from the United States, Europe, Asia, and Australia," *Ophthalmology,* vol. 117, no. 2, pp. 313-319. e1, 2010.

[21] J. Clarkson *et al.*, "Baseline and early natural history report: the Central Vein Occlusion Study," *Archives of Ophthalmology,* vol. 111, no. 8, pp. 1087-1095, 1993.

[22] E. D. C.-C. S. Group, "Risk factors for branch retinal vein occlusion," *American journal of ophthalmology,* vol. 116, no. 3, pp. 286-296, 1993.

[23] E. Z. Rath, R. N. Frank, D. H. Shin, and C. Kim, "Risk factors for retinal vein occlusions: a case-control study," *Ophthalmology,* vol. 99, no. 4, pp. 509-514, 1992.

[24] Y. U. Shin *et al.*, "Prevalence and associated factors of retinal vein occlusion in the Korean National Health and Nutritional Examination Survey, 2008–2012: A cross-sectional observational study," *Medicine,* vol. 95, no. 44, 2016.
[25] L. L. Lim *et al.*, "Prevalence and risk factors of retinal vein occlusion in an Asian population," *British Journal of Ophthalmology,* vol. 92, no. 10, pp. 1316-1319, 2008.
[26] A. Hirota, H. K. Mishima, and Y. Kiuchi, "Incidence of retinal vein occlusion at the Glaucoma Clinic of Hiroshima University," *Ophthalmologica,* vol. 211, no. 5, pp. 288-291, 1997.
[27] O. A. Saatci, S. T. Ferliel, M. Ferliel, S. Kaynak, and M. H. Ergin, "Pseudoexfoliation and glaucoma in eyes with retinal vein occlusion," *International ophthalmology,* vol. 23, no. 2, pp. 75-78, 1999.
[28] P. R. O'Mahoney, D. T. Wong, and J. G. Ray, "Retinal vein occlusion and traditional risk factors for atherosclerosis," *Archives of Ophthalmology,* vol. 126, no. 5, pp. 692-699, 2008.
[29] P. Song, Y. Xu, M. Zha, Y. Zhang, and I. Rudan, "Global epidemiology of retinal vein occlusion: a systematic review and meta-analysis of prevalence, incidence, and risk factors," *Journal of global health,* vol. 9, no. 1, 2019.
[30] S. S. Hayreh, M. B. Zimmerman, and P. Podhajsky, "Incidence of various types of retinal vein occlusion and their recurrence and demographic characteristics," *American journal of ophthalmology,* vol. 117, no. 4, pp. 429-441, 1994.
[31] S. S. Hayreh and M. B. Zimmerman, "Branch retinal vein occlusion: natural history of visual outcome," *JAMA ophthalmology,* vol. 132, no. 1, pp. 13-22, 2014.
[32] B. V. O. S. Group, "Argon laser scatter photocoagulation for prevention of neovascularization and vitreous hemorrhage in branch vein occlusion. A randomized clinical trial. Branch Vein Occlusion Study Group," *Arch Ophthalmol,* vol. 104, pp. 34-41, 1986.
[33] S. S. Hayreh and M. S. Hayreh, "Hemi-central retinal vein occlusion: pathogenesis, clinical features, and natural history," *Archives of ophthalmology,* vol. 98, no. 9, pp. 1600-1609, 1980.
[34] J. Garcia-Arumi, A. Boixadera, V. Martinez-Castillo, H. Blasco, A. Lavaque, and B. Corcostegui, "Radial optic neurotomy for management of hemicentral retinal vein occlusion," *Archives of Ophthalmology,* vol. 124, no. 5, pp. 690-695, 2006.
[35] A. R. Hunt *et al.*, "Hemiretinal vein occlusion 12-month outcomes are unique with vascular endothelial growth factor inhibitors: data from the Fight Retinal Blindness! Registry," *British Journal of Ophthalmology,* 2022.
[36] M. Laouri, E. Chen, M. Looman, and M. Gallagher, "The burden of disease of retinal vein occlusion: review of the literature," *Eye,* vol. 25, no. 8, pp. 981-988, 2011.
[37] S. S. Hayreh and M. B. Zimmerman, "Fundus changes in branch retinal vein occlusion," *Retina (Philadelphia, Pa.),* vol. 35, no. 5, p. 1016, 2015.
[38] D. Varma, S. Cugati, A. Lee, and C. Chen, "A review of central retinal artery occlusion: clinical presentation and management," *Eye,* vol. 27, no. 6, pp. 688-697, 2013.
[39] R. L. McIntosh *et al.*, "Natural history of central retinal vein occlusion: an evidence-based systematic review," *Ophthalmology,* vol. 117, no. 6, pp. 1113-1123. e15, 2010.
[40] Z. Oztas, C. Akkin, S. Nalcaci, O. Ilim, and F. Afrashi, "Branch retinal vein occlusion: The importance of the topographical distribution of retinal vessels among risk factors," *Eye,* vol. 31, no. 5, pp. 726-731, 2017.
[41] H. R. Novotny and D. L. Alvis, "A method of photographing fluorescence in circulating blood in the human retina," *Circulation,* vol. 24, no. 1, pp. 82-86, 1961.
[42] H. Osamura, T. Shiba, T. Itokawa, T. Matsumoto, and Y. Hori, "Relationships among ocular blood flow shown by laser speckle flowgraphy, retinal arteriosclerotic change, and chorioretinal circulation time obtained by fluorescein angiography," *Journal of Ophthalmology,* vol. 2017, 2017.
[43] M. R. Stein and C. W. Parker, "Reactions following intravenous fluorescein," *American Journal of Ophthalmology,* vol. 72, no. 5, pp. 861-868, 1971.

[44] C. S. Lee *et al.*, "Generating retinal flow maps from structural optical coherence tomography with artificial intelligence," *Scientific reports,* vol. 9, no. 1, pp. 1-11, 2019.
[45] P. J. Patel *et al.*, "Spectral-domain optical coherence tomography imaging in 67 321 adults: associations with macular thickness in the UK Biobank Study," *Ophthalmology,* vol. 123, no. 4, pp. 829-840, 2016.
[46] L. Nicholson, S. J. Talks, W. Amoaku, K. Talks, and S. Sivaprasad, "Retinal vein occlusion (RVO) guideline: executive summary," vol. 36, ed: Nature Publishing Group, 2022, pp. 909-912.
[47] T. Theelen and M. M. Teussink, "Inspection of the human retina by optical coherence tomography," in *Retinal Gene Therapy*: Springer, 2018, pp. 351-358.
[48] S. Kishi, "Impact of swept source optical coherence tomography on ophthalmology," *Taiwan journal of ophthalmology,* vol. 6, no. 2, pp. 58-68, 2016.
[49] V. Doblhoff-Dier *et al.*, "Measurement of the total retinal blood flow using dual beam Fourier-domain Doppler optical coherence tomography with orthogonal detection planes," *Biomedical optics express,* vol. 5, no. 2, pp. 630-642, 2014.
[50] G. Tsai, T. Banaee, F. F. Conti, and R. P. Singh, "Optical coherence tomography angiography in eyes with retinal vein occlusion," *Journal of ophthalmic & vision research,* vol. 13, no. 3, p. 315, 2018.
[51] T. Sugiyama, "Basic technology and clinical applications of the updated model of laser speckle flowgraphy to ocular diseases," in *Photonics*, 2014, vol. 1, no. 3: MDPI, pp. 220-234.
[52] X. Wei, P. K. Balne, K. E. Meissner, V. A. Barathi, L. Schmetterer, and R. Agrawal, "Assessment of flow dynamics in retinal and choroidal microcirculation," *Survey of Ophthalmology,* vol. 63, no. 5, pp. 646-664, 2018.
[53] C. Riva, B. Ross, and G. B. Benedek, "Laser Doppler measurements of blood flow in capillary tubes and retinal arteries," *Investigative Ophthalmology & Visual Science,* vol. 11, no. 11, pp. 936-944, 1972.
[54] C. Riva and B. Petrig, "Blue field entoptic phenomenon and blood velocity in the retinal capillaries," *JOSA,* vol. 70, no. 10, pp. 1234-1238, 1980.
[55] V. Rajan, B. Varghese, T. G. van Leeuwen, and W. Steenbergen, "Review of methodological developments in laser Doppler flowmetry," *Lasers in medical science,* vol. 24, no. 2, pp. 269-283, 2009.
[56] L. Nicholson, S. J. Talks, W. Amoaku, K. Talks, and S. Sivaprasad, "Retinal vein occlusion (RVO) guideline: executive summary," *Eye,* vol. 36, no. 5, pp. 909-912, 2022/05/01 2022, doi: 10.1038/s41433-022-02007-4.
[57] F. Coscas, G. Coscas, E. Souied, S. Tick, and G. Soubrane, "Optical coherence tomography identification of occult choroidal neovascularization in age-related macular degeneration," *American journal of ophthalmology,* vol. 144, no. 4, pp. 592-599. e2, 2007.
[58] A. A. Khanifar, A. F. Koreishi, J. A. Izatt, and C. A. Toth, "Drusen ultrastructure imaging with spectral domain optical coherence tomography in age-related macular degeneration," *Ophthalmology,* vol. 115, no. 11, pp. 1883-1890. e1, 2008.
[59] R. Klein, B. E. Klein, M. D. Knudtson, S. M. Meuer, M. Swift, and R. E. Gangnon, "Fifteen-year cumulative incidence of age-related macular degeneration: the Beaver Dam Eye Study," *Ophthalmology,* vol. 114, no. 2, pp. 253-262, 2007.
[60] S. R. Sadda *et al.*, "Relationship between angiographic and optical coherence tomographic (OCT) parameters for quantifying choroidal neovascular lesions," *Graefe's Archive for Clinical and Experimental Ophthalmology,* vol. 248, no. 2, pp. 175-184, 2010.
[61] T. Y. Wong and I. U. Scott, "Retinal-vein occlusion," *New England Journal of Medicine,* vol. 363, no. 22, pp. 2135-2144, 2010.

[62] L. P. Aiello *et al.*, "Vascular endothelial growth factor in ocular fluid of patients with diabetic retinopathy and other retinal disorders," *New England Journal of Medicine,* vol. 331, no. 22, pp. 1480-1487, 1994.
[63] A. Stahl, H. Agostini, L. L. Hansen, and N. Feltgen, "Bevacizumab in retinal vein occlusion-results of a prospective case series," *Graefe's Archive for Clinical and Experimental Ophthalmology,* vol. 245, no. 10, pp. 1429-1436, 2007.
[64] T. Yilmaz and M. Cordero-Coma, "Use of bevacizumab for macular edema secondary to branch retinal vein occlusion: a systematic review," *Graefe's Archive for Clinical and Experimental Ophthalmology,* vol. 250, no. 6, pp. 787-793, 2012.
[65] L. Hall *et al.*, "Prospective evaluation of intravitreal bevacizumab for ischemic central retinal vein occlusion," *International Journal of Retina and Vitreous,* vol. 5, no. 1, pp. 1-8, 2019.
[66] P. Hykin *et al.*, "Clinical effectiveness of intravitreal therapy with ranibizumab vs aflibercept vs bevacizumab for macular edema secondary to central retinal vein occlusion: a randomized clinical trial," *JAMA ophthalmology,* vol. 137, no. 11, pp. 1256-1264, 2019.
[67] M. Ashraf and A. A. Souka, "Steroids in central retinal vein occlusion: is there a role in current treatment practice?," *Journal of Ophthalmology,* vol. 2015, 2015.
[68] Z. F. Bashshur, R. N. Ma'luf, S. Allam, F. A. Jurdi, R. S. Haddad, and N. Baha'N, "Intravitreal Triamcinolone for the Management of Macular Edema Dueto Nonischemic Central Retinal Vein Occlusion," *Archives of ophthalmology,* vol. 122, no. 8, pp. 1137-1140, 2004.
[69] T. Higashiyama, O. Sawada, M. Kakinoki, T. Sawada, H. Kawamura, and M. Ohji, "Prospective comparisons of intravitreal injections of triamcinolone acetonide and bevacizumab for macular oedema due to branch retinal vein occlusion," *Acta Ophthalmologica,* vol. 91, no. 4, pp. 318-324, 2013.
[70] T. Qian, M. Zhao, and X. Xu, "Comparison between anti-VEGF therapy and corticosteroid or laser therapy for macular oedema secondary to retinal vein occlusion: A meta-analysis," *Journal of Clinical Pharmacy and Therapeutics,* vol. 42, no. 5, pp. 519-529, 2017.
[71] C. Li *et al.*, "Efficacy of panretinal laser in ischemic central retinal vein occlusion: A systematic review," *Experimental and Therapeutic Medicine,* vol. 17, no. 1, pp. 901-910, 2019.
[72] P. A. Campochiaro *et al.*, "Scatter photocoagulation does not reduce macular edema or treatment burden in patients with retinal vein occlusion: the RELATE trial," *Ophthalmology,* vol. 122, no. 7, pp. 1426-1437, 2015.
[73] J. Callizo *et al.*, "Bevacizumab versus bevacizumab and macular grid photocoagulation for macular edema in eyes with non-ischemic branch retinal vein occlusion: results from a prospective randomized study," *Graefe's Archive for Clinical and Experimental Ophthalmology,* vol. 257, no. 5, pp. 913-920, 2019.
[74] H. Liu, W. Zhang, Z. Xu, R. W. Caldwell, R. B. Caldwell, and S. E. Brooks, "Hyperoxia causes regression of vitreous neovascularization by downregulating VEGF/VEGFR2 pathway," *Investigative ophthalmology & visual science,* vol. 54, no. 2, pp. 918-931, 2013.
[75] I. Šínová *et al.*, "Correlation between ischemic index of retinal vein occlusion and oxygen saturation in retinal vessels," *American journal of ophthalmology,* vol. 188, pp. 74-80, 2018.
[76] J.-A. C. Pournaras, I. K. Petropoulos, J.-L. Munoz, and C. J. Pournaras, "Experimental retinal vein occlusion: effect of acetazolamide and carbogen (95% O2/5% CO2) on preretinal PO2," *Investigative ophthalmology & visual science,* vol. 45, no. 10, pp. 3669-3677, 2004.
[77] J. G. Arroyo, K. Dastgheib, and D. L. Hatchell, "Antithrombotic effect of ticlopidine in an experimental model of retinal vein occlusion," *Japanese journal of ophthalmology,* vol. 45, no. 4, pp. 359-362, 2001.

[78] E. M. Kohner, J. Pettit, A. Hamilton, C. Bulpitt, and C. Dollery, "Streptokinase in central retinal vein occlusion: a controlled clinical trial," *Br Med J,* vol. 1, no. 6009, pp. 550-553, 1976.
[79] A. Lazo-Langner, J. Hawel, W. Ageno, and M. J. Kovacs, "Low molecular weight heparin for the treatment of retinal vein occlusion: a systematic review and meta-analysis of randomized trials," *Haematologica,* vol. 95, no. 9, p. 1587, 2010.
[80] H. Ameri, J.-G. Kim, T. Ratanapakorn, G. J. Chader, and M. S. Humayun, "Intravitreal and subretinal injection of tissue plasminogen activator (tPA) in the treatment of experimentally created retinal vein occlusion in rabbits," *Retina,* vol. 28, no. 2, pp. 350-355, 2008.
[81] L.-O. Hattenbach *et al.*, "Retinal vein occlusion and low-dose fibrinolytic therapy (ROLF): a prospective, randomized, controlled multicenter study of low-dose recombinant tissue plasminogen activator versus hemodilution in retinal vein occlusion," *Retina,* vol. 29, no. 7, pp. 932-940, 2009.
[82] W. Ageno, J. Beyer-Westendorf, D. A. Garcia, A. Lazo-Langner, R. D. McBane, and M. Paciaroni, "Guidance for the management of venous thrombosis in unusual sites," *Journal of thrombosis and thrombolysis,* vol. 41, no. 1, pp. 129-143, 2016.
[83] E. Y. Chew, "Laser Photocoagulation and Intravitreal Injection of Triamcinolone for Retinal Vein OcclusionsTREATMENT FOR RETINAL VEIN OCCLUSIONS," *JAMA,* vol. 302, no. 15, pp. 1693-1695, 2009.
[84] L. J. Cehofski, B. Honoré, and H. Vorum, "A review: proteomics in retinal artery occlusion, retinal vein occlusion, diabetic retinopathy and acquired macular disorders," *International journal of molecular sciences,* vol. 18, no. 5, p. 907, 2017.
[85] M. Rohm *et al.*, "Predicting visual acuity by using machine learning in patients treated for neovascular age-related macular degeneration," *Ophthalmology,* vol. 125, no. 7, pp. 1028-1036, 2018.
[86] S. M. Waldstein *et al.*, "Evaluating the impact of vitreomacular adhesion on anti-VEGF therapy for retinal vein occlusion using machine learning," *Scientific reports,* vol. 7, no. 1, pp. 1-11, 2017.
[87] D. Nagasato *et al.*, "Deep neural network-based method for detecting central retinal vein occlusion using ultrawide-field fundus ophthalmoscopy," *Journal of ophthalmology,* vol. 2018, 2018.
[88] D. Nagasato *et al.*, "Deep-learning classifier with ultrawide-field fundus ophthalmoscopy for detecting branch retinal vein occlusion," *International Journal of Ophthalmology,* vol. 12, no. 1, p. 94, 2019.
[89] J. De Fauw *et al.*, "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nature medicine,* vol. 24, no. 9, pp. 1342-1350, 2018.
[90] M. Gallardo *et al.*, "Machine Learning Can Predict Anti–VEGF Treatment Demand in a Treat-and-Extend Regimen for Patients with Neovascular AMD, DME, and RVO Associated Macular Edema," *Ophthalmology retina,* vol. 5, no. 7, pp. 604-624, 2021.
[91] H. J. Hogg, S. J. Talks, M. Pearce, and S. Di Simplicio, "Real-world visual and neovascularisation outcomes from anti-VEGF in central retinal vein occlusion," *Ophthalmic Epidemiology,* vol. 28, no. 1, pp. 70-76, 2021.
[92] C. Papudesu *et al.*, "Ellipsoid zone status and its association with visual acuity in eyes with macular edema in the Study of COmparative Treatments for REtinal Vein Occlusion 2 (SCORE 2)," *Investigative Ophthalmology & Visual Science,* vol. 60, no. 9, pp. 2590-2590, 2019.
[93] E. Matthe, P. Eulitz, and O. Furashova, "Spectral-domain optical coherence tomography (SD-OCT) findings in retinal vein occlusion–Can retinal ischemia be detected in OCT findings?," *Investigative Ophthalmology & Visual Science,* vol. 60, no. 9, pp. 2587-2587, 2019.
[94] M. Ashraf, A. Souka, and R. Singh, "Central retinal vein occlusion: modifying current treatment protocols," *Eye,* vol. 30, no. 4, pp. 505-514, 2016.

[95] P. A. Campochiaro *et al.*, "Long-term outcomes in patients with retinal vein occlusion treated with ranibizumab: the RETAIN study," *Ophthalmology,* vol. 121, no. 1, pp. 209-219, 2014.
[96] J. S. Heier *et al.*, "Ranibizumab for macular edema due to retinal vein occlusions: long-term follow-up in the HORIZON trial," *Ophthalmology,* vol. 119, no. 4, pp. 802-809, 2012.
[97] G. M. Noh, "Choroidal Thickness Changes as a Predictor of Visual Acuity Improvement in Retinal Vein Occlusion Patients," *Investigative Ophthalmology & Visual Science,* vol. 60, no. 9, pp. 2595-2595, 2019.
[98] H. J. Hogg, S. Di Simplicio, and M. S. Pearce, "Ranibizumab and aflibercept intravitreal injection for treatment naïve and refractory macular oedema in branch retinal vein occlusion," *European Journal of Ophthalmology,* vol. 31, no. 2, pp. 548-555, 2021.
[99] T. Hasegawa, T. Ueda, M. Okamoto, and N. Ogata, "Presence of foveal bulge in optical coherence tomographic images in eyes with macular edema associated with branch retinal vein occlusion," *American Journal of Ophthalmology,* vol. 157, no. 2, pp. 390-396. e1, 2014.
[100] Y. Muraoka *et al.*, "Morphologic and functional changes in retinal vessels associated with branch retinal vein occlusion," *Ophthalmology,* vol. 120, no. 1, pp. 91-99, 2013.
[101] J. Yerushalmy, "Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques," *Public Health Reports (1896-1970),* pp. 1432-1449, 1947.
[102] C. S. Lee, D. M. Baughman, and A. Y. Lee, "Deep learning is effective for classifying normal versus age-related macular degeneration OCT images," *Ophthalmology Retina,* vol. 1, no. 4, pp. 322-327, 2017.
[103] H. Zhang, Z. Chen, Z. Chi, and H. Fu, "Hierarchical local binary pattern for branch retinal vein occlusion recognition with fluorescein angiography images," *Electronics Letters,* vol. 50, no. 25, pp. 1902-1904, 2014.
[104] Z. Fazekas, A. Hajdu, I. Lázár, G. Kovács, and B. Csákány, "Influence of using different segmentation methods on the fractal properties of the identified retinal vascular networks in healthy retinas and in retinas with vein occlusion," 2015.
[105] R. Zhao, Z. Chen, and Z. Chi, "Convolutional neural networks for branch retinal vein occlusion recognition?," in *2015 IEEE International Conference on Information and Automation*, 2015: IEEE, pp. 1633-1636.
[106] M. Niemeijer, B. Van Ginneken, J. Staal, M. S. Suttorp-Schulten, and M. D. Abràmoff, "Automatic detection of red lesions in digital color fundus photographs," *IEEE Transactions on medical imaging,* vol. 24, no. 5, pp. 584-592, 2005.
[107] C.-C. J. Kuo, "Understanding convolutional neural networks with a mathematical model," *Journal of Visual Communication and Image Representation,* vol. 41, pp. 406-413, 2016.
[108] J. Y. Choi, T. K. Yoo, J. G. Seo, J. Kwak, T. T. Um, and T. H. Rim, "Multi-categorical deep learning neural network to classify retinal images: A pilot study employing small database," *PloS one,* vol. 12, no. 11, p. e0187336, 2017.
[109] B. Choudhury, "Retinal Image Analysis and Diagnosis of Retinal Blood Vascular Diseases using Deep Learning Model," Swinburne University of Technology Sarawak Campus, 2019.
[110] R. Rasti *et al.*, "Deep learning-based single-shot prediction of differential effects of anti-VEGF treatment in patients with diabetic macular edema," *Biomedical Optics Express,* vol. 11, no. 2, pp. 1139-1152, 2020.
[111] D. Feng *et al.*, "A preliminary study of predicting effectiveness of anti-VEGF injection using OCT images based on deep learning," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020: IEEE, pp. 5428-5431.

[112] H. Bogunović *et al.*, "Prediction of anti-VEGF treatment requirements in neovascular AMD using a machine learning approach," *Investigative Ophthalmology & Visual Science,* vol. 58, no. 7, pp. 3240-3248, 2017.
[113] S. S. Kar, D. D. Sevgi, V. Dong, S. K. Srivastava, A. Madabhushi, and J. P. Ehlers, "Multi-compartment spatially-derived radiomics from optical coherence tomography predict anti-VEGF treatment durability in macular edema secondary to retinal vascular disease: preliminary findings," *IEEE Journal of Translational Engineering in Health and Medicine,* vol. 9, pp. 1-13, 2021.
[114] P. Prahs *et al.*, "OCT-based deep learning algorithm for the evaluation of treatment indication with anti-vascular endothelial growth factor medications," *Graefe's Archive for Clinical and Experimental Ophthalmology,* vol. 256, no. 1, pp. 91-98, 2018.
[115] S. M. Khan *et al.*, "A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability," *The Lancet Digital Health,* vol. 3, no. 1, pp. e51-e66, 2021.
[116] P. Gholami, P. Roy, M. K. Parthasarathy, and V. Lakshminarayanan, "OCTID: Optical coherence tomography image database," *Computers & Electrical Engineering,* vol. 81, p. 106532, 2020.
[117] J. J. Balaji and V. Lakshminarayanan, "ODTiD: Optic Nerve Head SD-OCT Image Dataset," *Clinical Ophthalmology (Auckland, NZ),* vol. 15, p. 4239, 2021.
[118] S. Morales *et al.*, "Retinal layer segmentation in rodent OCT images: Local intensity profiles & fully convolutional neural networks," *Computer Methods and Programs in Biomedicine,* vol. 198, p. 105788, 2021.
[119] M. Golabbakhsh and H. Rabbani, "Vessel-based registration of fundus and optical coherence tomography projection images of retina using a quadratic registration model," *IET Image Processing,* vol. 7, no. 8, pp. 768-776, 2013.
[120] G. C. Bedke, M. E. Jadhav, P. Punde, and S. Dongaonkar, "Retinal OCT images for glaucoma," in *2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC)*, 2020: IEEE, pp. 6-8.
[121] Y. Ma *et al.*, "ROSE: a retinal OCT-angiography vessel segmentation dataset and new model," *IEEE transactions on medical imaging,* vol. 40, no. 3, pp. 928-939, 2020.
[122] S. J. Kim, K. J. Cho, and S. Oh, "Development of machine learning models for diagnosis of glaucoma," *PloS one,* vol. 12, no. 5, p. e0177726, 2017.
[123] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang, "A cross-modality learning approach for vessel segmentation in retinal images," *IEEE transactions on medical imaging,* vol. 35, no. 1, pp. 109-118, 2015.
[124] D. S. W. Ting *et al.*, "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes," *Jama,* vol. 318, no. 22, pp. 2211-2223, 2017.
[125] M. H. Sarhan *et al.*, "Machine learning techniques for ophthalmic data processing: a review," *IEEE Journal of Biomedical and Health Informatics,* vol. 24, no. 12, pp. 3338-3350, 2020.
[126] X. Su, X. Yan, and C. L. Tsai, "Linear regression," *Wiley Interdisciplinary Reviews: Computational Statistics,* vol. 4, no. 3, pp. 275-294, 2012.
[127] G. A. Seber and A. J. Lee, *Linear regression analysis*. John Wiley & Sons, 2012.
[128] B. G. Tabachnick, L. S. Fidell, and J. B. Ullman, *Using multivariate statistics*. pearson Boston, MA, 2007.
[129] C. Dismuke and R. Lindrooth, "Ordinary least squares," *Methods and Designs for Outcomes Research,* vol. 93, pp. 93-104, 2006.
[130] X. Yan and X. Su, *Linear regression analysis: theory and computing*. world scientific, 2009.
[131] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis*. John Wiley & Sons, 2021.

[132] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science,* vol. 349, no. 6245, pp. 255-260, 2015.
[133] Z.-H. Zhou, *Machine learning*. Springer Nature, 2021.
[134] T.-H. Lee, A. Ullah, and R. Wang, "Bootstrap aggregating and random forest," in *Macroeconomic forecasting in the era of big data*: Springer, 2020, pp. 389-429.
[135] L. Breiman, "Random forests," *Machine learning,* vol. 45, no. 1, pp. 5-32, 2001.
[136] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R news,* vol. 2, no. 3, pp. 18-22, 2002.
[137] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *the Journal of machine Learning research,* vol. 12, pp. 2825-2830, 2011.
[138] A. H. Murphy, "Skill scores based on the mean square error and their relationships to the correlation coefficient," *Monthly weather review,* vol. 116, no. 12, pp. 2417-2424, 1988.
[139] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate research,* vol. 30, no. 1, pp. 79-82, 2005.
[140] L. Breiman, "Random forests," *Machine learning,* vol. 45, pp. 5-32, 2001.
[141] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU press, 2013.
[142] D. R. Cutler *et al.*, "Random forests for classification in ecology," *Ecology,* vol. 88, no. 11, pp. 2783-2792, 2007.
[143] R. W. Hoerl, "Ridge regression: a historical context," *Technometrics,* vol. 62, no. 4, pp. 420-425, 2020.
[144] C. Hans, "Bayesian lasso regression," *Biometrika,* vol. 96, no. 4, pp. 835-845, 2009.
[145] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological),* vol. 58, no. 1, pp. 267-288, 1996.
[146] J. Gaudart, B. Giusiano, and L. Huiart, "Comparison of the performance of multi-layer perceptron and linear regression for epidemiological data," *Computational statistics & data analysis,* vol. 44, no. 4, pp. 547-570, 2004.
[147] H. Taud and J. Mas, "Multilayer perceptron (MLP)," in *Geomatic approaches for modeling land change scenarios*: Springer, 2018, pp. 451-455.
[148] H. J. MacFie and D. Hedderley, "Current practice in relating sensory perception to instrumental measurements," *Food quality and preference,* vol. 4, no. 1-2, pp. 41-49, 1993.
[149] P. Toscas, F. Shaw, and S. Beilken, "Partial least squares (PLS) regression for the analysis of instrument measurements and sensory meat quality data," *Meat science,* vol. 52, no. 2, pp. 173-178, 1999.
[150] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and intelligent laboratory systems,* vol. 58, no. 2, pp. 109-130, 2001.
[151] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, 2005, vol. 1: Ieee, pp. 886-893.
[152] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition,* vol. 29, no. 1, pp. 51-59, 1996.
[153] T. Ojala and M. Pietikäinen, "Unsupervised texture segmentation using feature distributions," *Pattern recognition,* vol. 32, no. 3, pp. 477-486, 1999.
[154] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 24, no. 7, pp. 971-987, 2002.
[155] I. Terol-Villalobos and P. Hawkes, "Morphological image enhancement and segmentation with analysis," *New York: Academic,* pp. 207-273, 2005.
[156] J. Serra, "Mathematical Morphology Vol. I. London," *UK: Academic,* 1982.

[157] P. Soille, *Morphological image analysis: principles and applications* (no. 3). Springer, 1999.
[158] E. R. Dougherty and R. A. Lotufo, *Hands-on morphological image processing*. SPIE press, 2003.
[159] W. Ye and K.-K. Ma, "Blurriness-guided unsharp masking," *IEEE Transactions on Image Processing,* vol. 27, no. 9, pp. 4465-4477, 2018.
[160] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics,* no. 6, pp. 610-621, 1973.
[161] M. O'Byrne, B. Ghosh, V. Pakrashi, and F. Schoefs, "Texture analysis based detection and classification of surface features on ageing infrastructure elements," in *BCRI2012 Bridge & Concrete Research in Ireland*, 2012.
[162] S. Singh, D. Srivastava, and S. Agarwal, "GLCM and its application in pattern recognition," in *2017 5th International Symposium on Computational and Business Intelligence (ISCBI)*, 2017: IEEE, pp. 20-25.
[163] F. Chollet, *Deep learning with Python*. Simon and Schuster, 2021.
[164] A. Ameri, M. A. Akhaee, E. Scheme, and K. Englehart, "Regression convolutional neural network for improved simultaneous EMG control," *Journal of neural engineering,* vol. 16, no. 3, p. 036015, 2019.
[165] X. Lei, H. Pan, and X. Huang, "A dilated CNN model for image classification," *IEEE Access,* vol. 7, pp. 124087-124095, 2019.
[166] R. Firoz, M. S. Ali, M. N. U. Khan, M. K. Hossain, M. K. Islam, and M. Shahinuzzaman, "Medical image enhancement using morphological transformation," *Journal of Data Analysis and Information Processing,* vol. 4, no. 1, pp. 1-12, 2016.
[167] H. Sadreazami, M. Bolic, and S. Rajan, "Contactless fall detection using time-frequency analysis and convolutional neural networks," *IEEE Transactions on Industrial Informatics,* vol. 17, no. 10, pp. 6842-6851, 2021.
[168] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Gray scale and rotation invariant texture classification with local binary patterns," in *European conference on computer vision*, 2000: Springer, pp. 404-420.
[169] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
[170] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature,* vol. 521, no. 7553, pp. 436-444, 2015/05/01 2015, doi: 10.1038/nature14539.
[171] X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in *2009 IEEE 12th international conference on computer vision*, 2009: IEEE, pp. 32-39.
[172] Y. Zhang *et al.*, "Prediction of Visual Acuity after anti-VEGF Therapy in Diabetic Macular Edema by Machine Learning," *Journal of Diabetes Research,* vol. 2022, 2022.
[173] T.-E. Tan, T. Y. Wong, and D. S. W. Ting, "Artificial Intelligence for Prediction of Anti–VEGF Treatment Burden in Retinal Diseases: Towards Precision Medicine," *Ophthalmology Retina,* vol. 5, no. 7, pp. 601-603, 2021.
[174] D. R. Cox, "The regression analysis of binary sequences," *Journal of the Royal Statistical Society: Series B (Methodological),* vol. 20, no. 2, pp. 215-232, 1958.
[175] T. G. Nick and K. M. Campbell, "Logistic regression," *Topics in biostatistics,* pp. 273-301, 2007.
[176] L. Breiman, "Bagging predictors," *Machine learning,* vol. 24, no. 2, pp. 123-140, 1996.
[177] A. Parmar, R. Katariya, and V. Patel, "A review on random forest: An ensemble classifier," in *International Conference on Intelligent Data Communication Technologies and Internet of Things*, 2018: Springer, pp. 758-763.
[178] B. P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor, "Boosted decision trees as an alternative to artificial neural networks for particle identification," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 543, no. 2-3, pp. 577-584, 2005.

[179] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 161-168.
[180] Q. Wu, C. J. Burges, K. M. Svore, and J. Gao, "Adapting boosting for information retrieval measures," *Information Retrieval,* vol. 13, no. 3, pp. 254-270, 2010.
[181] A. V. Dorogush, V. Ershov, and A. Gulin, "CatBoost: gradient boosting with categorical features support," *arXiv preprint arXiv:1810.11363,* 2018.
[182] A. A. Ibrahim, R. L. Ridwan, M. M. Muhammed, R. O. Abdulaziz, and G. A. Saheed, "Comparison of the CatBoost classifier with other machine learning methods," *International Journal of Advanced Computer Science and Applications,* vol. 11, no. 11, 2020.
[183] R. E. Wright, "Logistic regression," 1995.
[184] M. Pal, "Random forest classifier for remote sensing classification," *International journal of remote sensing,* vol. 26, no. 1, pp. 217-222, 2005.