



Faculty of Cognitive Sciences and Human Development

**Automated Detection and Classification of Retinal Vein Occlusion Using
Ultra-widefield Retinal Fundus Images and Transfer Learning**

Ivy Ong Siaw Yin

**Master of Science
2024**

Automated Detection and Classification of Retinal Vein Occlusion Using Ultra-widefield Retinal Fundus Images and Transfer Learning

Ivy Ong Siaw Yin

A thesis submitted

In fulfillment of the requirements for the degree of Master of Science

(Intelligent Systems)

Faculty of Cognitive Sciences and Human Development

UNIVERSITI MALAYSIA SARAWAK

2024

DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Malaysia Sarawak. Except where due acknowledgements have been made, the work is that of the author alone. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.



.....

Signature

Name: Ivy Ong Siaw Yin

Matric No.: 21020483

Faculty of Cognitive Sciences and Human Development

Universiti Malaysia Sarawak

Date : 16 September 2024

ACKNOWLEDGEMENT

I am grateful to the Almighty God for guiding me as I completed my thesis. Completing my thesis would not have been possible without the guidance, knowledge, and patience of my respected supervisor, Assoc. Prof. Dr. Lee Nung Kion. I want to thank the financial support from the Ministry of Higher Education Malaysia, Fundamental Research Grant Scheme FRGS/1/2020/ICT02/UNIMAS/01/1. Additionally, I would like to thank Prof. Dr. Lim Lik Thai, Assoc. Prof. Dr. Muhammad Hamdi bin Mahmood, Assoc. Prof. Dr. Puppala Sai Perumalla Ravindra, and Dr. Eva Ong for providing the materials and help needed to carry out the research.

ABSTRACT

Retinal vein occlusion (RVO) is a retinal disease resulting from blockage of the retinal veins, leading to the development of haemorrhages and cotton wool spots within the retina. Retinal fundus imaging is essential for diagnosing these abnormalities. The study explores the potential of artificial intelligence (AI) and machine learning (ML) to automate the analysis of these images for detecting RVO signs. Detecting RVO signs in retinal images is challenging due to the irregular shapes, sizes, and locations of cotton wool spots and haemorrhages. Retinal images also include various anatomical structures and may suffer from poor quality, introducing artefacts. Manually segmentation is labour-intensive, requiring significant expertise and time. Moreover, existing models for retinal disease classification lack transparency, providing only classification labels without explanations or visualisations, raising concerns about their reliability. Additionally, the effectiveness of segmenting RVO signs in ultra-widefield (UWF) fundus images is uncertain due to a lack of studies and benchmarking. The study aims to develop automated methods for the segmentation and classification of RVO in UWF fundus images. It seeks to improve the interpretability and reliability of automated retinal disease screening by visualising pertinent disease features and examining the effectiveness of segmentation models trained on regular fundus datasets for UWF images. The methodology involves training segmentation models using both regular and UWF fundus images. The study also proposes using transfer learning, where a model pre-trained on a regular fundus dataset is fine-tuned with a UWF dataset to improve segmentation performance. The approach seeks to utilise knowledge from pretraining to enhance the performance of the segmentation model. The study also evaluates the classification model trained with lesion masks to classify images accurately into the respective categories. The automated algorithm can detect the presence of RVO in retinal

fundus images and accurately pinpoint the locations of associated lesions. This improves the reliability and interpretability of automated retinal disease screening, aiding clinicians in making informed decisions. The study showcases the effectiveness of deep neural networks for feature extraction and classification, underscoring the potential of AI in medical imaging.

Keywords: Retinal vein occlusion disease, image segmentation, image classification, artificial intelligence, deep learning

Pengesanan dan Klasifikasi Oklusi Vena Retina Automatik Menggunakan Imej Fundus Retina Medan Ultra Lebar dan Pembelajaran Pemindahan

ABSTRAK

Oklusi urat retina (RVO) adalah penyakit retina yang berpunca daripada penyumbatan urat retina yang menyebabkan pembentukan pendarahan dan bintik kapas dalam retina. Imej fundus retina memainkan peranan penting dalam diagnosis tanda-tanda RVO. Kajian ini meneroka potensi kecerdasan buatan (AI) dan pembelajaran mesin (ML) untuk membuat analisis imej-imej ini bagi mengesan tanda-tanda RVO secara automatik. Cabaran untuk mengesan tanda-tanda RVO, iaitu bintik bulu kapas dan pendarahan, termasuk bentuk dan saiz yang tidak tertentu, serta perbezaan dalam lokasi tanda-tanda RVO. Imej fundus retina juga merangkumi struktur anatomi dan kualiti imej yang rendah boleh memperkenalkan artefak. Segmentasi secara manual memerlukan tenaga kerja yang intensif, memerlukan kepakaran dan masa yang panjang. Model klasifikasi penyakit retina sedia ada juga kurang ketelusan kerana hanya mengeluarkan label klasifikasi tanpa menawarkan sebarang penjelasan atau visualisasi tanda penyakit dalam imej. Keterbatasan ini menimbulkan kebimbangan mengenai kebolehpercayaan saringan penyakit retina automatik.. Selain itu, keberkesanan segmentasi tanda-tanda RVO dalam imej fundus ultra-widefield (UWF) tidak dikenal pasti kerana kekurangan kajian dan penanda aras. Kajian ini bertujuan untuk memperkenalkan kaedah automatik untuk segmentasi dan klasifikasi RVO dalam imej fundus UWF. Ia bertujuan untuk meningkatkan kebolehfahaman dan kebolehpercayaan penyaringan penyakit retina automatik dengan menggambarkan tanda-tanda penyakit yang berkaitan dan mengkaji keberkesanan model segmentasi yang dilatih pada set data fundus biasa untuk imej UWF. Metodologi ini melibatkan latihan model segmentasi menggunakan imej fundus biasa dan UWF. Kajian ini juga mencadangkan penggunaan pembelajaran

pemindahan, di mana model yang telah dilatih terlebih dahulu pada set data fundus biasa diselaraskan dengan set data UWF untuk meningkatkan prestasi segmentasi. Pendekatan ini bertujuan untuk menggunakan pengetahuan daripada latihan awal untuk meningkatkan prestasi model segmentasi. Kajian ini juga menilai model klasifikasi yang dilatih dengan tanda-tanda RVO untuk mengklasifikasikan imej dengan tepat ke dalam kategori yang sesuai. Algoritma automatik ini boleh mengesan RVO dalam imej fundus retina dan mengenal pasti lokasi lesi yang berkaitan dengan tepat. Ini meningkatkan kebolehpercayaan dan kebolehfahaman penyaringan penyakit retina automatik, membantu pakar klinik membuat keputusan yang tepat. Kajian ini menunjukkan keberkesanan rangkaian neural mendalam untuk pengekstrakan ciri dan klasifikasi, menekankan potensi AI dalam pengimejan perubatan.

Kata kunci: *Penyakit oklusi urat retina, segmentasi imej, klasifikasi imej, kepintaran buatan*

TABLE OF CONTENTS

	Page
DECLARATION	i
ACKNOWLEDGEMENT	ii
ABSTRACT	iii
ABSTRAK	v
TABLE OF CONTENTS	vii
LIST OF TABLES	xi
LIST OF FIGURES	xiii
LIST OF ABBREVIATIONS	xv
CHAPTER 1 INTRODUCTION	1
1.1 Overview	1
1.2 Background of the Study	2
1.3 Retinal Vein Occlusion	5
1.4 Problem Statements	10
1.5 Research Questions	14
1.6 Objective	15
1.6.1 Specific Objectives	15
1.7 Contributions	16

1.8	Definition of Terms	16
1.8.1	Conceptual Definitions	16
1.8.2	Operational Definitions	17
1.9	Conclusion	17
CHAPTER 2 LITERATURE REVIEW		18
2.1	Overview	18
2.2	Computer Vision	19
2.3	Deep Learning	20
2.4	Convolutional Neural Networks (CNN)	21
2.5	Literature Review	25
2.5.1	Multi-class Diabetic Retinopathy Classification with Convolutional Neural Network	26
2.5.2	Diabetic Retinopathy Detection with Deep Convolutional Network and Transfer Learning	28
2.5.3	Diabetic Retinopathy Detection with Deep Convolutional Network	28
2.5.4	Central Retinal Vein Occlusion with CNN	31
2.5.5	Branch Retinal Vein Occlusion Detection with Deep Learning	33
2.5.6	Retinal Vein Occlusion Diagnosis and Segmentation with Deep Learning	34
2.5.7	U-Net: Convolutional Networks for Biomedical Segmentation	36
2.5.8	Lesions Segmentation with Improved U-Net Model (EAD-Net)	39
2.5.9	Diabetic Retinopathy Grading System and Multiple Retinal Lesions Detection	45

2.5.10	U-Net Based InceptionV3 Model for Diabetic Retinopathy Segmentation	49
2.6	Discussion	51
2.7	Conclusion	61
CHAPTER 3 METHODOLOGY		62
3.1	Overview	62
3.2	Research Flow Chart	62
3.3	Software	65
3.4	Image Acquisition and Verification	66
3.5	Image Pre-processing	68
3.5.1	Image Pre-processing for the Segmentation Model Trained Using Regular Dataset	68
3.5.2	Image Pre-processing for the Segmentation Model Trained Solely Using UWF Dataset	70
3.5.3	Image Pre-processing for the Segmentation Model Trained Using Transfer Learning	72
3.5.4	Image Pre-processing for the Classification Model	73
3.6	Image Post-processing	73
3.7	Data Augmentation	74
3.8	Implementation	74
3.9	U-Net Architecture	76
3.10	CNN Architecture	78

3.11	Performance Evaluation	79
CHAPTER 4 RESULTS AND DISCUSSION		82
4.1	Introduction	82
4.2	Results of the Segmentation Model Trained Using Regular Fundus Images	82
4.3	Results of the Segmentation Model Trained Using UWF Fundus Images	85
4.4	Results of the Segmentation Model Using Transfer Learning Approach	88
4.5	Discussion and Comparison of the Segmentation Models of this Study	90
4.6	Discussion and Comparison of Segmentation Model with Existing Methods	94
4.7	Results of the Classification Model	99
4.8	Discussion and Comparison of the Classification Model with Existing Methods	100
CHAPTER 5 CONCLUSION		102
5.1	Main Findings	102
5.2	Research Contributions	103
5.3	Limitations and Suggestions	104
REFERENCES		106
APPENDICES		115

LIST OF TABLES

	Page
Table 1.1: Characteristics of RVO	10
Table 2.1: The architecture of the models	27
Table 2.2: Deep learning and feature extraction methods used	29
Table 2.3: The architecture of the CNN model	30
Table 2.4: The architecture of the proposed CNN model	32
Table 2.5: Architecture of VGG16	34
Table 2.6: Results of VGG16 and SVM	34
Table 2.7: Statistics of the dataset	36
Table 2.8: Performance of the models on RVO classification	36
Table 2.9: Performance of the models on lesion segmentation	36
Table 2.10: Hyperparameters of the network used in Wan et al. (2021)	41
Table 2.11: Performance of the proposed method on the e_ophtha_EX dataset	42
Table 2.12: Performance of the proposed method on local intelligent ophthalmology dataset	42
Table 2.13: Performance of the proposed method on the IDRiD dataset	43
Table 2.14: Overall steps of the proposed method	44
Table 2.15: Performance comparison between EAD-Net and U-Net	44
Table 2.16: Parameters of the proposed method by Abdelmaksoud et al. (2021)	47
Table 2.17: Results of binary classification on datasets	48
Table 2.18: Performance of segmentation of lesions on IDRiD and DIARETDB1 datasets	49
Table 2.19: Results of SVM on grading DR on segmented images	49
Table 2.20: Evaluation metrics for IDRiD dataset	51
Table 2.21: Evaluation metrics for the DIARETDB1 dataset	51

Table 3.1:	The architecture of the proposed CNN model	79
Table 4.1:	Validation result of the segmentation model with different numbers of convolution blocks and filters	83
Table 4.2:	Evaluation results of segmentation model trained using a regular dataset on UWF testing images	84
Table 4.3:	Validation result of the segmentation model with different numbers of convolution blocks and filters	86
Table 4.4:	Hyper-parameter of the U-Net model	86
Table 4.5:	Evaluation results of segmentation model trained using UWF dataset on UWF testing images	87
Table 4.6:	Evaluation results of segmentation model trained using transfer learning on UWF testing images	88
Table 4.7:	Comparison of the segmentation models of this study	92
Table 4.8:	Comparison of evaluation results for segmentation model with existing methods	97
Table 4.9:	Comparison of the segmentation model with Ronneberger et al.	99
Table 4.10:	Parameters of the CNN model	100
Table 4.11:	Evaluation results on the classification of RVO and non-RVO on the UWF testing set	100
Table 4.12:	Comparison of evaluation result for classification model with existing methods	101

LIST OF FIGURES

	Page
Figure 1.1: Retinal fundus image of CRVO with retinal haemorrhage. The retinal fundus image shows CRVO with scattered retinal haemorrhages. (Retina Gallery, 2014)	6
Figure 1.2: Fundus image of the right eye with CRVO. Appearance of substantial intraretinal haemorrhaging, flame haemorrhages, and optic nerve oedema associated with CRVO. (Ip & Hendrick, 2019)	7
Figure 1.3: Retinal fundus image of the cotton wool spot. The image shows a retinal fundus image with a cotton wool spot (circled in yellow). (Glycosmedia, n.d)	8
Figure 1.4: Retinal fundus image with tortuous vessels. The retinal fundus image shows tortuous vessels (black arrow). (Pachade et al., 2021)	8
Figure 1.5: Retinal fundus image of a healthy patient, without retinal vein occlusion.	9
Figure 1.6: Regular fundus image	12
Figure 1.7: Ultra-widefield fundus image	13
Figure 2.1: CNN image classification pipeline. (Rawat & Wang, 2017)	22
Figure 2.2: An example of a 2-D convolution.(Wang & Yang, 2019)	23
Figure 2.3: U-Net architecture. The blue box represents a multi-channel feature map with its number of channels on top of the box. The number at the lower left edge of the box is the x-y size. The white box represents copied feature maps. The arrows signify different operations.	38
Figure 2.4: Segmentation of lesions. The result of the segmentation of microaneurysm (red colour), haemorrhages (green), hard exudates (blue), and soft exudates (yellow). (Wan et al., 2021)	40
Figure 3.1: Research flow chart	63
Figure 3.2: Retinal vein occlusion detection model flow chart	65
Figure 3.3: An UWF image with annotated haemorrhage by an expert. The regions of haemorrhage are marked in red colour.	67
Figure 3.4: An UWF image reviewed by the second expert. The yellow circles are marked by the second expert to indicate annotations that are missed.	68

Figure 3.5: Pre-processed regular fundus image with cotton wool spot. a) Original image b) Pre-processed image	69
Figure 3.6: Pre-processed regular fundus image with haemorrhage. a) Original image b) Pre-processed image	70
Figure 3.7: Overall pre-processing steps of UWF retinal fundus images with cotton wool spots. a) Original image b) Generated ellipse mask c) Subtracted image d) Resized and rescaled image	72
Figure 3.8: U-Net architecture. U-Net architecture with input image resolution of 512 x 448 x 1. The blue box represents a feature map with its number of filters on top of the box. The black dotted box represents copied feature maps. The coloured arrows signify different operations.	76
Figure 4.1: Segmented haemorrhages by the model trained using a regular dataset on testing UWF image compared to the actual ground truth mask	84
Figure 4.2: Segmented cotton wool spots by the model trained using a regular dataset on a testing UWF image	85
Figure 4.3: Segmented haemorrhages by the model trained using the UWF dataset on a testing UWF image	87
Figure 4.4: Segmented cotton wool spots by the model trained using the UWF dataset on a testing UWF image	88
Figure 4.5: Segmented haemorrhages by the model trained using transfer learning on a testing UWF image	89
Figure 4.6: Segmented cotton wool spots by the model trained using transfer learning on a testing UWF image	90
Figure 4.7: Comparison of the IoU score of the segmentation methods	92
Figure 4.8: Predicted and actual ground truth of haemorrhages for a UWF image by the model trained using the UWF dataset	93
Figure 4.9: Predicted and actual ground truth of cotton wool spots for an UWF image by the model using the UWF dataset	93

LIST OF ABBREVIATIONS

AUPR	Area Under Precision-Recall Curve
ANN	Artificial Neural Networks
BP	Bifurcation Points
BRVO	Branch Retinal Vein Occlusion
CRVO	Central Retinal Vein Occlusion
CNN	Convolutional Neural Networks
DR	Diabetic Retinopathy
DIARETDB	Diabetic Retinopathy Database
DRIVE	Digital Retinal Images for Vessel Extraction
EM	Electron Microscopic
FOV	Field-of-view
GB	Gigabytes
GPUs	Graphical Processing Units
GLCM	Grey Level Co-occurrence Matrix
Google Colab	Google Colaboratory
HEs	Haemorrhages
HRF	High-Resolution Fundus
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IDRID	Indian Diabetic Retinopathy Image Dataset
ISBI	International Symposium on Biomedical Imaging
IoU	Intersection over Union
MAs	Microaneurysm
RAM	Random Access Memory

ReLU	Rectified Linear Units
ROI	Region of Interest
ResNet-50	Residual Network-50
RVO	Retinal Vein Occlusion
ssTEM	Serial Section Transmission Electron Microscopy
STARE	Structured analysis of the Retina
UWF	Ultra-widefield

CHAPTER 1

INTRODUCTION

1.1 Overview

In the vast field of healthcare, to name a few branches of medicine, there are ophthalmology, cardiology, pathology, and oncology where each branch of medicine is focused on the prevention, diagnosis, and treatments of the respective branch. There will always be discoveries and advances in healthcare, whether it be new ways to improve diagnosis and treatments. All the efforts are to strive for the prevention of disease, relief of pain and suffering of disease, and maintenance of health, that is essentially what medicine is (Encyclopædia Britannica, 2020).

Besides that, as artificial intelligence develops, its application in healthcare has been widely studied to find ways to solve real-world problems and endlessly explore a better way to solve a problem. From there, many studies showed plausible results in solving problems such as classification, segmentation, risk prediction, anomaly detection and low medical image quality. For example, there are studies on detection of exudates, haemorrhages, and microaneurysm (Khojasteh et al., 2018), image enhancement method for degraded fundus images (Jin et al., 2018), coronary artery calcium score prediction (Son et al., 2020), and determining the severity of diabetic retinopathy and grading macular oedema (Sahlsten et al., 2019).

Furthermore, some of these studies involve deep learning which has also been emerging in artificial intelligence, it is about learning patterns and classifying images, text, or sounds from large input data with a network of many processing layers (LeCun et al.,

2015). Generally, these are all possible with large datasets as artificial intelligence is about creating machines that can mimic human intelligence, and its performance relies on the input data. For application in ophthalmology, it might require medical imaging as input data which can be fundus images, optical coherence, tomography, and ultrasonography (Flaxel et al., 2020).

Ophthalmology is a branch of medicine that is related to diseases and disorders of the eyes. Application of artificial intelligence in ophthalmological diseases that have been the most delved into are glaucoma, cataract, diabetic retinopathy, and age-related macular degeneration, while there have been fewer than 10 papers on retinal vein obstruction, retinal detachment, and papilledema (Lu et al., 2018). This study is about the application of artificial intelligence and deep learning to detect anomalies in retinal vein occlusion using retinal fundus images. This chapter covers the background of the study, statement of the problem, objectives of the study and contributions of the study.

1.2 Background of the Study

A type of retinopathy or retinal disease known as retinal vein occlusion (RVO) is caused by a blockage in the retinal vein as a result of hardening and blood clotting. The types of retinal vein occlusion are central retinal vein occlusion (CRVO), and branch retinal vein occlusion (BRVO) (Ip & Hendrick, 2019). As the name suggests, RVO refers to occlusion in different regions of the retina, whether in the main retinal vein within or posterior to the optic nerve head, branches of the central retinal vein (within a tributary), or at major bifurcation. Disrupted blood flow may cause conditions that will affect the retina such as macular oedema, optic neuropathy, or vitreous haemorrhage (Ip & Hendrick, 2019). Hence, RVO could lead to some degree of sudden vision loss or blindness and impairments.

The common signs of RVO are haemorrhages all through the retina as well as the retinal veins appearing dilated and tortuous which can be seen with retinal imaging (Mehta, 2020). Retinal haemorrhages happen when there is an occlusion in the blood vessels of the eye and blood leaks out from the site of occlusion. Cotton wool spots can also be present because lipids leak out due to a lack of oxygen supply to tissues resulting in damaged tissue (Kipli et al., 2018). Early diagnosis is crucial to prevent the condition of the patient from worsening where eventually the damage could be irreversible, that is blindness or some degree of vision impairment.

The Gutenberg Health Study found that 91.5% of patients with RVO had at least one cardiovascular risk factor (Ponto et al., 2015). This shows that many factors can increase the risk of developing RVO. These cardiovascular risk factors are also commonly the concerns that people are aware of as it is the cause of many premature deaths around the world which means it is quite prevalent. Thus, people should also be educated about RVO, its effect, prevention, and cure.

Besides that, retinal imaging has been relied on to give more insights into the abnormalities in the retina such as cotton wool spots, and haemorrhages that may help experts with the diagnosis and treatments. In addition, the application of artificial intelligence, specifically automated systems and machine learning algorithms is capable of analysing abnormalities of the retina with retinal images. Even with just image pre-processing techniques, it can significantly improve the quality of retinal imaging.

Anomaly detection has been applied and useful in areas such as cybersecurity, healthcare, finance, and science. Anomaly detection is about detecting any variations from the norms and the algorithms need to be designed specifically for each problem (Mehrotra

et al., 2017). In the healthcare field, anomaly detection algorithms have been used for patient monitoring and disease diagnosis using quantitative data and radiographic images. For example, monitoring premature ventricular contraction using electrocardiogram data and classifying benign or malignant tumours using radiographic images. This can potentially assist in early detection and help patients receive immediate care. There are many other applications of anomaly detection algorithms that are beneficial, and this research intends to discover its application in detecting abnormal features in retinal fundus images for retinal vein occlusion disease.

Image classification is one of the tasks in computer vision that involves analysing visual data, such as images or videos, and finding patterns that help to categorise the data. The goal is to make machines interpret visual data in a way that is similar to humans. The traditional approach for image classification is by using handcrafted features extracted from images before applying them to machine learning algorithms. Different feature extraction techniques can extract different types of features such as colour, texture, shape, and conceptual features. Some of the techniques used for extracting these features are grey-level co-occurrence matrices, Gabor wavelet transform, and Histogram of oriented gradients (Salau & Jain, 2019). Then, traditional machine learning algorithms such as support vector machines, random forests, and decision trees are trained using the extracted features. The constraints of traditional machine learning stem from their reliance on handcrafted feature extraction, which is time-consuming, labour-intensive, and limited in scope. The extracted features may not capture the relevant and discriminative information needed especially in a large dataset (Janiesch et al., 2021).

The deep learning approach can surpass the constraints of the traditional machine learning approach for image classification tasks by autonomously extracting complex features from input data. It uses a hierarchical approach when extracting feature maps of simpler or low-level features followed by high-level features which are combined to capture complex patterns of the input data. Some of the deep learning algorithms are convolutional neural networks (CNN), generative adversarial neural networks, and recurrent neural networks (Janiesch et al., 2021). There are also limitations of using deep learning techniques in image classification which are requiring a large quantity of data, requiring powerful hardware such as graphical processing units and memory, and difficulty of generalizing to datasets of different sources aside from the training dataset (Ahmad et al., 2023; Sarker, 2021).

1.3 Retinal Vein Occlusion

Retinal vein occlusion (RVO) is a type of retinopathy or retinal disease that is caused by partially or fully blocked retinal veins (Flaxel et al., 2020). As the name suggests, the type of occlusion of the retinal vein is determined by the regions of the retina where there is obstruction. The site of occlusion is either at the central or branch of the retinal vein and they correspond to central retinal vein occlusion (CRVO) and branch retinal vein occlusion (BRVO). RVO could result in some degree of sudden vision loss and impairments due to retinal veins hardening and blood clotting (thrombosis) (Sivaprasad et al., 2015), whether in the main retinal vein that is within or posterior to the optic nerve head, or within a tributary of the central retinal vein (Ip & Hendrick, 2019).

The common signs of RVO are haemorrhages all through the retina as well as the retinal veins appearing dilated and tortuous. Retinal haemorrhages happen when there is an

occlusion in the blood vessels in the eye that causes higher intravenous pressure (Flaxel et al., 2020) and blood will leak out from the site of occlusion. Moreover, bleeding in the photosensitive tissues leads to retinal haemorrhage which varies in size, location, and distribution as it could be flamed-shaped, dot and blot, or boat-shaped in the vitreous cavity, main or branches of retinal veins. Figure 1.1 depicts a retinal fundus image with scattered retinal haemorrhages and Figure 1.2 shows extensive haemorrhages in all quadrants throughout the retina.

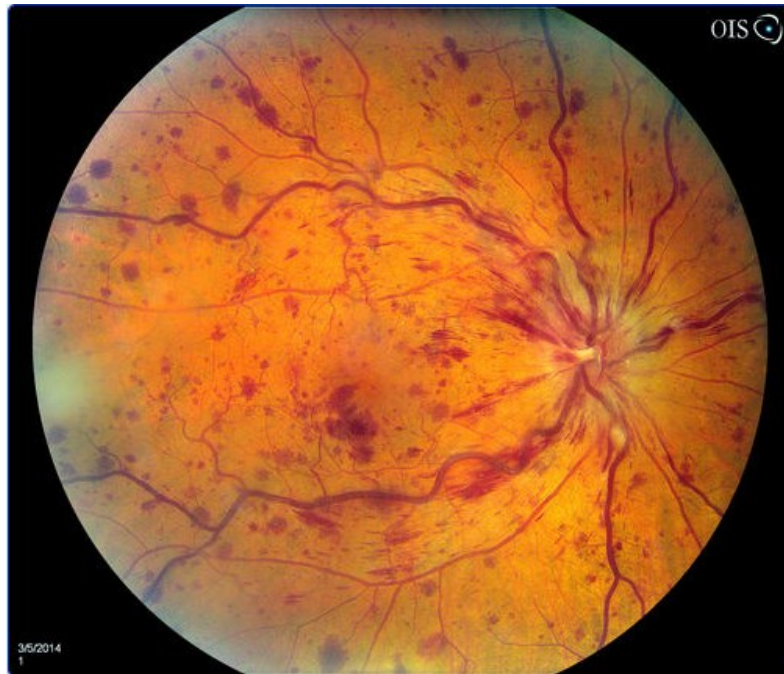


Figure 1.1: Retinal fundus image of CRVO with retinal haemorrhage. The retinal fundus image shows CRVO with scattered retinal haemorrhages. (Retina Gallery, 2014)

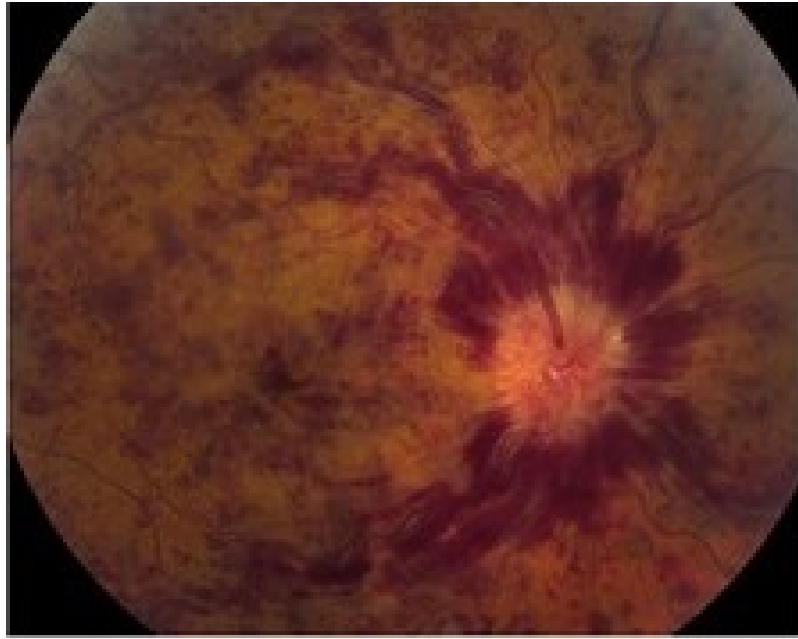


Figure 1.2: Fundus image of the right eye with CRVO. Appearance of substantial intraretinal haemorrhaging, flame haemorrhages, and optic nerve oedema associated with CRVO. (Ip & Hendrick, 2019)

Besides that, the cotton wool spot is also one of the abnormal features caused by RVO, it is seen as white and fluffy patches on the retina as shown in Figure 1.3 (Pachade et al., 2021). The presence of cotton wool spots is a result of lipids leaking out from the site due to a lack of oxygen supply to tissues and damaged tissue (Kipli et al., 2018). When there is any presence of cotton wool spots, intraretinal haemorrhage, or dilated and tortuous retinal vein, it is classified as an ischaemic type (Sivaprasad et al., 2015). While the non-ischaemic type shows there are mild dilated and tortuous veins, with haemorrhages in all quadrants. Retinal veins will also appear dilated and tortuous that follow a serpentine path (path that curves back and forth or twists and turns) as shown in Figure 1.4, this sign is associated with diseases such as hypertension and diabetes due to blockage by thrombus (Mehta, 2020; Pachade et al., 2021).

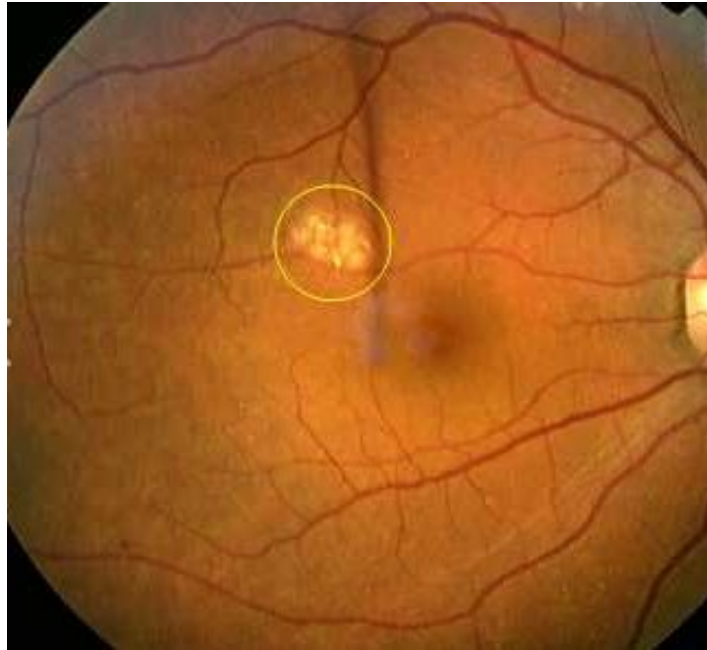


Figure 1.3: Retinal fundus image of the cotton wool spot. The image shows a retinal fundus image with a cotton wool spot (circled in yellow). (Glycosmedia, n.d)

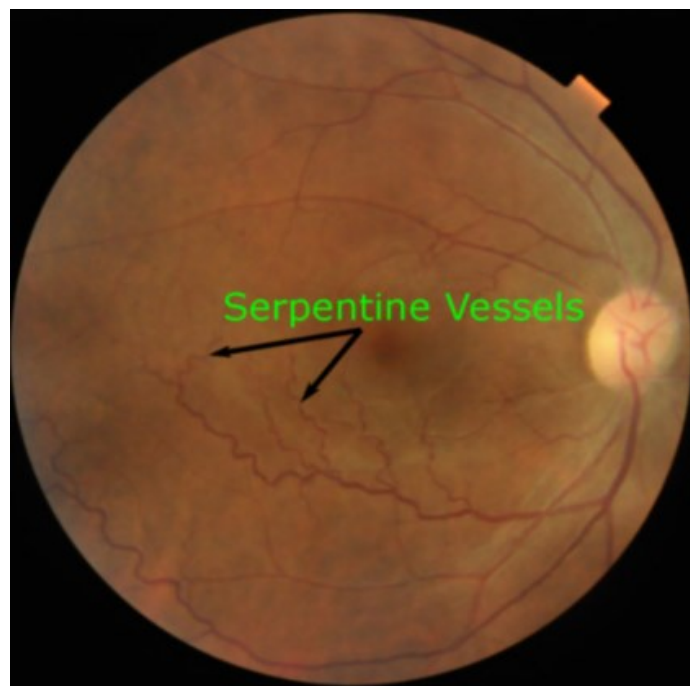


Figure 1.4: Retinal fundus image with tortuous vessels. The retinal fundus image shows tortuous vessels (black arrow). (Pachade et al., 2021)

According to Mehta (2020), the symptom of CRVO and BRVO is sudden vision loss, and the signs are the presence of dilated, tortuous retinal vein, retinal oedema, and haemorrhages scattered throughout the retina. Additionally, Flaxel et al. (2020) mentioned that vitreous haemorrhage is more likely to develop when BRVO occurs, and CRVO are more likely to cause macular oedema. Nonetheless, the severity of RVO and the appearance of these features differs in every case but with analysis of retinal imaging experts can evaluate and determine the condition of each patient. With treatment, haemorrhages and cotton wool spots could also be cleared (Flaxel et al., 2020).

For reference, a retinal fundus image of a non-diseased retina which has none of the signs of RVO is shown in Figure 1.5. In this study, these abnormalities are the anomalies of RVO that this study aims to detect on the retinal fundus image.

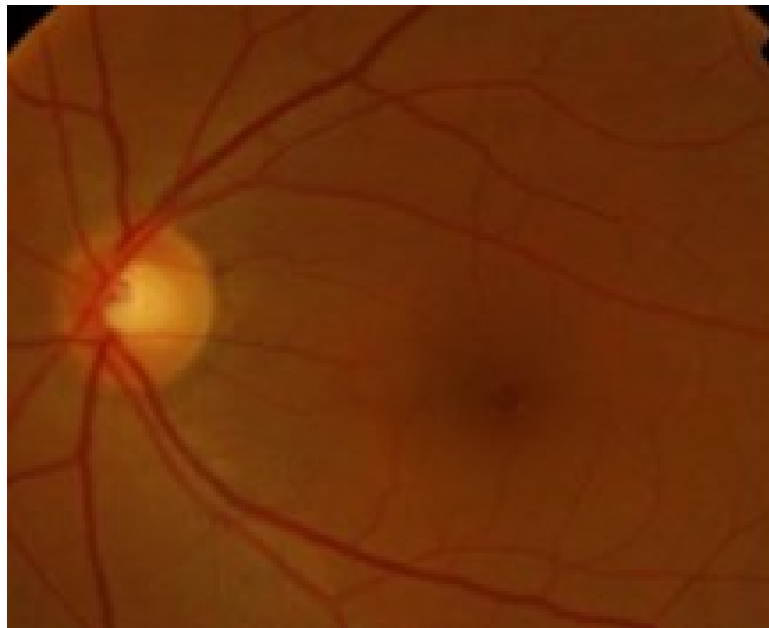


Figure 1.5: Retinal fundus image of a healthy patient, without retinal vein occlusion.

Table 1.1 is used to determine and summarise the criteria of signs of retinal vein occlusion by referring to The Royal College of Ophthalmologists (Nicholson et al., 2022) and ClinicalKey, (2021). The categorisation of retinal vein occlusion into central or branch is dependent on the area of the appearance of the signs which is either all over the retina for CRVO or around the affected branch vein for BRVO (ClinicalKey, 2021; Nicholson et al., 2022). In this research, only the haemorrhages and cotton wool spots are segmented as abnormal features caused by retinal vein occlusion due to time constraints and hardware limitations.

Table 1.1: Characteristics of RVO

The Royal College of Ophthalmologists	ClinicalKey
Deep retinal haemorrhages	Haemorrhages
Cotton wool spots	Cotton wool spots
Disc oedema	Macular oedema
Dilated and tortuous retinal veins	Dilated and tortuous retinal veins
Retinal oedema	Optic nerve head neovascularization and/or neovascularization elsewhere
Capillary non-perfusion	Optic disk oedema

1.4 Problem Statements

Existing retinal disease classification models, such as those proposed by (Butt et al., 2019; Choudhury et al., 2017; Hagos & Kant, 2019; Xu et al., 2017), lack interpretability. These models typically provide only a diagnosis without offering insights into the underlying reasoning. This "black box" nature, characteristic of Convolutional Neural Network (CNN)

classification models, weakens trust and reliance due to the inability to understand how conclusions are reached. Additionally, identifying the cause of errors becomes challenging. To address these issues, visualising the presence of abnormal signs associated with the disease could significantly enhance model interpretability. Furthermore, pinpointing the location of these abnormalities is crucial for clinicians to assess disease severity and monitor progression. Interpretability is essential in medical diagnosis due to the critical decisions involved. It develops trust and safety by allowing clinicians and patients to understand the rationale behind diagnoses. Moreover, it facilitates the identification and rectification of biases within these models.

The performance of segmenting signs of RVO in ultra-widefield (UWF) fundus images is uncertain due to the limited number of studies in this area. While deep learning models have been widely applied in ophthalmology for conditions such as diabetic retinopathy, age-related macular degeneration, glaucoma, and cataracts (Shetty et al., 2021), there is a noticeable gap in research specifically targeting the segmentation of RVO signs using UWF images. Most existing studies focus on image-level classification and segmentation of lesions related to other retinal diseases using regular fundus images, which capture up to a 75° field of view (Abdelmaksoud et al., 2021; Ali, 2019; Choudhury et al., 2017; Khojasteh et al., 2018; Luo et al., 2021; Partovi et al., 2016; Wan et al., 2021; Zhang et al., 2021). In contrast, UWF fundus images provide a broader field of view of up to 200°, allowing for the capture of more peripheral retinal abnormalities (Zhang et al., 2021). Figure 1.6 shows an example of a regular fundus image with a 50° field of view, while Figure 1.7 illustrates a UWF fundus image with a 200° field of view. The angle of the retina captured depends on the imaging device used. Given the potential for RVO signs to occur in the peripheral retina, UWF images are preferred for screening to improve diagnostic accuracy

and efficiency. Employing deep learning to segment retinal vein occlusion signs in UWF images could significantly enhance the accuracy and efficiency of diagnostic tools. Therefore, addressing this research gap by focusing on the development and evaluation of deep learning models specifically for RVO segmentation in UWF images is essential.



Figure 1.6: Regular fundus image

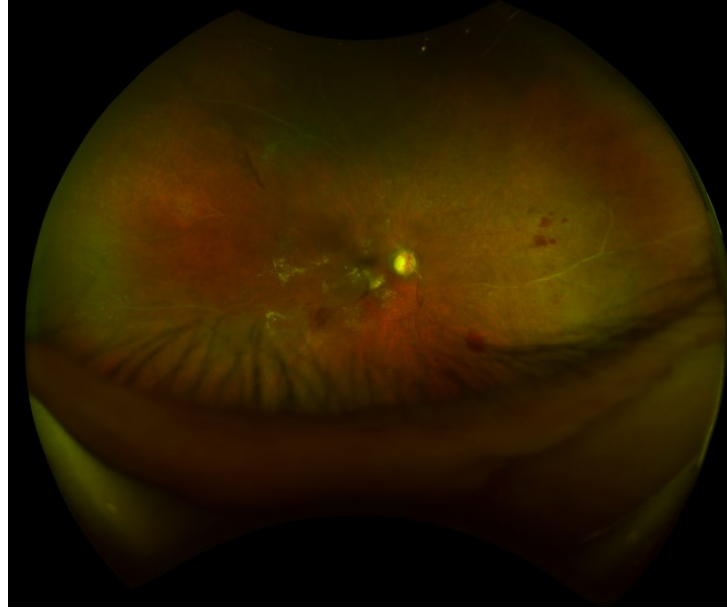


Figure 1.7: Ultra-widefield fundus image

While deep learning models like U-Net have been widely and successfully applied to the segmentation of various ocular diseases, such as diabetic retinopathy and blood vessels, there is a noticeable gap in their application for segmenting abnormal features caused by RVO on UWF fundus images. U-Net, a convolutional neural network model introduced by (Ronneberger et al., 2015), has demonstrated its capability in segmenting biomedical images with precise localization of target features, as evidenced by its success in neuronal structure segmentation and winning the ISBI cell tracking challenge. Despite its effectiveness in other areas, the feasibility and effectiveness of the U-Net model specifically for RVO segmentation on UWF images have not been thoroughly studied. Addressing this gap is essential for advancing the field and improving the efficiency and accuracy of retinal disease research.

The process of manually segmenting and identifying abnormal features in retinal images presents significant challenges for researchers in the field of medical image analysis.

Retinal fundus images of patients are typically taken using a fundus camera by trained technicians or nurses under the supervision of an ophthalmologist. It captures the retina and provides insights into the condition of the eye for diagnosis, monitoring of eye conditions, and treatment of eye diseases. The retinal images are analysed by ophthalmologists to manually segment and identify the abnormal features associated with eye diseases. It requires expertise and attention to detail to identify and trace the boundaries of abnormalities on the image. The complexity of the features, the number of abnormalities, the varying quality of the images, and the diverse shapes and sizes of the features make it particularly challenging for this study to ensure that automated segmentation algorithms can accurately and efficiently identify abnormal features in retinal images. Addressing this problem is essential for advancing the field and improving the efficiency of retinal disease research.

1.5 Research Questions

The research questions for this study are as follows:

- i. How can the interpretability of deep learning models in classifying retinal disease be enhanced to improve the reliability and explainability of the diagnostic process?
- ii. What are the pre-processing steps that are effective and suitable for retinal fundus images?
- iii. How well does the proposed method perform in segmenting signs of retinal vein occlusion (RVO), such as haemorrhages and cotton wool spots, in ultra-widefield (UWF) fundus images?
- iv. How well does the proposed method perform in classifying the RVO disease?

1.6 Objective

This research primarily aims to determine an approach for segmenting and classifying signs of retinal vein occlusion in ultra-widefield retinal fundus images to improve model interpretability and diagnostic reliability.

1.6.1 Specific Objectives

The specific research objectives are as follows:

- i. To develop methods that improve the interpretability of deep learning models used for classifying retinal diseases.
- ii. To determine the most suitable pre-processing steps for retinal fundus images.
- iii. To measure the effectiveness of the proposed method in segmenting signs of RVO, specifically haemorrhages and cotton wool spots, in UWF fundus images.
- iv. To measure the effectiveness of the proposed method classifying RVO disease in UWF fundus images.

1.7 Contributions

The contributions of this research are as follows:

- i. An algorithm to detect RVO and visualise the abnormalities in UWF retinal fundus images to improve the interpretability of deep learning models.
- ii. The baseline effectiveness of the deep learning neural networks model, U-Net for segmentation of haemorrhage and cotton wool spot for RVO on UWF retinal fundus images will be established.
- iii. A method to train U-Net for segmentation of haemorrhage and cotton wool spot of retinal vein occlusion with a limited amount of UWF images.

1.8 Definition of Terms

1.8.1 Conceptual Definitions

- i. Retinal vein occlusion (RVO) is an ocular disease that results from a blockage in the retinal vein due to the formation of a blood clot (Nicholson et al., 2022).
- ii. Abnormal signs of RVO refer to the presence of haemorrhage and cotton wool spot in the retina which appears due to obstructed blood flow (Mehta, 2020; Pachade et al., 2021).
- iii. Ultra-widefield fundus images are images of the retina with at least 100° field of view which are used for monitoring eye conditions and screening for eye diseases (Patel et al., 2020).
- iv. Segmentation of abnormal signs of RVO refers to identifying the regions and the boundaries of haemorrhages and cotton wool spots on fundus images using image processing and deep learning techniques.

- v. Classification of RVO refers to identifying whether a fundus image has the presence of abnormalities associated with RVO and a fundus image classified as non-RVO means it does not have any of the abnormalities associated with RVO.

1.8.2 Operational Definitions

- i. Abnormal signs of RVO refer to the presence of retinal haemorrhage and cotton wool spots on fundus images.
- ii. Ultra-widefield fundus images are colour fundus images taken by a fundus camera with at least 100° to 220° field of view.
- iii. Segmentation of abnormal signs of RVO is about classifying every pixel of the abnormal signs of RVO on the fundus images to get a segmented map with the exact location of the abnormal signs. This process involves image pre-processing and the deep neural networks for segmentation which is the U-Net model.
- iv. Classification of RVO is about categorising fundus images into the class label, RVO or non-RVO based on the presence of abnormal signs of RVO. It involves training and testing CNN classification models.

1.9 Conclusion

There is more to explore on RVO and automated detection of RVO can solve problems that people are facing such as inaccessibility to ophthalmology services, time-consuming processes, inaccuracy, and inconvenience. This chapter has covered the background of this study, retinal vein occlusion, problem statements, research questions, contributions and objectives of this study.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

Generally, anomaly detection is about finding something that deviates from the norms or what is expected by comparing it with the rest of the same type of data and with experience or knowledge (Mehrotra et al., 2017). The spectrum of anomaly detection is very broad as anomaly detection has been used in the fields of finance, healthcare, cybersecurity, consumer home safety, manufacturing, and industry. The anomalous events and the type of data used to analyse anomalies vary in different fields.

For instance, detecting anomalies for theft of credit cards involves checking the owner's usual purchase amount, type of purchase, and place of purchase. For cybersecurity, the common application of anomaly detection is identifying fraudulent emails and there are many matters to look out for such as uncommon typographical errors, unknown senders, and uncommon salutations. Most email service providers use anomaly detection to analyse received emails and categorise potential fraudulent emails under spam or junk email folders to prevent users from unknowingly engaging in fraudulent activities.

In the context of healthcare, anomaly detection has been used for monitoring patients' vitals, and detecting diseased cells, or unusual mutations and not limited to these applications as more applications of anomaly detection are being explored to possibly assist experts in making decisions. These different types of anomaly detection require different types of data to analyse, the data used could be one-dimensional or multidimensional qualitative and quantitative data such as medical images, records, or trends.

This study is focused on the application of anomaly detection techniques in the healthcare field and the study aims to detect anomalies of retinal vein occlusion. The fundus image is frequently used to diagnose retinal disease because it provides professionals with information and allows them to examine any abnormalities in the retina. When aberrant characteristics related to retinal vein occlusion are present on retinal fundus imaging, abnormalities of retinal vein occlusion can be observed.

2.2 Computer Vision

Computer vision is a field of artificial intelligence that focuses on creating machines that can analyse images such as identifying objects in an image or video and understand the context so it can draw conclusions as humans do (Rasche, 2019). For example, identifying the car plate numbers of a certain car provided an image of cars on a highway. Computer vision has been applied in many real-world applications such as object recognition, face recognition, autonomous vehicles, image and video search, and medical image analysis. The techniques that are often used and associated are image processing and pattern recognition. Image processing is about enhancing the original image or extracting relevant features by applying some methods such as image thresholding, image conversion, image enhancement, and image arithmetic that manipulate the image.

Besides that, pattern recognition is always associated with computer vision because computer vision tasks such as object classification require identifying features that form patterns and their relationships in an object. There are a few types of recognition processes for stationary scenes that are of interest in this research which are classification, and segmentation. Image classification is about labelling an entire image to a class while image detection is finding regions of interest in an image, and image segmentation is about

segmenting an image into distinct objects. There is various computer vision approaches to solve recognition tasks such as feature engineering and feature learning or deep learning. Feature engineering is about using hand-crafted feature representation which involves image processing, feature extraction, feature integration and feature matching to make classification. Whereas feature learning refers to deep learning to automatically extract the feature representation from input data using convolution layers for classification. The feature engineering process provides more clarity in which features are extracted but it takes more effort and time, and it can be more complicated dealing with complex data.

2.3 Deep Learning

Machine learning refers to a sort of computer software that can learn by itself from given data whereas deep learning is a subset of machine learning (Wehle, 2017). Artificial neural networks arise from the idea of biological neural networks of the human brain, and it has more than one hidden layer. It can process numeric and structured input data (Srivignesh Rajan, 2020). Artificial neural networks can have many hidden layers, which distinguishes them from traditional neural networks. An artificial neural network is a feedforward neural network with backward propagation which has three layers, there are the input layer, hidden layers, and output layer. Each layer has neurons with weights which are interconnected between these layers, the weights of neurons will be adjusted to improve the model and that is the way the model learns.

In the 2000s, deep learning has become widely used and explored in the fields of machine learning and artificial intelligence due to its breakthroughs (Minar & Naher, 2018). Deep learning can be categorised into supervised or unsupervised learning, both with the same goal of predicting class or numerical value; supervised learning is when the input data

is labelled while unsupervised learning occurs when input data is unlabelled, so the network has to train itself to identify the features to classify classes (Lin et al., 2020).

Deep learning employs multiple processing layers to learn patterns and it can effectively classify unstructured data such as images, text or sounds from large input data, unlike traditional artificial neural networks (LeCun et al., 2015). Feature representation learning is a big part of deep learning as it determines the performance of a model and feature extraction is done by the model itself which means it does not need hand-picked features.

Deep neural networks have many hidden layers, more than a traditional neural network would have; to identify features from input data and they will process more complex functions (Minar & Naher, 2018). Convolutional neural networks are a specific architecture of deep learning. They are multi-layered feedforward neural networks with various layers of convolution and pooling layers that can reduce the number of parameters to minimise computing power in processing data and most importantly convolutional layers extract important features from input data (Sumit Saha, 2018).

Convolutional networks are usually used for visual recognition tasks, and it has achieved great results, but pixel-level classification is preferred when it comes to classification tasks for biomedical images. Thus, Ronneberger et al. (2015) developed the U-Net architecture for precise segmentation of medical images.

2.4 Convolutional Neural Networks (CNN)

CNN has several convolutional and pooling layers and fully connected layers that are stacked together to identify and classify features (Rawat & Wang, 2017). Figure 2.1 shows an example of a CNN image classification pipeline. The input and output layers of a

CNN are the same as the artificial neural network (ANN), however the hidden layers distinguish them apart. Convolutional and pooling layers make up the hidden layers of CNN.

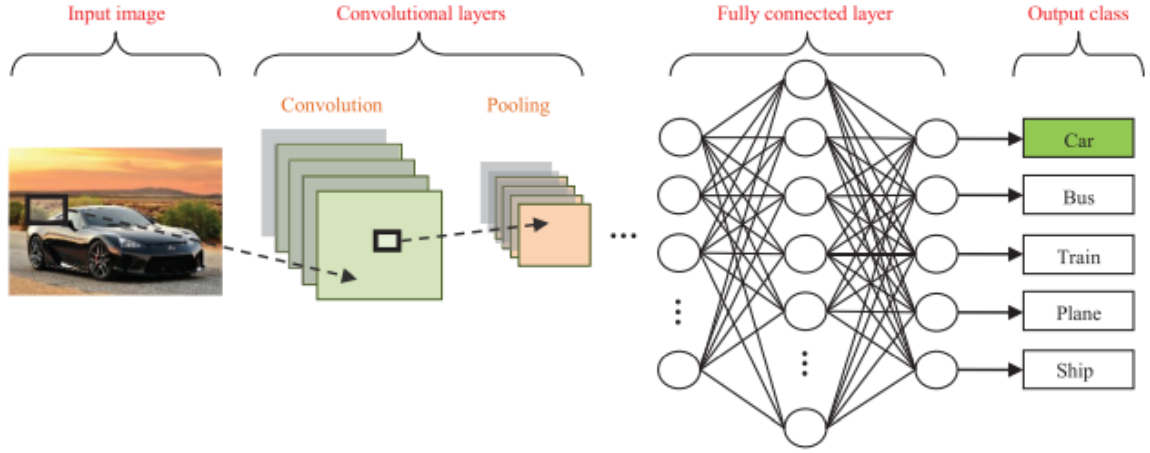


Figure 2.1: CNN image classification pipeline. (Rawat & Wang, 2017)

Firstly, a two-dimensional image is obtained to be fed into the model at the input layer. The image will be divided into several smaller images and each of them will be applied with a convolution kernel or filter which is a weight matrix. This operation is as defined in Equation 2.1. The results are convolution products which then combine to form feature maps.

$$g(x, y) = w * f(x, y) = \sum_{dx=-a}^a \sum_{dy=-b}^b w(dx, dy) f(x + dx, y + dy) \quad \text{Equation 2.1}$$

where $g(x, y)$ is the filtered image, $f(x, y)$ is the original image, w is the convolutional kernel, dx and dy refer to the shifts in the x axis and y axis within the convolutional kernel, and a and b determine the range of dx and dy . $-a \leq dx \leq a$ and $-b \leq dy \leq b$ are the elements of the convolutional kernel, $w(dx, dy)$ represents the weights assigned to a specific position relative to the centre of the kernel, and

$f(x + dx, y + dy)$ is the value of the input function at the point shifted by dx in the x direction and dy in the y direction from the current position (x, y) .

Figure 2.2 demonstrates how the convolution products are obtained as it is different from traditional neural networks where matrix multiplication is typically used (Wang & Yang, 2019). Each convolution kernel involves element-wise multiplication and adding all the values to get the output. This way the model is more efficient because it will not take up too much storage and reduce the computational power for the calculation, this is known as sparse interactions or connectivity, CNN also has the characteristic of weight sharing because the weight of each kernel is fixed when it is applied to other parts of the image thus reducing the parameters and overcoming the overfitting problem (Wang & Yang, 2019).

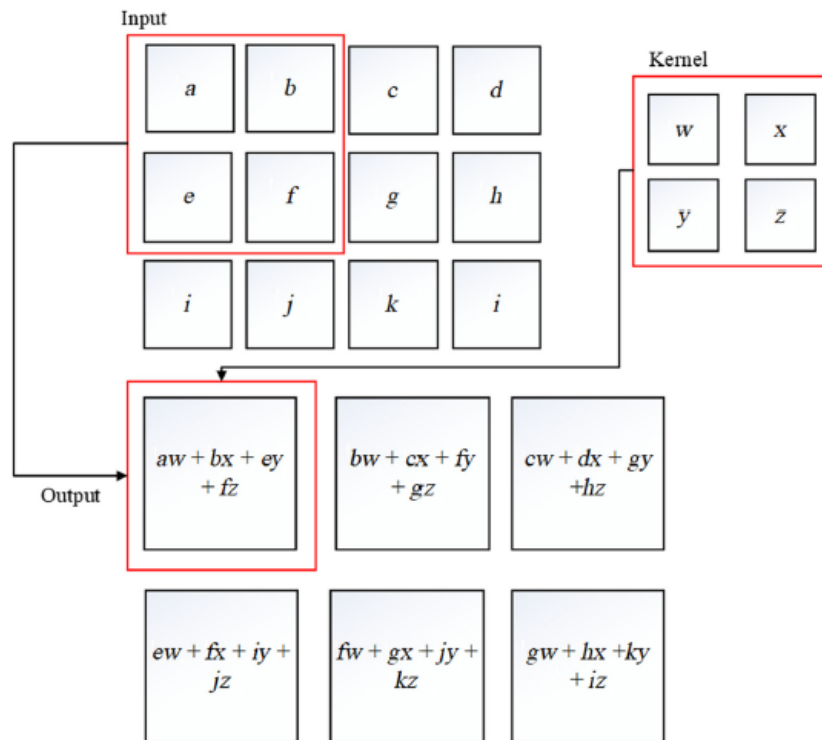


Figure 2.2: An example of a 2-D convolution.(Wang & Yang, 2019)

At first, the weights of the convolution kernel are initialised, and it will be adjusted using gradient descent as Equation 2.2, in the backward propagation phase (Kousai, 2019).

$$w^+ = w - \eta \Delta C \quad \text{Equation 2.2}$$

where w^+ is the new weights, w is the current weights, η is the learning rate, and C is the gradient of the Cost function.

Then, the activation function, Rectified Linear Units (ReLU) is applied to the feature maps, and it is used to reduce non-linearity by replacing negative pixel values with zeros as Equation 2.3 (Piyush Mishra, 2020).

$$ReLU(x) = \{0, x < 0\}, \{x, x \geq 0\} \quad \text{Equation 2.3}$$

Usually after a convolutional block, the feature maps go through a pooling layer. At the pooling layer, pooling operations such as average pooling or maximum pooling are done by applying a kernel on the feature map. Max pooling involves selecting the highest pixel value from a specific region of the feature map as the highest value represents the most relevant feature of the image and the region is usually 2 x 2 or 3 x 3 (Kousai, 2019). The general equation of pooling can be defined as in Equation 2.4.

$$Output = \left[\left\{ \frac{n_h - f}{s + 1} \right\} \times \left\{ \frac{n_w - f}{s + 1} \right\} \times n_c \right] \quad \text{Equation 2.4}$$

where the input of the pooling layer is n_h, n_w, n_c , f is the size of the kernel, and s is the stride.

A 2×2 maximum pooling with stride 2 means the highest value of a 2×2 region is preserved and stride 2 means moving 2 units from the starting point for the next region thus the output is a 2×2 feature map. Then, the feature map goes through several convolutional blocks to obtain a final feature map of the input image. At the final layer, there is a fully connected layer to perform classification tasks and a Softmax activation function is used to classify the input. However, the approach of CNN has some downsides due to the computational power needed to process the dense layer with a large number of parameters as segmentation occurs at each pixel. Hence, a fully convolutional network with downsampling and upsampling is designed to extract relevant features and localisation thus more suitable for image segmentation.

2.5 Literature Review

This section comprises reviews of previous studies on using deep learning algorithms for the classification of retinal disease that is image-level classification using retinal fundus images. The algorithm learns to differentiate features of images of each class through its training process. Besides that, this section reviews some of the methods that have been used in previous studies for pixel-level segmentation on retinal fundus images. There are algorithms for segmenting retinal lesions and features such as hard exudates, haemorrhages, and microaneurysms that are associated with retinal diseases.

2.5.1 Multi-class Diabetic Retinopathy Classification with Convolutional Neural Network

Diabetic retinopathy is a type of retinal disease, and the stage of this condition can be assessed by analysing a fundus image. Butt et al. (2019) used a convolution neural network (CNN) for categorising retinal fundus images into 5 severity classes: no DR, mild, moderate, severe, and proliferative DR. They also compared the performance of the model with different pre-processing steps and CNN model architecture. Table 2.1 shows the three different architectures of the CNN model. In the public dataset, EyePACS, 35126 images are employed to train and test the model with a 4:1 ratio. The pre-processing steps involve extracting the RGB channels and grayscale channels from the images and resizing them to 100 x 100 pixels due to hardware limitations for training. Three different architectures of the CNN model are trained on each of the channels.

Table 2.1: The architecture of the models

Model 1	Model 2	Model 3
2 convolution layers with 32 filters of 3 x 3 size with ReLU activation	A convolution layer with 32 filters of 5 x 5 size with ReLU activation	A convolution layer with 48 filters and 32 filters for the 3 consecutive convolution layers with ReLU activation
After each convolution layer, a 3 x 3 max pooling layer is added	2 x 2 max pooling layer	After each convolution layer, a 2 x 2 max pooling layer is added
Dropout method of 20% is applied	A convolution layer with 32 filters and 3 x 3 size with ReLU activation	A fully connected dense layer of 512 and 128 with ReLU activation
Flattening to 1-dimensional vector	2 x 2 max pooling layer	Final dense layer with Softmax activation
Fully connected dense layer of 256 and 64 with ReLU activation	Flattening to 1-dimensional vector	
Final dense layer with Softmax activation	Fully connected dense layer of 128 and 64 with ReLU activation	
	Final dense layer with Softmax activation	

The performance of each model on each of the datasets is evaluated and it proved that adding more convolution layers and pooling layers in Model 3 caused overfitting and thus did not improve the performance. Besides that, training with more epochs showed better results. Model 2 outperformed the other models with a 97.08% accuracy score on the blue channel dataset which was trained with 128 epochs and a batch size of 40.

2.5.2 Diabetic Retinopathy Detection with Deep Convolutional Network and Transfer Learning

Hagos & Kant (2019) pointed out that a large dataset is essential for medical image classification tasks so the algorithm can learn better but obtaining such a dataset is challenging. Thus, they suggested using a pre-trained Inception-V3 that was trained for the ImageNet Large Scale Visual Recognition Challenge 2015 (ILSVRC) to classify images into healthy and unhealthy or diabetic retinopathy (DR) classes.

The method begins with data collection and preparation where training images are composed of 1250 normal images and 250 images with 4 different grades of DR are obtained from a public dataset, while the testing dataset has 1000 images of normal and 4000 images with DR grading. Data pre-processing includes removing unnecessary background by cropping and the images are all resized to 300 x 300 pixels then the local average is applied. The pre-processed images are fed into an Inception-V3 model with ImageNet-trained weights for feature extraction and training the model with a stochastic gradient descent optimizer. A ReLU activation layer and a Softmax layer are added to the model. The learning rate was set to 0.0005 and the cosine loss function was used. From the evaluation of the model on testing images, it scored 90.9% accuracy and loss of 0.0394 which was better than a study that used more than 35000 images and with data augmentation.

2.5.3 Diabetic Retinopathy Detection with Deep Convolutional Network

Xu et al. (2017) studied image classification tasks with deep convolutional neural networks instead of traditional machine learning which relies on hand-picked feature extraction. Hence, a deep CNN model is trained for classifying fundus images with diabetic retinopathy and compared with feature-based extraction methods.

The dataset comprises 800 training images and 200 testing images of normal and diabetic retinopathy cases. The different methods used are 2 deep learning approaches and 4 feature extraction approaches as shown in Table 2.2. The CNN model architecture used is described in Table 2.3. For the deep learning approach with CNN and data augmentation, the data augmentation includes rotation, flipping, shearing, rescaling, and translation.

Table 2.2: Deep learning and feature extraction methods used

Method	Accuracy (%)
Hard exudates + Gradient Boosting Machines	89.4
Red lesions + Gradient Boosting Machines	88.7
Microanueyrysm + Gradient Boosting Machines	86.2
Blood vessel detection + Gradient Boosting Machines	79.1
CNN without data augmentation	91.5
CNN with data augmentation	94.5

Table 2.3: The architecture of the CNN model

Output Shape	Description
224 x 224 x 3	Input
222 x 222 x 32	3 x 3 convolution, 32 filter
220 x 220 x 32	3 x 3 convolution, 32 filter
110 x 110 x 32	2 x 2 max-pooling
108 x 108 x 64	3 x 3 convolution, 64 filter
106 x 106 x 64	3 x 3 convolution, 64 filter
53 x 53 x 64	2 x 2 max-pooling
53 x 53 x 128	3 x 3 convolution, 128 filter
51 x 51 x 128	3 x 3 convolution, 128 filter
49 x 49 x 128	2 x 2 max-pooling
24 x 24 x 256	3 x 3 convolution, 256 filter
22 x 22 x 256	3 x 3 convolution, 256 filter
11 x 11 x 256	2 x 2 max-pooling
4096	Flattened and fully connected
1024	Fully connected
2	Softmax

From the comparison of the performance of the methods, the deep learning approach showed better results as compared to the feature extraction approach. Applying data

augmentation when using the CNN model showed a slightly higher accuracy score (94.5%) which could have reduced overfitting.

2.5.4 Central Retinal Vein Occlusion with CNN

The abnormal signs caused by central retinal vein occlusion are haemorrhages and tortuous and dilated veins which vary in size, colour, and texture. To extract all these features with classic feature extraction requires complex methods but using a CNN model can save that trouble as CNN will extract and learn the features automatically. Thus, Choudhury et al. (2017) trained a CNN model to classify CRVO and normal fundus images.

The dataset used is obtained from 4 public sources: STARE database, DRIVE database, Retina Image Bank, and Dr. Hossein Rabbani. A total of 108 CRVO images and 133 normal images are collected. The pre-processing steps are converting to a TIF file, resizing the images to 60 x 60 pixels, extracting the green channel, applying average filter and contrast limited adaptive histogram equalization. The architecture of the proposed CNN model which is based on LeNet is as in Table 2.4.

Table 2.4: The architecture of the proposed CNN model

Layer	Description
1	5 x 5 convolution layer, 32 filters, stride of 1
2	ReLU activation
3	2 x 2 max pooling layer, stride of 2
4	5 x 5 convolution layer, 128 filters
5	ReLU activation
6	2 x 2 max pooling layer, stride of 2
7	5 x 5 convolution layer, 512 filters
8	ReLU activation
9	2 x 2 max pooling layer, stride of 2
10	5 x 5 convolution layer, 1024 filters
11	ReLU activation
12	Fully connected layer

The model is trained with 110 CRVO and 100 normal images with 70 epochs. For testing, it scored 97.56% for accuracy, 96.97% for specificity and 100% for sensitivity as calculated using Equation 3.6. Although only 8 CRVO images and 33 normal images are used for testing, the accuracy is higher compared to other deep learning methods for detecting retinal vein occlusion. It is beneficial to use a deep learning algorithm to learn the features of CRVO due to the difficulty of designing an algorithm to extract each of the features of CRVO.

2.5.5 Branch Retinal Vein Occlusion Detection with Deep Learning

There is limited research on using UWF images for retinal vein occlusion classification tasks although UWF images capture more of the retina than regular fundus images. Nagasato et al. (2019) investigated deep learning and machine learning approaches to identify branch retinal vein occlusion (BRVO) which is one of the types of RVO on ultra-widefield fundus (UWF) images and used heatmap to visualise the region of haemorrhages.

The dataset used consists of 237 BRVO images and 229 non-BRVO images and it is obtained from a local hospital. 9-fold cross-validation is applied to the dataset so the dataset is split into 9 subsets and one of the subsets will be used for testing at each fold. Data augmentation techniques such as noise addition, inversion, brightness adjustment, gamma correction, and histogram equalization were applied to the training set to increase the number of images. The images were also resized to 256 x 192 pixels and normalized before training with the proposed model.

The deep learning approach is by using the VGG-16 model, and the architecture is shown in Table 2.5. The pre-trained weight on ImageNet is used during training so only the 5th convolution block and fully connected layers are trainable. The parameters for SVM are a cost parameter of 10 and a gamma value of 0.001. The results of both methods are as in Table 2.6. It shows that the deep learning approach where it can successfully extract features by itself performs better than SVM for this classification task.

Table 2.5: Architecture of VGG16

Layer	Description
Convolution block 1	2 convolution layers with 64 filters followed by max pooling
Convolution block 2	2 convolution layers with 128 filters followed by max pooling
Convolution block 3	3 convolution layers with 256 filters followed by max pooling
Convolution block 4	3 convolution layers with 512 filters followed by max pooling
Convolution block 5	3 convolution layers with 512 filters followed by max pooling
Fully connected layer	2 fully connected layers, dropout of 25% at the 1 st layer
Output layer	Softmax activation

Table 2.6: Results of VGG16 and SVM

Metrics	VGG16	SVM
Sensitivity (%)	94.0	80.5
Specificity (%)	97	84.3
Area under curve	0.976	0.857

2.5.6 Retinal Vein Occlusion Diagnosis and Segmentation with Deep Learning

Chen et al. (2021) proposed a retinal vein occlusion disease classification model as there is limited research on RVO as compared to other retinal diseases even though it causes serious effects. They compared 4 CNN models: ResNet-50, Inception-V3, DenseNet-121, and SE-ReNeXt-50 in terms of performance on identifying regular fundus images into CRVO, BRVO, normal and non-RVO classes. These models are remarkable at the ImageNet

challenge. Besides that, they segment lesions that are associated with RVO which are cotton wool spots, haemorrhages, hard exudates, and dilated and tortuous blood vessels. They compared 4 segmentation models: FCN-32s, DeepLab-v3, DANet, and Lesion-Net-8s on the performance of segmenting these lesions.

The dataset is obtained from the public database, eyePACS and privately collected data which has a total of 8600 regular fundus images. The statistics of the dataset and the division of data into training, validation and testing subsets are shown in Table 2.7. The data are colour retinal fundus images and they are fed into the models without any pre-processing steps. The performance comparison on RVO classification of the average score of the 4 models shows that Inception-v3 is more effective as tabulated in Table 2.8 and the performance comparison on the average score of lesions segmentation of the 4 models shows that the DeepLab-v3 model achieved better results as shown in Table 2.9. The DeepLab-v3 achieved a sensitivity of 0.81 and 0.92 for the segmentation of cotton wool spot and flame-shaped haemorrhage. The dataset used for lesion segmentation is the images from eyePACS but the ground truth of all the signs is manually annotated by ophthalmologists and it is not available to the public. This study shows the capability of deep learning on RVO classification and segmentation thus the ability to use it for automated diagnosis of RVO in a clinical setting.

Table 2.7: Statistics of the dataset

Category	Training	Validation	Test	Total
Normal	4083	2055	2022	8160
CRVO	69	30	36	135
BRVO	86	33	56	175
Other	62	30	38	130
Total	4300	2148	2152	8600

Table 2.8: Performance of the models on RVO classification

	ResNet-50	Inception-v3	DenseNet-121	SE-ResNeXt-50
Sensitivity	0.85	0.93	0.90	0.88
Specificity	0.97	0.99	0.98	0.98

Table 2.9: Performance of the models on lesion segmentation

	FCN-32s	DeepLab-v3	DANet	Lesion-Net-8s
Sensitivity	0.62	0.74	0.72	0.81
Specificity	0.97	0.97	0.95	0.92

2.5.7 U-Net: Convolutional Networks for Biomedical Segmentation

U-Net, a method for biomedical image segmentation by Ronneberger et al. (2015) achieved first place in the IEEE International Symposium on Biomedical Imaging (ISBI) Cell Tracking Challenge 2015. Since then, it has been extensively explored in the medical

field for segmentation tasks and modifications have been made to the architecture to better perform for different tasks. This section reviews the original paper on the U-Net model, convolutional networks for biomedical image segmentation.

Ronneberger et al. (2015) proposed to improve the fully convolutional network, which was also a method known for image segmentation, so the training does not require a large dataset while producing accurate segmentations and relatively faster. The network architecture has a contracting and expansive path as depicted in Figure 2.3. The contracting path has a common architecture of a convolutional network where there are 2 convolution layers (3 x 3) with rectified linear unit activation function (ReLU) and a max pooling (2 x 2, stride of 2). The number of kernels is doubled after each downsampling step, whereas the resolution of the feature map is halved by a 2 x 2 convolution in the expansive step. In the expansive path, there are also 2 convolutions (3 x 3) with ReLU activation function. Cropping is also applied as border pixels are lost in every convolution. Finally, a convolution layer (1 x 1) maps each of the 64 component feature vectors to the classes.

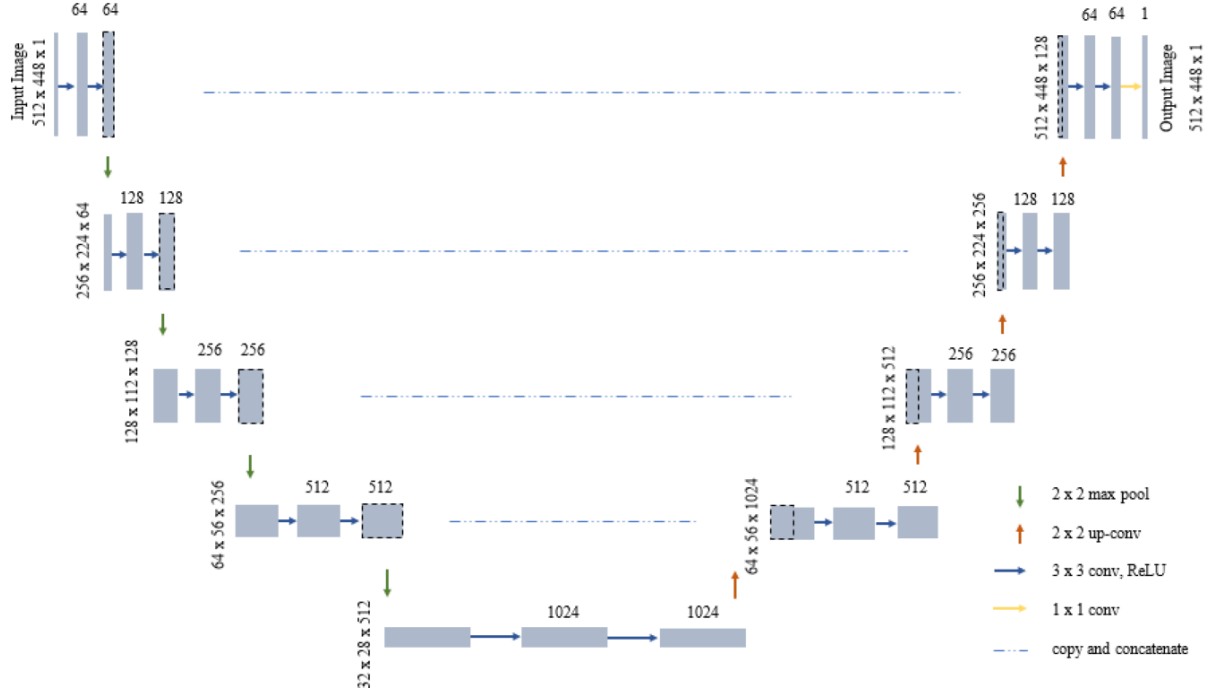


Figure 2.3: U-Net architecture. The blue box represents a multi-channel feature map with its number of channels on top of the box. The number at the lower left edge of the box is the x-y size. The white box represents copied feature maps. The arrows signify different operations.

Before training the model with images, training images undergo elastic deformations which is a data augmentation technique for the model to learn invariance. The dataset that is used is the serial section transmission electron microscopy (ssTEM) of the *Drosophila* first instar larva ventral nerve cord (VNC) and fully annotated ground truths that are obtained from the EM segmentation challenge. The training and testing data consist of 30 greyscale images of 512 x 512 pixels with ground truths while the ground truths of testing data are not available as it is for evaluating the model. Besides that, a weighted loss is used to separate background labels between touching objects of the same class.

The model is evaluated by comparing the predicted membrane probability map with other methods. The U-Net model has achieved the best score of warping error of 0.0003529. The U-Net model is also evaluated for cell segmentation on “PhC-U373” (Glioblastoma-

astrocytoma U373 cells on a polyacrylamide substrate recorded by phase contrast microscopy) and “DIC-HeLa” (HeLa cells on a flat glass recorded by differential interference contrast (DIC) microscopy) which were the challenge for the ISBI cell tracking in 2014 and 2015. The model was trained with 35 and 20 training images respectively. The average intersection over union (IOU) was the highest average as compared to other methods scoring an IoU of 0.9203 for the PhC-U373 dataset and 0.7756 for DIC-HeLa. In summary, the U-Net model showed great performance on biomedical segmentation tasks with only a few annotated images. Hence it has been adapted for many other biomedical segmentation tasks

2.5.8 Lesions Segmentation with Improved U-Net Model (EAD-Net)

Wan et al. (2021) proposed a segmentation method of microaneurysms (MAs), haemorrhages (HEs), and exudates which could be beneficial for this study as it solves similar problems and has achieved good performance. Figure 2.4 shows the segmentation of microaneurysms, haemorrhages, hard exudates, and soft exudates. Their study presented a method to segment features of diabetic retinopathy automatically by using an improved U-net model as manual segmentation requires a lot of time and effort.

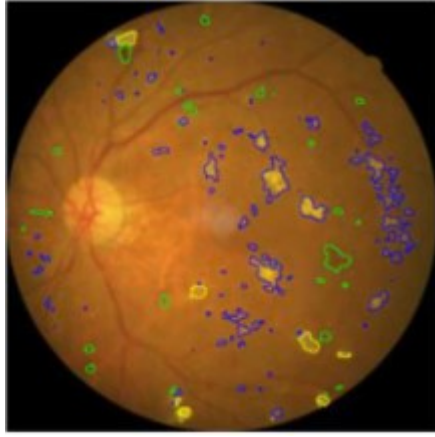


Figure 2.4: Segmentation of lesions. The result of the segmentation of microaneurysm (red colour), haemorrhages (green), hard exudates (blue), and soft exudates (yellow). (Wan et al., 2021)

The datasets used are e_optha_EX, Indian Diabetic Retinopathy Image Dataset (IDRiD), and datasets that are obtained from their local affiliated eye hospital which are retinal fundus images that have been annotated with MAs, HEs, soft and hard exudates. e_optha_EX has 82 annotated images, IDRiD has 81 annotated images, and the local dataset has 262 images. The pre-processing steps are removing redundant black edges, filling in the short side of the image, and then all the images and ground truth images are normalised into 1024 x 1024 pixels. Data augmentation transformation is also applied to obtain a larger dataset and due to limited computational power, the images were reduced to 512 x 512 pixels. Then, the images from IDRiD, e_optha_EX and the local dataset are divided into a ratio of 2:1 for training and testing.

Wan et al. (2021) proposed a model known as EAD-Net which is a U-Net-based model. The proposed model used is a traditional U-Net that has been modified to perform better. The basis of U-Net is that it has an encoder and decoder. There are improvements in the encoder module, decoder module and added dual attention module. The encoder part of

the U-Net is replaced with a convolution block with a residual structure and a 2 x 2 pooling layer is added. The other pooling layers are replaced with convolution and identity blocks in the downsampling process as too many pooling layers may not be ideal for recovering small lesions due to semantic information loss. The combination of convolution block with identity block is adapted from the ResNet model as skip connection can reduce the depth of the architecture for faster convergence and may reduce overfitting. The dual attention module is also added for better feature representations. The decoder is composed of convolutional layers and batch normalisation that are for the upsampling process. The hyperparameters of the network are shown in Table 2.10.

Table 2.10: Hyperparameters of the network used in Wan et al. (2021)

Hyperparameter	Value
Batch size	2
Dropout rate	0.5
Optimizer	Adam
Loss function	Binary cross entropy and dice loss
Learning rate	0.0001, lowered by 10 times after 5 epochs of no improvement

The performance of the network is evaluated on three datasets: the public e_ophtha_EX dataset, the public IDRiD dataset, and the local intelligent ophthalmology dataset. The public e_ophtha_EX dataset and local intelligent ophthalmology dataset are

evaluated based on sensitivity, specificity, precision, accuracy, and F1 score as tabulated in Table 2.11 and Table 2.12. The public IDRiD dataset is evaluated based on the area under the precision-recall curve (AUPR) as depicted in Table 2.13.

Table 2.11: Performance of the proposed method on the e_ophtha_EX dataset

Evaluation metric	Value (%)
Sensitivity	92.77
Specificity	99.98
Precision	89.06
Accuracy	99.97
F1 score	90.87

Table 2.12: Performance of the proposed method on local intelligent ophthalmology dataset

Lesion	Sensitivity (%)	Specificity (%)	Precision (%)	Accuracy (%)	F1 score
MAAs	17.32	99.98	59.26	99.91	26.82
HEs	83.59	99.95	87.75	99.89	85.62
Hard exudates	84.60	99.99	93.51	99.98	88.83
Soft exudates	84.92	99.99	92.78	99.98	88.68

Table 2.13: Performance of the proposed method on the IDRiD dataset

Lesion	AUPR value
MAs	0.2408
HEs	0.5649
Hard exudates	0.7818
Soft exudates	0.6083

The result of the proposed method on e_optha_EX and IDRiD showed effective performance as compared to other state-of-the-art methods. The summarised method is shown in Table 2.14. There are limitations to this method where the performance of MAs and small exudates are not satisfactory. There are misidentifications and omissions of MAs and misidentification of blood vessels as HEs. This could be due to the small dataset of MAs compared to other lesions in the study hence the method is more effective on haemorrhages, and hard exudates. Wan et al. (2021) suggested ensemble networks or more pre-processing steps can be used to improve the performance. Table 2.15 presents a comparison between using EAD-Net and U-Net on the local dataset. It shows that EAD-Net achieved better scores in sensitivity, precision-recall, and F1 score. This effective novel lesion segmentation method could be applied to segment features of retinal vein occlusion especially on haemorrhages and hard exudates to achieve great results for this study. However, the dataset needs to be manually annotated which will take some time and effort.

Table 2.14: Overall steps of the proposed method

Methods	Steps
Pre-processing	Removing redundant black edges Resizing the images to 512 x 512 pixel Applying data augmentation transformation
Segmentation of exudate, haemorrhage, and microaneurysm	Train and test the U-Net model with an added dual attention module

Table 2.15: Performance comparison between EAD-Net and U-Net

Lesion	Model	Sensitivity (%)	Specificity (%)	Accuracy (%)
MAs	U-Net	13.17	99.97	99.90
	EAD-Net	17.32	99.98	99.91
HEs	U-Net	73.43	99.93	99.83
	EAD-Net	83.59	99.95	99.89
Hard exudates	U-Net	68.38	99.99	99.96
	EAD-Net	84.60	99.99	99.98
Soft exudates	U-Net	76.89	99.99	99.98
	EAD-Net	84.92	99.99	99.98

2.5.9 Diabetic Retinopathy Grading System and Multiple Retinal Lesions Detection

Abdelmaksoud et al. (2021) proposed a method that can automatically detect between healthy or diabetic retinopathy from fundus images and segment four lesions of DR which are exudates, microaneurysms, haemorrhages and blood vessels by using U-Net then it assigns the gradings of diabetic retinopathy (DR). They chose to use retinal fundus images because they can show blood vessels, optic discs, macula, and vascular abnormalities. Besides that, this study is necessary because usually an expert is needed to detect diabetic retinopathy, manually detecting abnormalities is time-consuming and can be inconsistent. From earlier methods, there is misidentification of lesions due to similarities, problems due to contrast, noise, illumination, and artefacts in fundus images, overfitting of DL models, and deep fine-tuned CNNs show promising results which outperform CNNs, thus they proposed a multi-label diagnosis system. For training and testing, public datasets such as DRIVE, STARE, HRF, IDRID, etc are used.

The method begins with pre-processing steps which are applying a median filter to reduce noise, applying histogram equalisation for brightness preservation based on dynamic stretching to enhance contrast and a green channel extracted from RGB enhanced image to be used in feature extraction by grey level run length matrix (GLRLM) and binary classification by SVM. Then, the feature extraction step is where GLRLM is used to extract the features of the green channel of all processed images to get a resulting feature vector for differentiating between healthy and DR images. The calculated feature vector is fed into a SVM classifier to classify into 2 classes: 0 for normal cases and 1 for DR cases. Those images labelled as 1 will then be segmented and assigned a DR grade.

There is also the post-processing stage where the pre-processed images with their ground truths are resized to 512 x 512 and binarised before they are used for the U-Net segmentation phase. The segmentation of lesions can help with the classification of DR grades as the severity of DR is decided by the presence of lesions. The next step is to train a model to segment exudates, microaneurysms, haemorrhages, and blood vessels, the model used is the U-Net model as it is known for segmenting images by training on ground truths of the images. The U-Net architecture consists of 3 blocks in the downsampling stage and 3 blocks in the upsampling stage. Each block in the downsampling stage has 2 convolution layers (3 x 3) with RELU activation function and a max pooling layer (2 x 2). 32 kernels are used at first then it is duplicated after each block reaches 512 kernels in the bottleneck phase. Whereas each block in the upsampling stage has 2 convolution layers (3 x 3) with RELU activation function and upsampling convolutional layer (2 x 2). The number of kernels used is reduced to half after each block. The last layer is a convolutional layer (1 x 1) for producing the lesions segmentation maps. The U-Net is trained 4 times on ground truths of the 4 lesions and the parameters used are specified in Table 2.16.

Table 2.16: Parameters of the proposed method by Abdelmaksoud et al. (2021)

Parameter	Value
Cross-validation	10-fold
Epochs	30
Steps per epoch	100
Batch size	16
Optimiser	Adam optimiser
Learning rate	1e-3

After that, the model can produce segmentations of exudates, blood vessels, haemorrhages, and microaneurysms. The segmented images are then classified into grades of DR (mild, moderate, severe, and proliferative DR) by using a multi-label classification support vector machine and before that, features are extracted from the segmented images to be used for input of the SVM. The features for the 4 lesions are extracted using a grey-level co-occurrence matrix (GLCM) to form the feature vector. The feature vector contains 12 GLCM descriptors, 4 ROIs (areas of lesions), and a bifurcation point (BP) count for each DR image.

The performance of the SVM classifier in identifying normal and DR images on HRF, IDRiD, MESSIDOR, STARE, and DIARETDB0 datasets with 5-fold cross-validation

is as shown in Table 2.17. From the comparison of results using different datasets, the SVM model that is trained on the HRF dataset is preferred and is used for the U-Net model.

Table 2.17: Results of binary classification on datasets

Database	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)
HRF	96.66	100.00	93.30	99.00
DIARETDB0	92.30	94.00	33.00	95.00
STARE	90.00	90.00	90.00	92.00
MESSIDOR	75.00	86.00	50.00	72.00
IDRiD	96.10	100.00	80.00	96.00

Then, the segmentation of blood vessels, bifurcation points, exudates, microaneurysms, and haemorrhages with the U-Net model are evaluated on the datasets. The U-Net is trained on the DRIVE dataset for blood vessel segmentation and tested on DRIVE, STARE, CHASEDB1, and HRF datasets. Table 2.18 shows the performance of segmentation of lesions on IDRiD and DIARETDB1 datasets. The average performance on these datasets is accuracy of 95.97%, area under curve of 95.78%, sensitivity of 66.4%, and specificity of 98.51%. The U-Net is trained on the IDRiD dataset and evaluated on the IDRiD and DIARETDB1 datasets for the segmentation of exudate, microaneurysm, and haemorrhage. The performance shows that the U-Net can segment the lesions very effectively. Then, the multilabel SVM model undergoes 10-fold cross-validation to grade the segmented images from DRIVE, STARE, MESSIDOR, IDRiD, CHASEDB1, HRF, DIARETDB1, and DIARETDB0 datasets and the average performance is as shown in Table

2.19. Evaluating on different datasets proved that it is robust and applicable in real-world conditions as it has been tested on fundus images of different sources and qualities.

Table 2.18: Performance of segmentation of lesions on IDRiD and DIARETDB1 datasets

Dataset	Lesion	Accuracy (%)	Sensitivity (%)	Specificity (%)	Area Under Curve (%)
IDRiD	Exudate	100	100	100	100
	Microaneurysm	99.9	99.8	99.5	98
	Haemorrhage	99.1	100	98	97.5
DIARETDB1	Exudate	98.72	97	96.2	97.8
	Microaneurysm	98	96	98	97
	Haemorrhage	96	95	97	97.5

Table 2.19: Results of SVM on grading DR on segmented images

Accuracy (%)	AUC (%)	Sensitivity (%)	Specificity (%)
95.05	91.85	86.11	86.8

2.5.10 U-Net Based InceptionV3 Model for Diabetic Retinopathy Segmentation

Kaur & Kaur (2022) explored segmentation for signs of diabetic retinopathy which are haemorrhages, microaneurysm, hard exudates, and soft exudates. They used U-Net with the pre-trained CNN network, InceptionV3 as the encoder of the U-Net where feature extraction takes place.

IDRiD and DIARETDB1 datasets are used for training, testing, and validation. The IDRiD dataset has 80 images with haemorrhages and 81 images with microaneurysm, hard exudates, and soft exudates. The DIARETDB1 dataset has 89 images with all the signs. 70% of the dataset is used for training, 20% for validation, and 10% for testing. The images are cropped to eliminate the excess black background and only the training dataset goes through the data augmentation process, which is random rotation of 45°, changes in contrast and brightness with a factor of 0.2, zooming of 1.2 and shearing with 0.2.

Then, the model is trained using the training dataset with a learning rate of 0.0001, batch size of 16, and Adam optimizer for 200 epochs. The segmentation scores for each of the signs achieved on IDRiD and DIARETDB1 datasets are presented in Table 2.20 and Table 2.21. IDRiD dataset shows better segmentation of haemorrhages and soft exudates with higher sensitivity scores whereas DIARETDB1 has a higher sensitivity score for segmentation of microaneurysm and hard exudates. They concluded that the IDRiD dataset is preferable for the detection of moderate DR while DIARETDB1 is good for the detection of severe DR. The advantages of using an InceptionV3 model for the encoder are reducing the duration of the training process and over-fitting issues because the weights used for training of the model are not randomised, it has pre-trained weights which are better for feature extraction. Applying the data augmentation process also helps diversify the training dataset and prevents over-fitting of the model.

Table 2.20: Evaluation metrics for IDRiD dataset

Signs	Sensitivity	Specificity	Accuracy
Haemorrhages	88.39	99.94	99.82
Microaneurysm	70.76	99.98	99.95
Hard exudates	85.90	99.95	99.83
Soft exudates	85.38	99.71	99.62

Table 2.21: Evaluation metrics for the DIARETDB1 dataset

Signs	Sensitivity	Specificity	Accuracy
Haemorrhages	81.00	99.87	96.41
Microaneurysm	80.86	99.90	98.77
Hard exudates	96.26	99.94	99.27
Soft exudates	73.64	99.00	99.67

2.6 Discussion

The previous section is about reviewing existing deep learning methods used for retinal disease classification. Table 2.23 summarises the methods for retinal disease classification used in previous studies. The process starts with deciding what type of data and the collection of the relevant data. Typically, pre-processing operations are performed on the data to enhance the photos and resize them to a standard size. Data augmentation is also used to increase the dataset. The data are then divided into a subset of training, validating, and testing images. Then the classification model architecture and evaluation

metrics that will be used have to be decided. After that, the effectiveness of the classification model in identifying the testing images is analysed and reported.

Mostly, an artificial neural network model is used for this retinal fundus image classification task. It is a supervised learning-based algorithm which means the algorithm is trained with labelled data and its corresponding target class. The model will learn to accurately predict the unseen images into the correct class by updating its weights, so the error is minimised.

Based on the reviewed studies, the dataset used for the classification of retinal diseases is labelled retinal fundus images either from public databases or privately collected data. The number of images ranges from 100 to 80000 which are divided for training, validating, and testing the model. The existing works have built their convolutional neural networks from scratch using pre-trained CNN models. Convolutional neural networks model can extract feature representations from input data using convolutional layers and perform classification using fully connected layers. While a pre-trained CNN model is a CNN model that was previously trained on a large dataset then it is used for a different task. Pre-processing steps and resizing are usually applied to the input images before inputting the images into the proposed model.

From the study by Butt et al. (2019), the suitable epoch size for a dataset with 28100 training images and 7025 testing images is 32 and above for better convergence. However, the training time and computational resources increase significantly as a larger dataset is used. It is also noted that adding more convolutional and pooling layers may result in overfitting for this specific study, but it relies on the characteristics of the dataset thus the

complexity of the architecture can be increased gradually while monitoring the performance of the model on the validation data to find the most effective architecture.

Based on previous studies that differentiate the effectiveness of deep learning and traditional machine learning methods, deep learning models achieved higher performance as compared to traditional machine learning algorithms with feature extraction methods. The deep learning method can save the time of manually selecting relevant features before feeding them into a classifier. Xu et al. (2017) showed that the CNN model with data augmentation performed significantly better than feature extraction with a machine learning algorithm. Despite that, the CNN model takes up to 2 days to train but the trained model can classify in a few seconds. Nagasato et al. (2019) also proved that the deep learning model with transfer learning performed better than the SVM model, however, a larger dataset with other eye diseases is needed for further evaluation.

Transfer learning is also useful when there are limited hardware and computational resources. There are many pre-trained CNN architectures such as the Inception-V3 model. The Inception-V3 model is good for feature extraction of target features with varying sizes because it applies various kernel sizes in parallel to the input image. This contributes to high performance on image classification tasks, especially for identifying the abnormal signs of eye diseases which differ in size (Hagos & Kant, 2019). They mentioned that deep learning is effective when training with insufficient data and a comparison of other pre-trained CNN models can be done to find the most effective for training with a smaller dataset.

Even though deep learning methods are effective in image classification tasks, there are several requirements and shortcomings to consider when using this method. Firstly, the data must be unbiased, high quality and sufficient to represent the population for the model

to achieve good generalization and avoid overfitting. The hardware storage and computational requirements also must be powerful enough to train and run the algorithms. From the existing works on the classification of retinal diseases, deep learning models are hard to understand in terms of the decision-making process. The output of a trained model is usually just a class label without explanation or visualisation of the important features in the input image that relate to the output. Hence, visualisation of the relevant features and their location on the input image provide more clarity on the model's decision, this is especially helpful for the healthcare field.

Therefore, existing studies on image segmentation of lesions related to retinal diseases are also reviewed. The process of building a segmentation model begins with data collection and preparation. The data consists of images and their corresponding annotated ground truth. The boundaries of the target sign must be accurately marked on the images to provide the exact location of the sign. The size of the ground truth must also match the original images and the order of the input images the model takes in must match the ground truth to ensure the model learns correctly. Before using the data for training, data augmentation and pre-processing can be applied to increase the dataset, enhance the images, and avoid overfitting. Then, the data can be divided into training, validation, and testing subsets for training the selected model architecture. After training, the model can be fine-tuned based on the result of the trained model on the validation subset in terms of the selected evaluation metrics. The intersection over union (IoU) is usually used for binary segmentation tasks. It calculates the number of correct pixels in the predicted and actual ground truth divided by the total number of pixels labelled as target signs in either the predicted or actual ground truth.

Based on the literature review of previous studies on image segmentation, the U-Net model is effective as it has shown high performance in segmenting medical images even with limited annotated datasets. Data augmentation can also be applied to increase the quantity and variability of the dataset. The architecture of the U-Net ensures low-level features such as the edges, textures, and patterns and high-level features which are the context or scene of the image are captured. This is done by concatenating the feature map from the decoder and encoder of the same level to preserve spatial and global information. The architecture of the U-Net can also be modified to improve the performance such as adding a dual attention module and adding more convolution layers, although increasing the parameters could require more computational power.

Besides that, segmentation of the lesions associated with eye diseases can help with the feature extraction stage for the classification model because the relevant features are already extracted. The output of the segmentation model will only contain the location and boundaries of the target signs which are crucial in the decision-making for diagnosis of eye diseases. In this study, segmentation of the abnormal signs of RVO is performed first then the segmented map is input into the classification model.

Table 2.23: Summary of literature review for image classification

Author(s)	Dataset	Proposed model	Type of classification and segmentation	Performance
Butt et al. (2019)	EyePACS Regular fundus images Training: 28100 images Testing: 7025 images	CNN	Diabetic retinopathy grading 5 classes: No DR, Mild DR, Moderate DR, Severe DR, and proliferative DR	Model 2 Accuracy: 97.08%
Hagos & Kant (2019)	EyePACS Regular fundus images Training: 1250 normal images and 1000 images with each DR grading Testing: 1000 normal and 4000 with each DR grading	Inception-V3	Diabetic retinopathy 2 class: DR and non-DR	Accuracy: 90.9% Loss: 0.0394
Xu et al. (2017)	Regular fundus images Training: 800 images Testing: 200 images	CNN	Diabetic retinopathy 2 class: Normal and DR	Accuracy: 94.5%

Table 2.23 continued

Choudhury et al. (2017)	Regular fundus images 108 images with central retinal vein occlusion and 133 normal images Training: 100 normal images and 100 CRVO images Testing: 33 normal and 8 CRVO	CNN	Central retinal vein occlusion 2 class: CRVO and normal	Accuracy: 97.56% Specificity: 96.97% Sensitivity: 100%
Nagasato et al. (2019)	Ultra-widefield fundus images 237 images with branch retinal vein occlusion and 229 images with non-BRVO Training: 414 Testing: 52	VGG-16	Branch retinal vein occlusion 2 class: BRVO and non-BRVO	Sensitivity: 94.0% Specificity: 97% AUC: 0.976
Chen et al. (2021)	eyePACS Regular fundus images 8160 normal, 135 CRVO, 175 BRVO, 130 non-RVO	Inception-v3	4 class: Central retinal vein occlusion, branch retinal vein occlusion, normal, and non-RVO	Sensitivity: 93% Specificity: 99% F-1 score: 0.95 AUC: 0.99

Table 2.23 continued

Ronneberger et al. (2015)	DIC-HeLa	U-Net model	DIC-HeLa and PhC-U373	DIC-HeLa
	35 training images			IoU: 0.7756
	PhC-U373			PhC-U373
	20 training images			IoU: 0.9203
Wan et al. (2021)	e_optha_EX: 47 regular fundus images with exudates	U-Net based model	Microaneurysm (MAs), haemorrhages (HEs), soft and hard exudates	Local dataset
	Training: 32 images			MAs
	Testing: 15 images			Sensitivity: 13.17%
				Specificity: 99.97%
	IDRiD: 81 regular fundus images with			HEs
	Microaneurysm (MAs), haemorrhages (HEs), soft and hard exudates			Sensitivity: 73.43%
				Specificity: 99.93%
	Training: 54 images			Hard exudates
	Testing: 27 images			Sensitivity: 68.38%
				Specificity: 99.99%
	Local dataset: 262 regular fundus images: 63 images with MAs, 84 images with HE, 86 with hard exudates, and 29 with soft exudates			Soft exudates
	Training: 175 images			Sensitivity: 76.89%
	Testing: 87 images			Specificity: 99.99%

Table 2.23 continued

Abdelmaksoud et al. (2021)	Regular fundus image	U-Net	Exudate, microaneurysm, and haemorrhages	IDRID
	Training: 54 images from IDRID			Exudate Sensitivity: 100%
	Testing: 27 images from IDRID, 89 images from DIARETDB1			Microaneurysm Sensitivity: 99.8%
				Haemorrhage Sensitivity: 100%
				DIARETDB1 Exudate Sensitivity: 97%
				Microaneurysm Sensitivity: 96%
				Haemorrhage Sensitivity: 95%

Table 2.23 continued

Kaur & Kaur (2022)	Regular fundus images	U-Net	Exudate, microaneurysm, and haemorrhages	IDRID
	IDRID			Exudate
	The training set			Sensitivity: 100%
	Haemorrhages: 56 images			Microaneurysm
	Microaneurysms: 57 images			Sensitivity: 99.8%
	Hard exudates: 57 images			Haemorrhage
	Soft exudates: 57 images			Sensitivity: 100%
	Validation set			DIARETDB1 Exudate
	Haemorrhages: 16 images			Sensitivity: 97%
	Microaneurysms: 16 images			Microaneurysm
	Hard exudates: 16 images			Sensitivity: 96%
	Soft exudates: 16 images			Haemorrhage
	Testing set			Sensitivity: 95%
	Haemorrhages: 8 images			
	Microaneurysms: 8 images			
	Hard exudates: 8 images			
	Soft exudates: 8 images			

2.7 Conclusion

From the reviewed works on image classification, the process of image classification is observed to understand the workings and its limitations are identified to formulate a solution. Existing works on image segmentation are also reviewed as they can be used to provide visualisation of the abnormal features associated with retinal disease. The process of image segmentation is identified as it will also be implemented in the proposed method. The proposed method is a combination of image segmentation and image classification.

CHAPTER 3

METHODOLOGY

3.1 Overview

From the reviewed studies, a suitable method is proposed to classify and segment signs of retinal vein occlusion which are haemorrhages and cotton wool spots on ultra-widefield retinal fundus images. This chapter discusses the method and steps taken to meet the research objectives.

3.2 Research Flow Chart

A flow chart systematically presents a workflow to solve a problem, guiding the research toward a solution. Figure 3.1 illustrates the flow of the research. The process begins with a review of existing studies on image classification in retinal diseases and the problems are identified. Image segmentation on lesions associated with retinal diseases is also explored through existing works to understand the workings of the segmentation method to construct a suitable method. After that, the relevant data is acquired, and the identified pre-processing steps are applied before implementing the selected segmentation and classification algorithms. The performance of the algorithms is analysed and optimised to achieve the optimal result before finalising the algorithms. Finally, the trained algorithms are evaluated on the testing dataset to report on the performance of the algorithms.

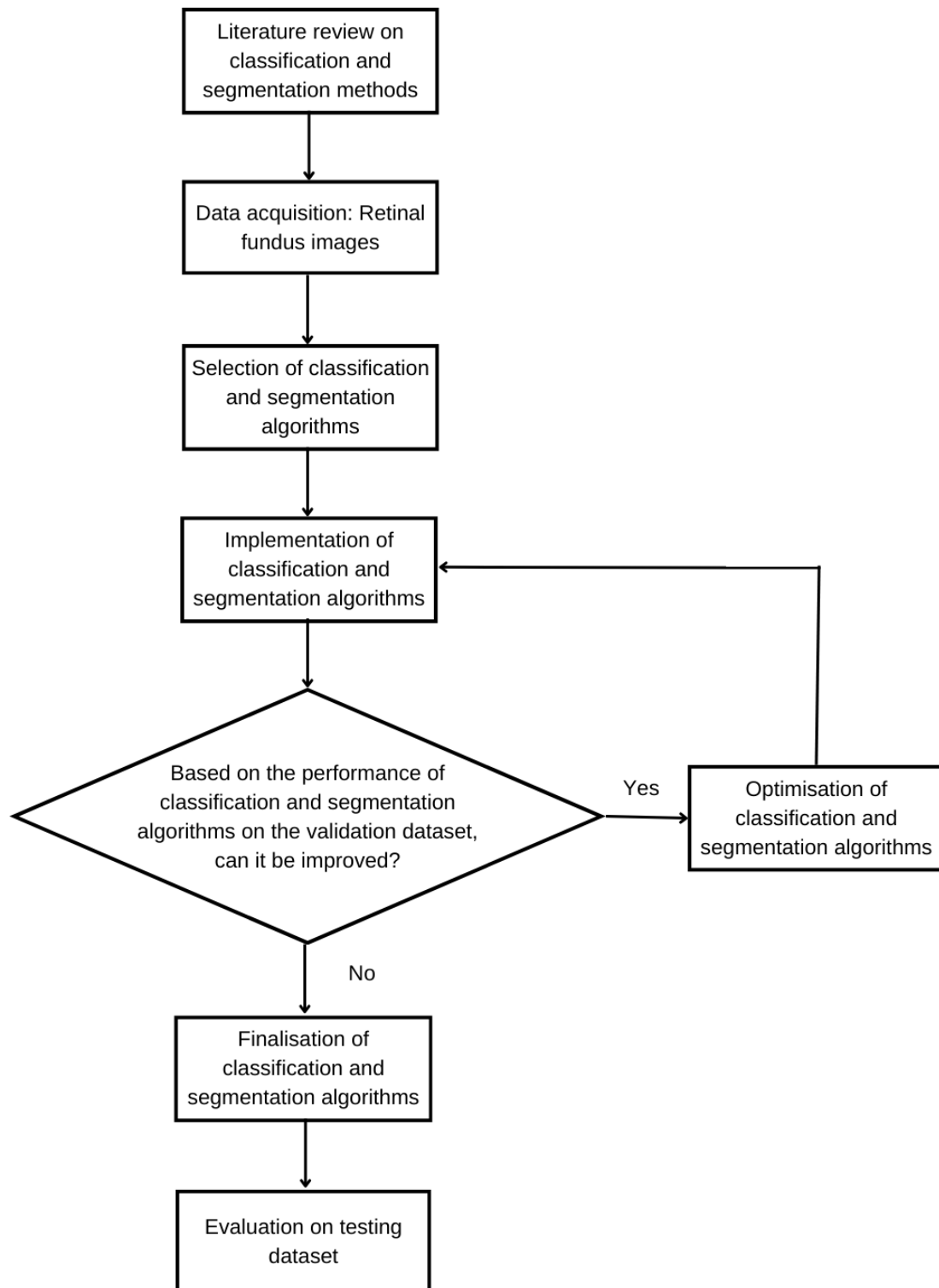


Figure 3.1: Research flow chart

Figure 3.2 outlines the proposed method. The proposed method starts with collecting regular and ultra-widefield fundus images from public and private databases. Then, a series of pre-processing steps are applied to the collected data to be compatible with the input requirements of the classification and segmentation models. After that, the performance of the models is monitored, optimised, and evaluated. Lastly, the models are applied to new, unseen ultra-widefield fundus images to segment any signs of RVO and categorise the segmented images into RVO or non-RVO.

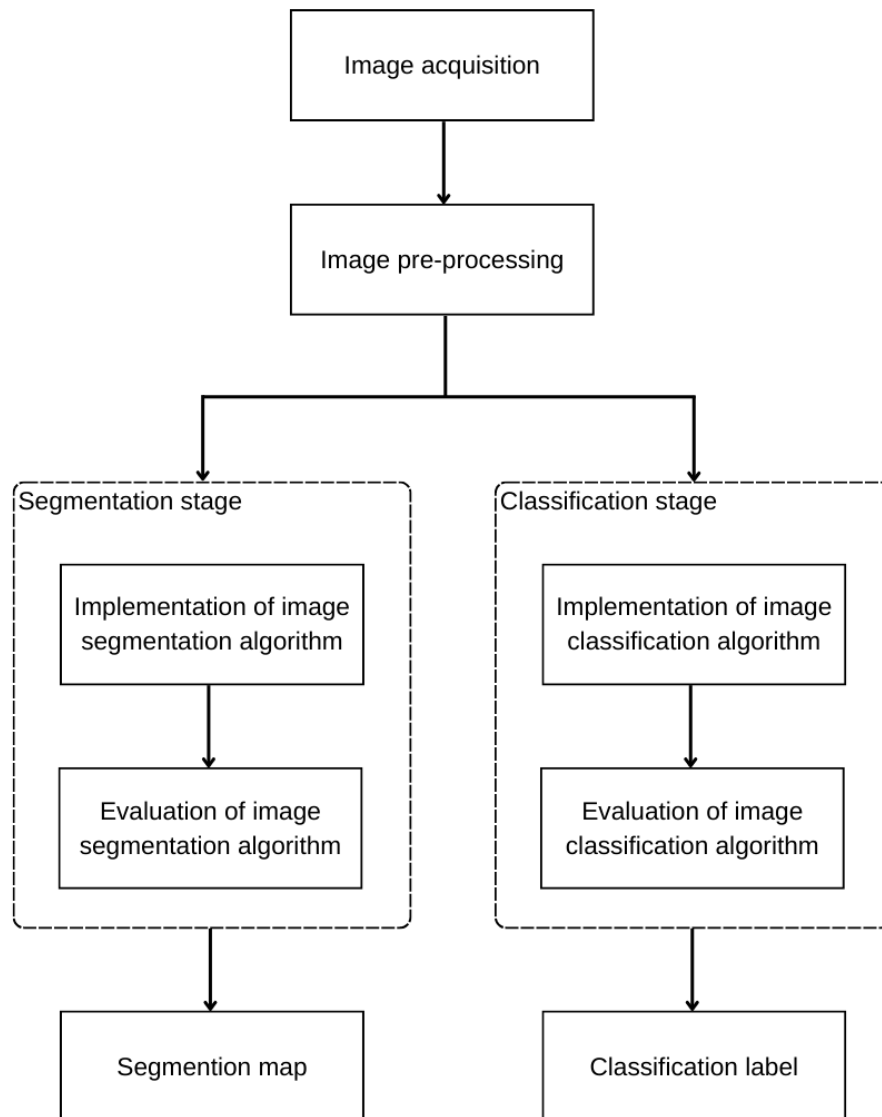


Figure 3.2: Retinal vein occlusion detection model flow chart

3.3 Software

The proposed method was completed using the free version of Google Colaboratory (Google Colab) instead of Jupyter Notebook due to insufficient resources to run locally, however, it functions similarly to Jupyter Notebook. It can execute Python code and support Python packages and libraries such as Tensorflow and Scikit-image which are often used for image processing and building deep neural networks. Working on Google Colab also provides free limited cloud storage and Graphical Processing Units (GPUs) resources. Thus,

it is suitable to use for machine learning projects which are computationally expensive and require a large storage. The Nvidia T4 with 16 gigabytes (GB) of memory, 2 CPU cores and 12 GB of Random Access Memory (RAM) are provided by the Google Colab. The simulation of the project is done on an HP 245 G8 Notebook PC with AMD Ryzen 5 5500U with Radeon Graphics, 8GB system memory, and a 64-bit Windows 11 operation system.

3.4 Image Acquisition and Verification

The dataset consists of regular and ultra-widefield colour fundus images. The regular fundus images with haemorrhage and cotton wool spots are obtained from a public dataset, the Indian Diabetic Retinopathy Image Dataset (IDRiD) (Porwal et al., 2018). It has 80 images with pixel-level annotation of haemorrhages and 40 images with cotton wool spots annotations. It is captured on a Kowa VX-10 α digital fundus camera with a 50° field of view and the pixel resolution is 4288 x 2848 and in jpg image file format. This dataset is used for training and validating a segmentation model, and the performance of this model is compared to a model trained solely using UWF images to observe the efficacy of leveraging regular fundus images in building a segmentation model for UWF images.

The UWF dataset consists of UWF colour fundus images, collected from a private eye specialist clinic. The images are taken using the Optos Daytona camera which captures a 200° field-of-view of the retina. The pixel resolution is 3082 x 2728 and of jpg image file format. There are 63 images with haemorrhage, 20 images with cotton wool spots, and 80 images without any signs. The abnormal features on the images are manually annotated for the training, validation, and evaluation of the segmentation model. The segmented UWF images with haemorrhage, cotton wool spot and without any signs are also used for evaluating the trained CNN model. The process of annotating the abnormal features begins

with identifying and marking the region of the abnormal features within the images. These annotated images are initially annotated by an expert and subsequently reviewed and verified by another expert. Figure 3.3 displays a UWF image that has been annotated by the first expert. After the initial annotation, the annotated images are sent to a second expert for further verification. If any signs have been missed or misidentified, the expert marks them on the image. Figure 3.4 shows an image that has been reviewed and verified by the second expert.



Figure 3.3: An UWF image with annotated haemorrhage by an expert. The regions of haemorrhage are marked in red colour.



Figure 3.4: An UWF image reviewed by the second expert. The yellow circles are marked by the second expert to indicate annotations that are missed.

3.5 Image Pre-processing

Pre-processing steps are implemented for both the regular and UWF datasets before the images are used for training and testing the segmentation models and classification models to enhance the performance.

3.5.1 Image Pre-processing for the Segmentation Model Trained Using Regular Dataset

The 40 regular images with cotton wool spots pixels and 80 regular images with haemorrhages are cropped to 3390 x 2848 pixels to remove excess black background and resized to 512 x 448 pixels to be compatible as the input of the U-Net model. The green channel of the images is extracted because it shows the most contrast between the feature and the background (Sharif et al., 2021). The cotton wool spots images are applied with a gamma correction of 1.60 value to adjust the brightness of the images as Equation 3.1. Then,

an unsharp mask with a radius of 1.0 pixels and weight of 0.6 is applied to sharpen the images as in Equation 3.2. An unsharp mask is done by subtracting the original image with a blurred version of the image that is obtained by applying a Gaussian blur filter. Figure 3.5 shows a regular fundus image with cotton wool spot before and after pre-processing.

$$q = \alpha^\gamma \quad \text{Equation 3.1}$$

where q is the output image, a is the input image, and γ is the coefficient.

$$q = p + a * (p - r) \quad \text{Equation 3.2}$$

where q is the output image, p is the input image, a is the amount parameter, and r is the blurred image.

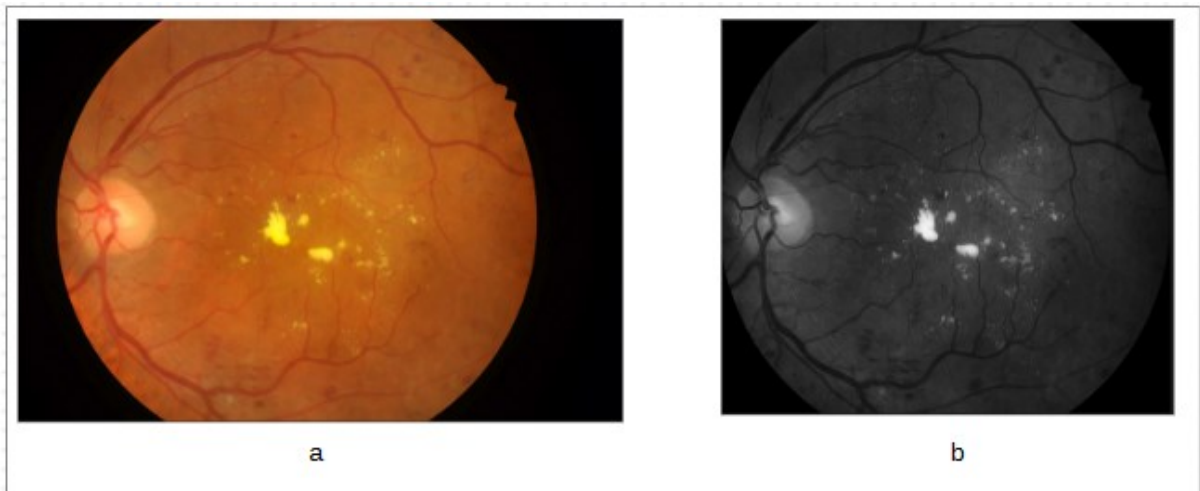


Figure 3.5: Pre-processed regular fundus image with cotton wool spot. a) Original image b) Pre-processed image

By substituting the median pixel value for each pixel that deviates by two pixels from the median, bright outliers such as exudates, optic discs, and dust are eliminated from the

extracted green channel of haemorrhage pictures. The area used to calculate the median value is 50 pixels.



Figure 3.6: Pre-processed regular fundus image with haemorrhage. a) Original image b) Pre-processed image

The UWF images for testing the segmentation model trained with regular fundus images also go through cropping, resizing, and extracting green channels. Additionally, inverse thresholding is applied to images with cotton wool spots to remove the optic disc. This replaces the bright optic disc pixels of more than 220 with the black pixels by setting the pixel to 0. The optic disc is a bright artefact that could be mistaken as the cotton wool spots due to its bright pixel.

3.5.2 Image Pre-processing for the Segmentation Model Trained Solely Using UWF Dataset

The ultra-widefield (UWF) images are cropped to eliminate redundant black backgrounds, eyelids, and eyelashes using an approach inspired by Ali (2019). It is done by subtracting the original image with an ellipse binary mask. The shape of the ellipse mask is determined by $(x_0 + 350, y_0 + 400, x_1 - 600, y_1 - 1000)$, where (x_0, x_1) and (y_0, y_1) is

the coordinates of where the colour pixels start along the x and y axes. This approach excludes the black background, and more cropping is done to remove excessive eyelashes and eyelids thus the input image only includes the retina. After that, the subtracted image is resized to 512 x 448 pixels. For the UWF images with cotton wool spots, the intensity of the image is rescaled to a range of 0 to 200 pixels for higher brightness and contrast. While the UWF images with haemorrhage are rescaled to a range of 0 to 150. Figure 3.5 depicts the UWF image before and after applying the pre-processing steps. Additionally, UWF testing images with cotton wool spots are applied with a gamma correction of 0.8 to further increase brightness for better segmentation.

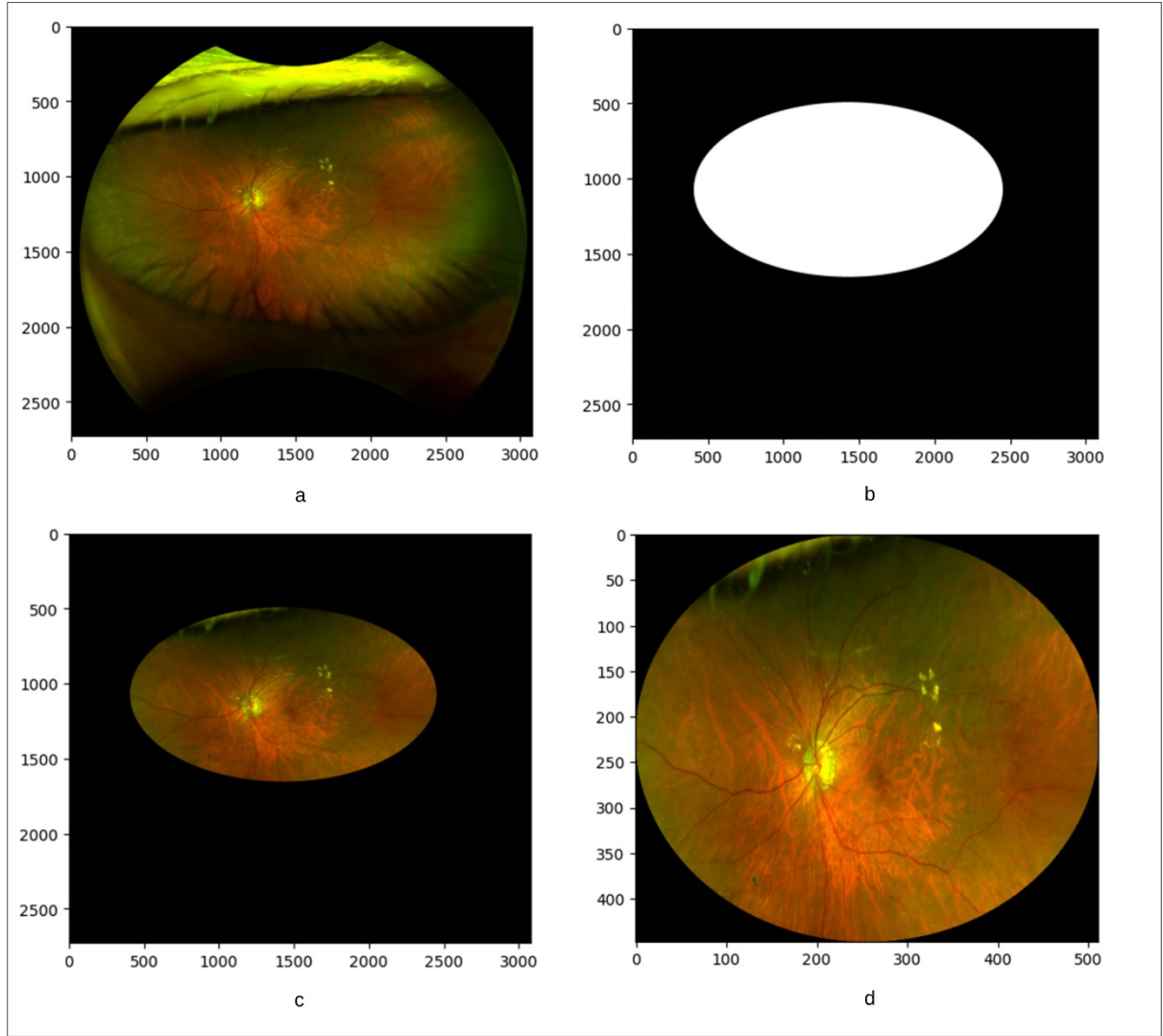


Figure 3.7: Overall pre-processing steps of UWF retinal fundus images with cotton wool spots. a) Original image b) Generated ellipse mask c) Subtracted image d) Resized and rescaled image

3.5.3 Image Pre-processing for the Segmentation Model Trained Using Transfer Learning

The transfer learning approach involves pre-training the segmentation model using regular images, and the weights are saved and loaded to train the UWF images. Firstly, the 40 regular images with cotton wool spots pixel and 80 regular images with haemorrhages are cropped to 3390 x 2848 pixels to remove the excess black background and resized to a

standard size of 512 x 448. Then the images are rescaled to a range of 0 to 150 to increase brightness and contrast.

The UWF images are also cropped to eliminate redundant black backgrounds, eyelids, and eyelashes. It is done by subtracting the original image from the generated ellipse binary mask. The shape of the mask is $(x_0 + 350, y_0 + 400, x_1 - 600, y_1 - 1000)$, where (x_0, x_1) and (y_0, y_1) is the coordinates of where the colour pixels start along the x and y axes. This step removes the black background and crops excessive eyelashes and eyelids. After that, the subtracted image is resized to 512 x 448 pixels. For the UWF images with haemorrhage, the intensity of the image is rescaled to a range of 0 to 150 pixels for higher brightness and contrast. While the UWF images with cotton wool spots are rescaled to a range of 0 to 200.

The UWF testing images are cropped, resized, and rescaled intensity accordingly. The images with cotton wool spots are also applied with a gamma correction of 0.8 to increase brightness.

3.5.4 Image Pre-processing for the Classification Model

The UWF ground truths of no-signs, haemorrhage and cotton wool spots are cropped and resized accordingly as the corresponding UWF images used in the segmentation model trained solely with UWF images.

3.6 Image Post-processing

There is a post-processing step when evaluating all the segmentation models for haemorrhages. The segmented images generated by the trained U-Net models for

haemorrhages are applied with morphological opening with a kernel size of 3 x 3 to remove noise.

3.7 Data Augmentation

The four sets of data transformations include translation of -35.5° along the x and y axes, flipping horizontally, flipping horizontally and rotation of 30° , and flipping horizontally, rotation of 30° , and translation of -15 along the x and y axes for all regular fundus images and their corresponding ground truths for the segmentation model trained using regular dataset. Thus, there are 400 images from the input of 80 images with haemorrhages and a total of 200 images with cotton wool spots from the input of 40 images.

The data transformation applied to UWF for training are flipping horizontally, rotation of 30° , and translation of 30 along the x and y axes. Then, there is a total of 192 UWF with haemorrhage and 60 images with cotton wool spots.

The regular and UWF fundus images used for transfer learning are applied with data transformation by flipping horizontally, rotation of 30° , and translation of 30 along the x and y axes. There is a total of 320 regular images and 160 regular images with haemorrhage and cotton wool spots, and 192 UWF images and 60 UWF with haemorrhage and cotton wool spots.

3.8 Implementation

After the data is prepared, the segmentation model with U-Net architecture is trained with the regular fundus images of 512 x 448 x 3 pixels for haemorrhage and cotton wool spot segmentation. The model is trained with a batch size of 2, binary focal cross entropy for the loss function, Adam optimizer, and learning rate of 0.0001. The model is trained and

saved at the epoch where the loss value is the lowest among the 200 epochs. The number of convolutional blocks and filters that achieve the highest performance are explored to decide on the best model that will be compared with the segmentation model trained on UWF images.

For comparison, a segmentation model was trained using UWF images of 512 x 448 x 3 pixels with a batch size of 2, binary focal cross entropy for the loss function, Adam optimizer, and learning rate of 0.0001. The model with U-Net architecture has also trained a maximum of 200 epochs for haemorrhage and cotton wool spots. The most suitable number of convolutional blocks and filters for each model is also decided by experimental findings.

The segmentation model trained using transfer learning is about using a pre-trained model on a related problem. It uses the knowledge of the segmentation model that has been learnt from a larger annotated dataset which is the regular dataset. The segmentation model is first trained with regular images of 512 x 448 x 3 pixels with a batch size of 2, binary focal cross entropy for the loss function, Adam optimizer, and learning rate of 0.0001. The architecture has 5 convolutional blocks, the number of filters in the first block is 64 and it doubles in the following block. It is trained with a maximum of 150 epochs with early stopping that monitors the lowest validation loss starting from 40 epochs and the trained model is saved. Then the trained model is loaded to train on UWF images with the same hyper-parameters for segmentation of haemorrhage and cotton wool spot.

For the CNN classification model, the input is the ground truths of 512 x 448 pixels. The loss function used is binary cross entropy with Adam optimizer and a learning rate of 0.0001. The model is trained up to 50 epochs and the epoch with the lowest loss value is

saved. The saved segmentation and classification models are then used to evaluate the UWF images for performance analysis.

3.9 U-Net Architecture

The model used for image segmentation is the widely recognised, U-Net architecture. It is a convolutional neural network model specifically for biomedical image segmentation developed by Ronneberger et al. (2015). The model has an encoder for the contracting path, a decoder for the expanding path and a bottleneck which connects the paths forming a ‘U’ shaped architecture as shown in Figure 3.3. The contracting path is where the downsampling and feature extraction take place.

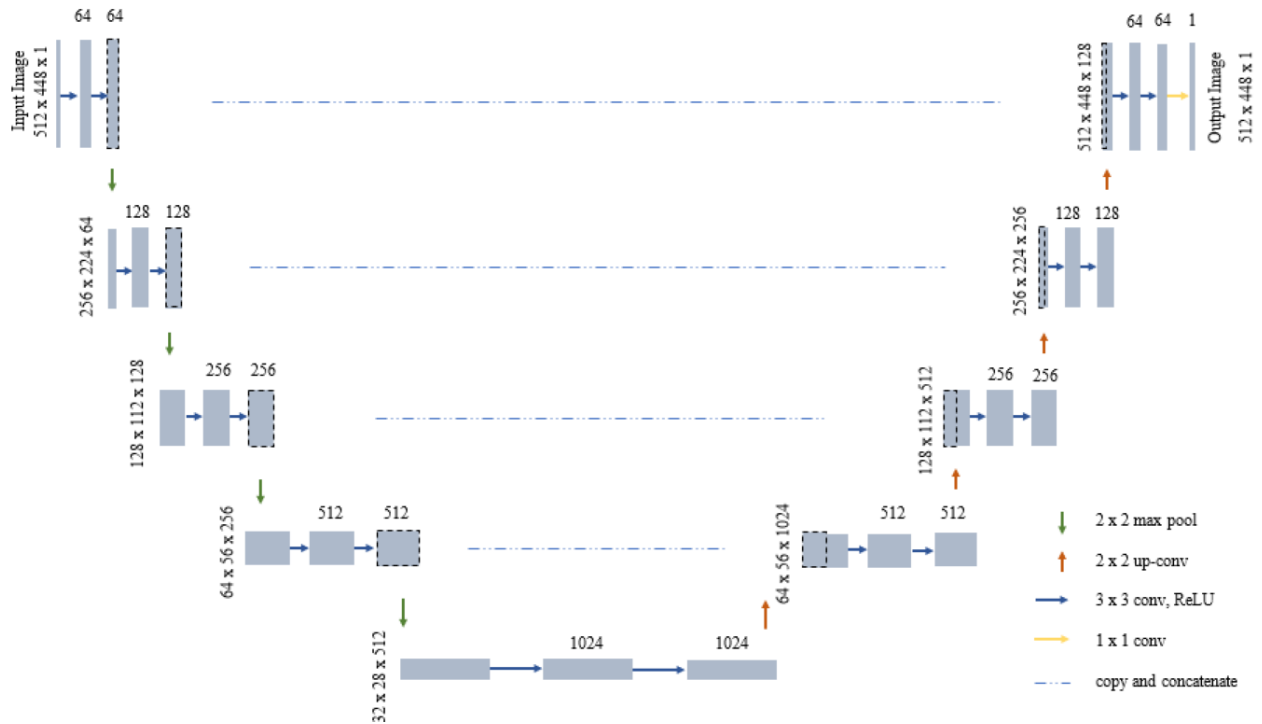


Figure 3.8: U-Net architecture. U-Net architecture with input image resolution of 512 x 448 x 1. The blue box represents a feature map with its number of filters on top of the box. The black dotted box represents copied feature maps. The coloured arrows signify different operations.

Generally, there are 4 contracting blocks and 4 expanding blocks, but the number of blocks can be increased or decreased. Each block is made up of two 3×3 convolution layers followed by a rectified linear unit (ReLU) activation function with batch normalisation. The number of filters is doubled after each block. For example, if the first block has 64 convolutional filters, then the second block has 128 convolutional filters.

After that, the feature map goes through a 2×2 max pooling layer with a stride of 2 for downsampling. After downsampling, the resolution of the feature map is halved. For example, assume the input image has a resolution of $512 \times 448 \times 1$ and it goes through the first contracting block, then the resulting feature map has a resolution of $256 \times 224 \times 64$. This process is done 4 times if there are 4 contracting blocks. After 4 downsampling steps, it reaches the bottleneck which is composed of two 3×3 convolutional filters with ReLU function.

For the expanding path, there is upsampling by a 2×2 transposed convolution layer. For instance, the feature map from the last contracting block is $32 \times 28 \times 1024$ then it up-samples to $64 \times 56 \times 1024$. After upsampling, there is a skip connection to concatenate between the up-sampled feature map and the feature map from the encoder at the same level which has the same resolution. Then, it goes through two 3×3 convolution layers with ReLU activation function and batch normalisation. The number of convolutional filters is half of the previous filters. For instance, if the previous number of filters of the last contracting block is 1024, then the number of filters of the first expanding block would be 512.

After that, it goes through up-sampling again. This process is repeated 4 times if there are 4 expanding blocks. After the last expanding block, there is one 1×1 convolution layer where the output of this layer will be the same resolution as the input image, for instance,

512 x 448 x 1. The output of this layer produces the segmented target features that are in the input image, for instance, segmented haemorrhages.

3.10 CNN Architecture

The second part of the framework is to build a classification model for categorising the segmented images into RVO or non-RVO based on the features segmented by the segmentation model. A custom CNN model is constructed and trained for this task. Table 3.1 shows the architecture of the CNN model. It has three 3 x 3 convolution layers and the number of filters in the first layer is 8 and it doubles in the subsequent layer. The convolution layer is followed by ReLU activation and max pooling of 2 x 2. Then, the pooled feature map is flattened and goes through a dense layer with ReLU and a dense layer of sigmoid activation to get the output.

Table 3.1: The architecture of the proposed CNN model

Architecture
Input
3 x 3 convolution, 8 filter
2 x 2 max-pooling
3 x 3 convolution, 16 filter
2 x 2 max-pooling
3 x 3 convolution, 32 filter
2 x 2 max-pooling
Flattening to 1-dimensional vector
Fully connected dense layer of 32 with ReLU activation
Final dense layer with sigmoid activation

3.11 Performance Evaluation

To evaluate the performance of the models, the following 4 values need to be calculated to compute the evaluation metrics:

1. True positives (TP) are the number of correctly classified pixels.
2. False positives (FP) are the number of pixels that are incorrectly classified into the positive class.
3. True negatives (TN) are the number of pixels that are correctly classified into the negative class.

4. False negatives (FN) are the number of pixels that are incorrectly classified into the negative class.

Intersection over Union (IoU) is a metric commonly used for object detection and object segmentation. It measures the overlap between the ground truth mask and the predicted mask in Equation 3.3 and Equation 3.4. The value lies between 0 and 1, 0 means the model could not correctly segment any pixels of the target feature while 1 means the model could correctly segment all the pixels of the target feature. IoU is used for comparing the segmentation of U-Net on UWF images with the method in the study by Ronneberger et al. (2015).

$$IoU = \frac{\text{Area of Intersection}}{\text{Area of Union}} \quad \text{Equation 3.3}$$

$$IoU = \frac{TP}{TP+FP+FN} \quad \text{Equation 3.4}$$

The performance of the models is monitored based on their accuracy, sensitivity, and specificity. Accuracy is the proportion of correctly classified pixels as Equation 3.5. Equation 3.6 represents the likelihood of actual positives, which is known as sensitivity, whereas specificity represents the probability of actual negatives as Equation 3.7. Precision is the ratio of correctly predicted positive pixels among the total number of positive pixels predicted as Equation 3.8.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad \text{Equation 3.5}$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad \text{Equation 3.6}$$

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

Equation 3.7

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

Equation 3.8

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

In this chapter, the results of implementing the proposed method are presented and discussed. The training and validation of the segmentation model with U-Net architecture and classification model with CNN architecture are recorded. Then, the models are evaluated on the UWF dataset, and the performances are compared with existing deep learning methods.

4.2 Results of the Segmentation Model Trained Using Regular Fundus Images

This study also investigates the effectiveness of using regular fundus images to build a segmentation model that can segment UWF images because annotated public datasets of UWF images are scarce and there is a publicly available dataset of regular fundus images annotated with retinal lesions. Thus, it could be difficult to acquire UWF images suitable for research purposes. The highest-performing segmentation model is chosen to compare its performance with the segmentation model trained using UWF images. Before that, the segmentation model is trained with a varying number of convolutional filters, and their results are compared to select the best model.

Table 4.1 presents the validation IoU scores of 3 models with varying numbers of convolutional blocks and filters. Model 3 with 5 convolutional blocks and 64 filters in the first block achieved the best validation IoU score which is 0.9034 for segmentation of haemorrhage and 0.9308 for cotton wool spot.

Table 4.1: Validation result of the segmentation model with different numbers of convolution blocks and filters

Model	No. of convolutional blocks and filters	Validation IoU score	
		Segmentation of haemorrhages	Segmentation of cotton wool spots
1	64, 128, 256	0.8038	0.9079
2	64, 128, 256, 512	0.8836	0.9273
3	64, 128, 256, 512, 1024	0.9034	0.9308

The segmentation models are then trained with 5 convolutional blocks. Based on the evaluation metrics in Table 4.2, the model cannot segment haemorrhage and cotton wool spots satisfactorily. The average IoU scores for the segmentation of cotton wool spot and haemorrhage were 0.5904 and 0.5744, respectively. Figure 4.1 and Figure 4.2 show the segmentation of haemorrhage and cotton wool spot on a testing UWF image which scored an IoU of 0.5798 and 0.6657. The segmentation of haemorrhages shows the misidentification of darker pixels around the excess eyelids and eyelashes. Whereas the region of the cotton wool spot is interfered by the excess eyelashes.

Table 4.2: Evaluation results of segmentation model trained using a regular dataset on UWF testing images

	IoU	Sensitivity (%)	Specificity (%)
Haemorrhages	0.5744	53.59	51.85
Cotton wool spots	0.5904	40.17	39.95

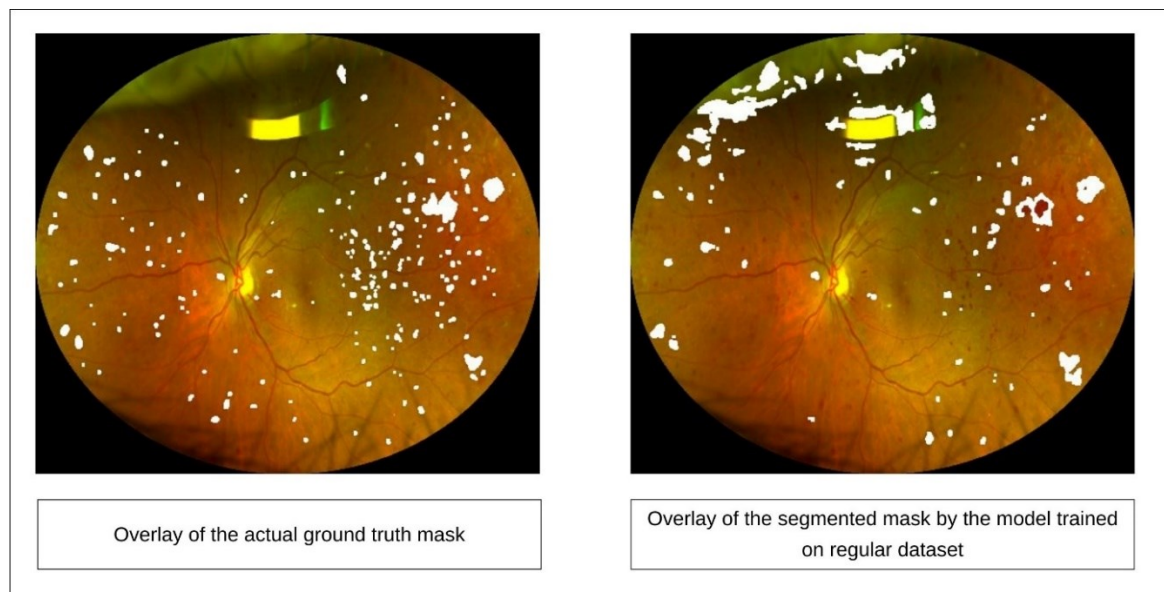


Figure 4.1: Segmented haemorrhages by the model trained using a regular dataset on testing UWF image compared to the actual ground truth mask



Figure 4.2: Segmented cotton wool spots by the model trained using a regular dataset on a testing UWF image

4.3 Results of the Segmentation Model Trained Using UWF Fundus Images

The most suitable number of convolutional blocks is chosen by experimental findings. The model with the optimal number of filters is chosen through a 3-fold cross-validation for cotton wool spot and a 4-fold cross-validation for haemorrhage. The validation IoU scores of three models with different numbers of convolutional blocks and filters are displayed in Table 4.3. Based on the results, the greatest validation IoU score for segmentation of haemorrhages is 0.8086 using 4 convolution blocks. Whereas the model with 3 convolutional blocks achieved the highest validation IoU of 0.7322 for cotton wool spots.

Table 4.3: Validation result of the segmentation model with different numbers of convolution blocks and filters

Model	No. of convolutional blocks and filters	Validation IoU score	
		Segmentation of haemorrhages	Segmentation of cotton wool spots
1	64, 128, 256	0.7516	0.7322
2	64, 128, 256, 512	0.8086	0.7241
3	64, 128, 256, 512, 1024	0.7657	0.7098

Consequently, the most optimal number of filters is used to train the segmentation model for haemorrhages and cotton wool spots. The batch size, learning rate, and lost function have been set as shown in Table 4.4. The trained models are then used to segment the testing UWF images to evaluate their performance.

Table 4.4: Hyper-parameter of the U-Net model

Parameter	Value
Batch size	2
Learning rate	0.0001
Loss function	Binary focal cross-entropy

Table 4.5 shows the evaluation results of the segmentation model trained on the UWF dataset on UWF testing images. Based on the results, the U-Net model can segment haemorrhage and cotton wool spots relatively accurately with an average IoU score of 0.8795 and 0.8235. Figure 4.3 and Figure 4.4 show a sample of segmentation of

haemorrhages and cotton wool spots by the U-Net model trained using UWF image which scored an IoU of 0.9324 and 0.8418 for the segmentation of haemorrhage and cotton wool spot respectively.

Table 4.5: Evaluation results of segmentation model trained using UWF dataset on UWF testing images

	IoU	Sensitivity (%)	Specificity (%)
Haemorrhages	0.8795	83.32	99.93
Cotton wool spots	0.8235	83.85	99.97

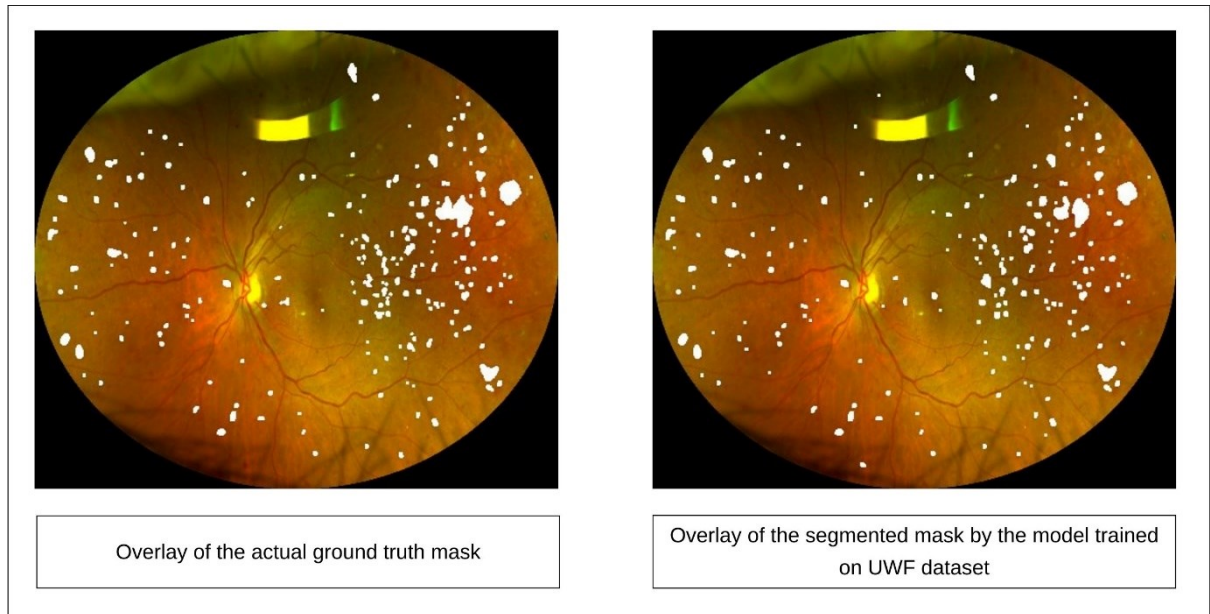


Figure 4.3: Segmented haemorrhages by the model trained using the UWF dataset on a testing UWF image

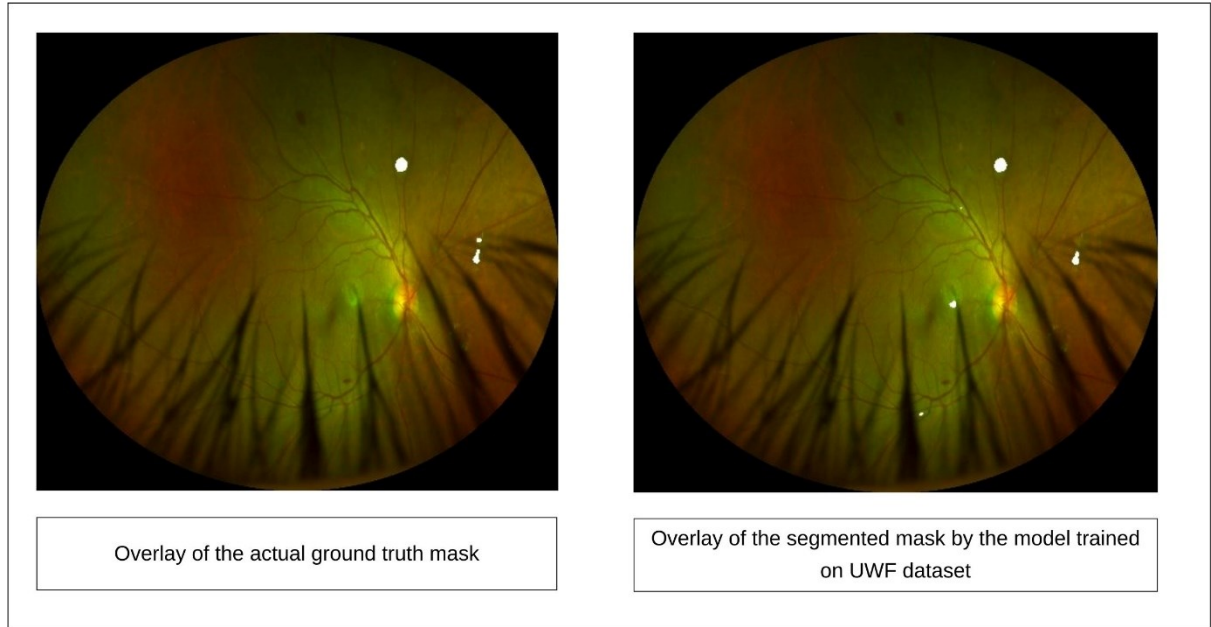


Figure 4.4: Segmented cotton wool spots by the model trained using the UWF dataset on a testing UWF image

4.4 Results of the Segmentation Model Using Transfer Learning Approach

A segmentation model is trained with a transfer learning approach where a segmentation model trained solely using regular fundus is trained and the pre-trained weights are loaded as the starting point when training the UWF images. The final model is saved for evaluating the UWF testing images to investigate the effectiveness of using regular fundus images and a transfer learning approach. Table 4.6 depicts the results of the segmentation model trained using transfer learning on UWF testing images.

Table 4.6: Evaluation results of segmentation model trained using transfer learning on UWF testing images

Lesion	IoU	Sensitivity (%)	Specificity (%)
Haemorrhages	0.7833	64.32	99.80
Cotton wool spots	0.7842	80.95	99.95

The model trained using transfer learning achieved an IoU of 0.7833 for haemorrhages and achieved an IoU of 0.7832 for cotton wool spots. Figure 4.5 and Figure 4.6 shows an example of segmentation of haemorrhage and cotton wool spot which scored IoU of 0.8111 and 0.7595. The prediction is better than the segmentation model trained using a regular dataset but slightly lower than the model trained using the UWF dataset.

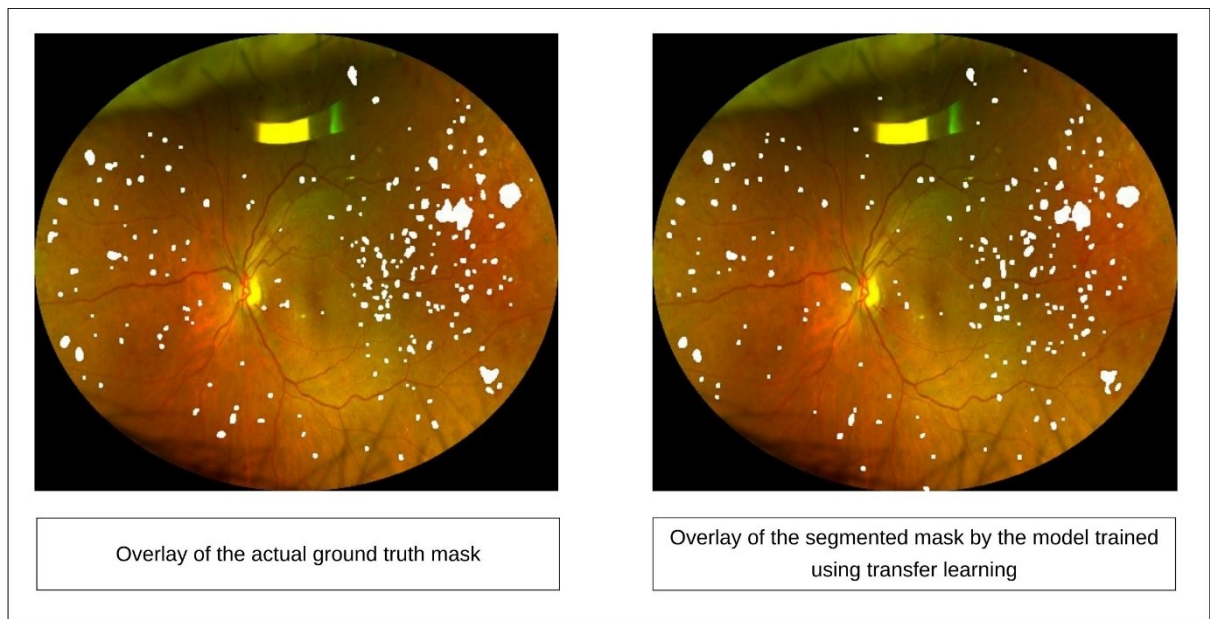


Figure 4.5: Segmented haemorrhages by the model trained using transfer learning on a testing UWF image

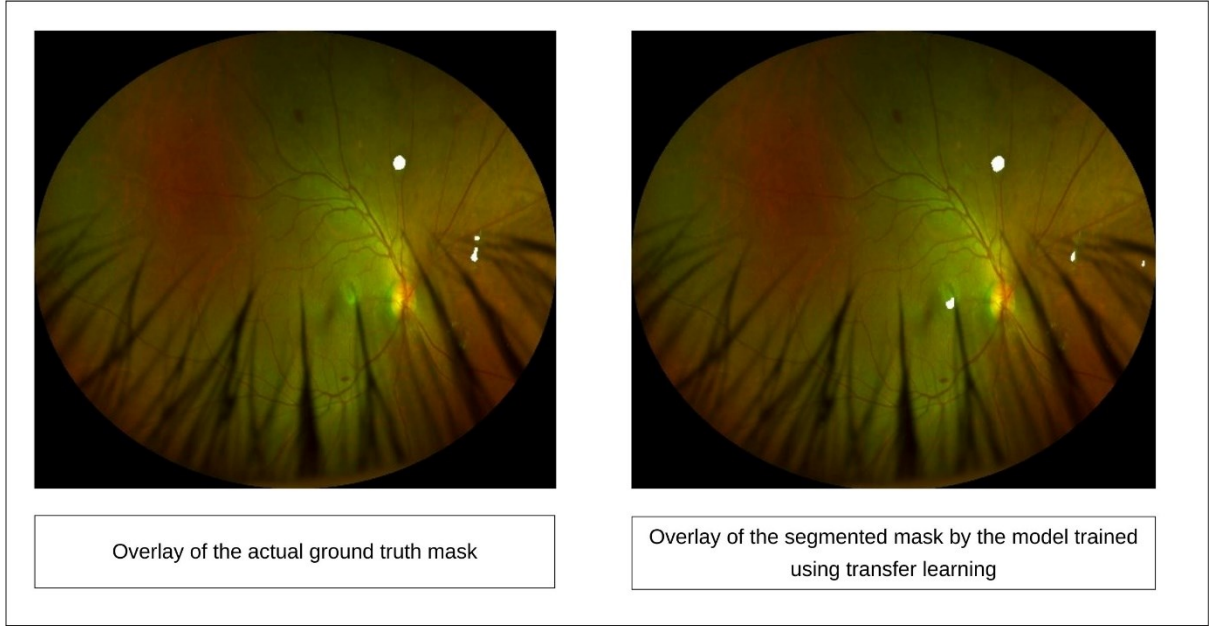


Figure 4.6: Segmented cotton wool spots by the model trained using transfer learning on a testing UWF image

4.5 Discussion and Comparison of the Segmentation Models of this Study

The selection of the appropriate number of filters is not governed by fixed rules, as it relies on factors such as the dataset, fine-tuning, and experimentation. Adding the number of filters can enhance performance to a certain extent, as the number of convolutional filters determines the generation of feature maps, but this requires additional training resources. The performance of the model may be enhanced due to a greater number of feature maps, as they capture relevant features from input images. Consequently, a higher number of feature maps aids in capturing intricate and abstract features, thereby reducing information loss. Although increasing filters do not directly address information loss during downsampling, it enables the capture of more representations.

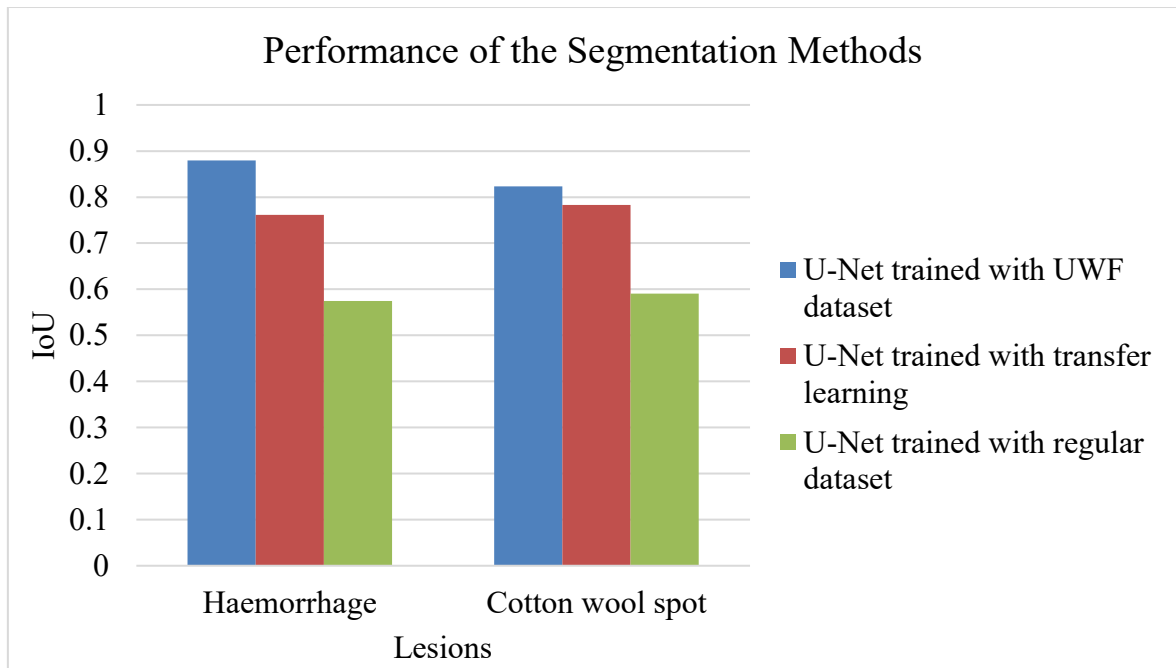
Based on the experimentation of the segmentation model trained with UWF images, it was discovered that employing three convolutional blocks for the segmentation of cotton

wool spots and four convolutional blocks for haemorrhages with 64 filters in the first block is suitable for both segmentation models. Further increase in the number of convolutional blocks results in a lower validation score. Lower performance of the model could be caused by overfitting because the model has more capacity to learn the training images and becomes over-generalized to the training images. It could also have reached a point where a greater number of blocks does not improve the performance anymore because it could be extracting noise instead of relevant features from the data. Moreover, the most optimal number of convolutional blocks for the segmentation model trained using a regular fundus dataset is five. This proves that the most optimal number of convolutional blocks and filters varies depending on the dataset and the task thus experiments and fine-tuning can help achieve the best model.

The evaluation results of the segmentation model trained using UWF datasets are the highest compared to the segmentation models trained with other methods of this study as shown in Table 4.7. Figure 4.7 depicts the comparison of the three segmentation models based on the IoU score. Thus, it is the most effective method, and it is used for the feature extractor before the classifier. Evaluation of unseen UWF images by the model trained using UWF images demonstrates the model's ability to effectively detect haemorrhages and cotton wool spots, even when both signs are present in the image, along with other features such as the optic disc, exudates, blood vessels, and laser marks. This could be attributed to the testing images originating from the same source as the training images. Consequently, the model has learned to extract similar features from the training images, enabling better generalization on testing images with comparable patterns and variations.

Table 4.7: Comparison of the segmentation models of this study

Method	Lesion	IoU	Sensitivity (%)	Specificity (%)
U-Net trained with UWF dataset	Haemorrhages	0.8795	83.32	99.93
	Cotton wool spots	0.8235	83.85	99.97
U-Net trained with transfer learning	Haemorrhages	0.7612	64.32	99.98
	Cotton wool spots	0.7832	80.95	99.95
U-Net trained with a regular dataset	Haemorrhages	0.5744	53.59	51.85
	Cotton wool spots	0.5904	40.17	39.95

**Figure 4.7:** Comparison of the IoU score of the segmentation methods

However, there are a few instances where darker regions around excess eyelashes, shadow and laser marks are identified as haemorrhages due to their darker pixels. Figure 4.8 shows a testing UWF image that scored an IoU of 0.6422 where laser marks are misidentified as haemorrhages by the model trained on the UWF dataset. It could be that fewer training images have laser marks along with the RVO lesions. Thus, a larger dataset with varying cases of fundus images could improve the performance. Additionally, bright pixels are occasionally misidentified as cotton wool spots, such as bright lesions in the images. Figure 4.9 shows the testing UWF where bright lesions on the image are misinterpreted as cotton wool spots by the model trained on UWF. This image scored an IoU of 0.7898.

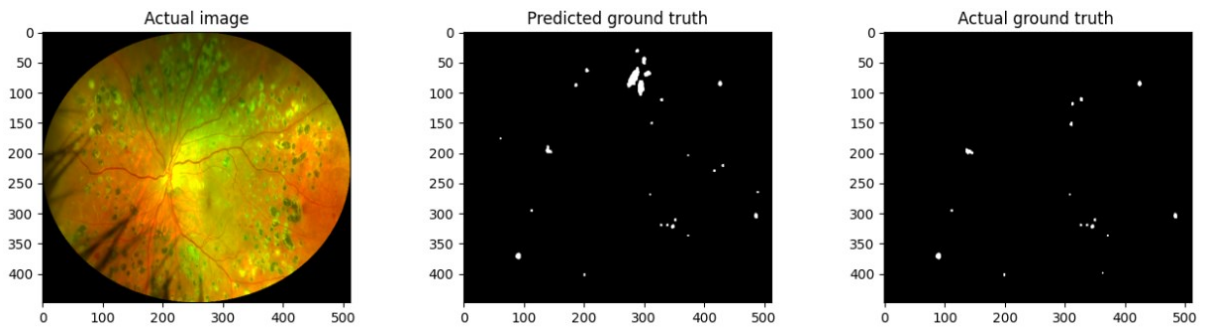


Figure 4.8: Predicted and actual ground truth of haemorrhages for a UWF image by the model trained using the UWF dataset

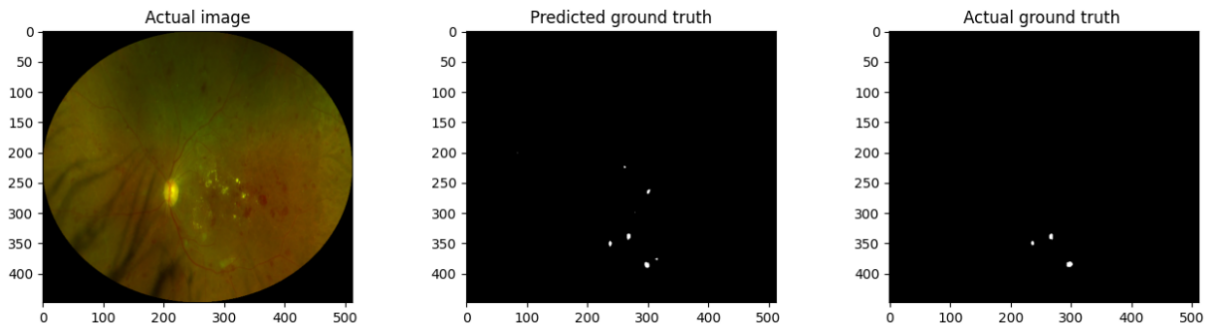


Figure 4.9: Predicted and actual ground truth of cotton wool spots for an UWF image by the model using the UWF dataset

The transfer learning approach shows improvement in the performance of segmenting the lesions. It is beneficial when there is a limited annotated dataset of UWF images available for training. The segmentation model can be trained with a more readily available and larger annotated regular dataset for training and use the pre-trained weights as the starting point for solving similar problems with smaller datasets. It allows faster convergence and better performance by replacing the weights initialisation with the pre-trained weights. From the study, the model trained with transfer learning performed significantly better than the model trained on solely regular datasets. Using the model trained on the solely regular dataset, most of the testing images had misidentification of darker pixels around the excess eyelids and eyelashes for the segmentation of haemorrhages. Segmenting cotton wool spots is interfered with by the excess eyelashes and other bright lesions. However, there are fewer instances of misidentification of haemorrhages around darkened regions and cotton wool spots among other bright lesions using the transfer learning approach. Although the performance fell short compared to the segmentation model trained on the UWF dataset. It is expected that the model can be improved using a larger dataset as usually thousands of data are required to train a neural network.

4.6 Discussion and Comparison of Segmentation Model with Existing Methods

For evaluation of the proposed method, the evaluation results of the segmentation model trained on UWF datasets are the highest compared to the segmentation models trained with other methods of this study. Thus, it is compared with existing methods that focus on segmenting features associated with retinal diseases, retinal fundus images, and the U-Net-based model. Table 4.8 shows the comparison of the segmentation model trained on UWF datasets with other studies. The sensitivity score refers to the ability of the segmentation model to correctly identify the target signs of the proposed method. The sensitivity score for

the segmentation of haemorrhage and cotton wool spots is above 80%, which shows that the model can segment the target signs.

As compared to a study that also evaluated on local dataset, the sensitivity score for haemorrhages and cotton wool spots segmentation is similar to Wan et al. (2021). Although the model by Wan et al., (2021) showed that it can segment features with distinct features, they noticed that smaller target signs are harder to segment accurately such as microaneurysm and exudates, and some lesions naturally appear blurry with no clear boundary. Besides that, microaneurysms and blood vessels are sometimes segmented as haemorrhages. A few factors that could result in this situation are the nature of medical images, the environment, the imaging device, and the imbalanced dataset.

A study by Chen et al. (2021) proposed using DeepLab-v3 which had the best performance and generalised better when compared to the other segmentation models they tested. The model scored a sensitivity score of 88% for the segmentation of flame-shaped haemorrhages and 82% for cotton wool spots on a test set of 25 flame-shaped haemorrhages and 11 cotton wool spot images. Whereas this study scored a sensitivity score of 83.32% for haemorrhages and 83.85% for cotton wool spots. The sensitivity score of Chen et al. (2021) is slightly higher than this study for haemorrhages segmentation but their study only included flame-shaped haemorrhages whereas the UWF dataset used for this study contained different types of haemorrhages such as pre-retinal haemorrhages, dot and blot, and flame-shaped haemorrhages and included images with laser marks and varying image quality which reflect the real-world data. They also had more testing images to evaluate the model and used higher-quality images.

Kaur and Kaur (2022) scored a slightly higher sensitivity score than the proposed method on segmentation of haemorrhage and cotton wool spot using a test set of 8 haemorrhage images and 8 cotton wool spot images. However, the validation IoU obtained from their proposed method was 0.8236 for the segmentation of haemorrhages and 0.7517 for cotton wool spots on the standard dataset, the IDRiD dataset. The U-Net model in this study scored higher validation IoU with 0.8714 for segmentation of haemorrhage and 0.8323 for cotton wool spot. Kaur and Kaur (2022) evaluated their model by testing images from the standard dataset resulting in better segmentation results. Moreover, comparing the validation IoU score, the proposed method scored higher for the segmentation of haemorrhages than Kaur and Kaur (2022), but the best validation IoU score for the segmentation of cotton wool spots is 0.9022 which is achieved by the DIARETDB1 dataset. Howbeit, Abdelmaksoud et al. (2021) mentioned that the DIARETDB1 dataset does not provide precise annotation of the signs because the ground truth is annotated by marking the centre of a sign and drawing a circle around it to cover the whole region of the sign, so it is not considered pixel-level annotation and not recommended for training purpose. Furthermore, the result from Abdelmaksoud et al. (2021) using the U-Net-based model that was trained on the IDRiD dataset scored perfect sensitivity scores on the IDRiD dataset and above 90% for the DIARETDB1 dataset.

Besides that, the proposed method scored higher IoU than the original paper on the human cancer cell dataset, the DIC-HeLa dataset. Even though the target signs are different, it is more difficult to segment haemorrhage and cotton wool spots since the data are retinal fundus images which present the human anatomy. There can be individual variations that should be considered, and the model must identify the location of the abnormalities amongst various physiological structures such as the optic disc, macula, blood vessels, and other

lesions that are present when the images are taken. Table 4.9 shows a comparison between the proposed method and Ronneberger et al. (2015).

Table 4.8: Comparison of evaluation results for segmentation model with existing methods

Authors	Retinal images	Methods	Dataset	Sensitivity (%)	Specificity (%)
Wan et al.	Regular fundus	U-Net with an attention module	Local dataset		
			Haemorrhages	83.59	99.95
			Cotton wool spots	84.92	99.99
Chen et al.	Regular fundus	DeepLab-v3	Private dataset		
			Haemorrhages	88.00	75.00
			Cotton wool spots	82.00	70.00
		Lesion-Net-8s	Haemorrhages	84.00	48.00
			Cotton wool spots	100.00	45.00
		DANet	Haemorrhages	84.00	59.00
			Cotton wool spots	82.00	56.00
		FCN-32s	Haemorrhages	84.00	68.00
			Cotton wool spots	55.00	64.00

Table 4.8 continued

Kaur and Kaur	Regular fundus	U-Net with pre-trained Inceptionv3 backbone	IDRiD		
			Haemorrhages	88.39	99.94
			Cotton wool spots	85.38	99.71
			DIARETDB1		
			Haemorrhages	81.00	99.87
			Cotton wool spots	73.64	99.00
Abdelmaksoud et al.	Regular fundus	U-Net	IDRiD		
			Haemorrhages	100.00	98.00
			Cotton wool spots	100.00	100.00
			DIARETDB1		
			Haemorrhages	95.00	97.00
			Cotton wool spots	97.00	96.20
This study	UWF fundus	U-Net trained on the UWF dataset	Private dataset		
			Haemorrhages	83.32	99.93
			Cotton wool spots	83.85	99.97

Table 4.9: Comparison of the segmentation model with Ronneberger et al.

Method	Dataset used	IoU
Ronneberger et al.	DIC-HeLa (HeLa cells on a flat glass)	0.7756
Proposed method	UWF dataset	
	Haemorrhages	0.8795
	Cotton wool spots	0.8235

4.7 Results of the Classification Model

The CNN model is trained and tested with the ground truth of normal, haemorrhages and cotton wool spots. The batch size, learning rate, and epochs have been set as shown in Table 4.10. The training and validation process is run up to 50 epochs and the lowest validation loss value recorded is at the 50th epoch with a loss value of 0.28486. At the 50th epoch, the validation accuracy, precision, and recall achieved 100%. The evaluation result of the classification model on the testing subset achieved 100% accuracy, sensitivity, and specificity as shown in Table 4.11. This indicates that the model could correctly identify all cases of RVO and non-RVO among the 20 test images (10 RVO and 10 non-RVO). While these results are promising, this study uses a relatively small dataset size. Nevertheless, the performance of the classification model shows that using the segmentation model as the feature extractor is very effective due to the representation of the relevant features in detecting the presence of RVO.

Table 4.10: Parameters of the CNN model

Parameter	Value
Epochs	50
Optimizer	Adam
Learning rate	0.0001
Loss function	Binary cross entropy

Table 4.11: Evaluation results on the classification of RVO and non-RVO on the UWF testing set

Accuracy (%)	Precision (%)	Sensitivity (%)
100	100	100

4.8 Discussion and Comparison of the Classification Model with Existing Methods

For evaluation results for classification, Table 4.12 compares with existing methods that classified retinal vein occlusion using CNN models. The segmentation map obtained from the U-Net model that is classified using the CNN classification model into RVO or non-RVO class achieved the best results as compared to other existing methods. It proved that the trained U-Net model is great for feature extraction because the classification model is trained on segmented images by U-Net and it had no misclassification when evaluating the segmented map of the UWF dataset. Since this is an image-based classification, the image is input into the model as it is, and the model learns the relevant features related to the classification class. Thus, it is helpful that the input image had prior feature extraction to remove noises before training the CNN classification model.

Table 4.12: Comparison of evaluation result for classification model with existing methods

Author	Class	Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
Choudhury et al.	Central retinal vein occlusion (CRVO) and non-CRVO	Custom CNN	97.56	100	96.97
Nagasato et al.	Branch retinal vein occlusion (BRVO) and non-BRVO	Custom CNN	-	94	97
Chen et al., 2021	BRVO, CRVO, non-RVO, and normal	ResNet-50	-	70	89
		Inception-v3	-	81	90
		DenseNet-121	-	71	86
		SE-ResNeXt-50	-	75	88
Proposed method	Retinal vein occlusion (RVO) and non-RVO	Custom CNN	100	100	100

CHAPTER 5

CONCLUSION

5.1 Main Findings

Rather than image-based disease classification, the diagnosis of retinal diseases requires exact localisation of the signs present on retinal images. This was achieved by using deep learning and image processing techniques. The proposed method is a lesion-based approach where the U-Net model segments haemorrhage and cotton wool spots, and the CNN model gives the classification of RVO. Regular fundus images are leveraged to build the segmentation model for ultra-widefield images, and the results showed that regular fundus images were not satisfactory, but they could be improved. The transfer learning approach proved to perform significantly better than using regular fundus images. The segmentation model used the easily accessible regular dataset to train, and the pre-trained weights were used to train the segmentation model for UWF images which is a similar task with smaller datasets.

The proposed method relies on pre-processing procedures for the ultra-widefield fundus images to remove unnecessary features such as redundant black background, excess eyelashes, and eyelids, and enhance the target features. The main step involves creating an ellipse mask by identifying the starting point of colour pixels in each image. This mask is then subtracted from the original image to remove the background. Additionally, the image undergoes enhancement by adjusting its intensity range. After that, data augmentation is applied to the retinal fundus images and ground truth to increase and diversify the dataset and the transformations are translation, rotation, and flipping. The U-Net segmentation model demonstrated the effectiveness in segmenting the haemorrhages and cotton wool spots

in ultra-wide fundus images and the CNN classification model is effective in differentiating the images into RVO and non-RVO.

In conclusion, this study introduced an approach for detecting retinal vein occlusion (RVO) signs in ultra-widefield (UWF) images. It encompassed preprocessing steps tailored to UWF images, assessed the U-Net model's performance in segmenting RVO signs within UWF images, and proposed a classification method for distinguishing images as either RVO or non-RVO. The developed algorithm can be implemented as an automated screening system for the RVO diagnostic process by inputting retinal images into the system to obtain a classification label, along with the precise location of the suspected lesions if present.

5.2 Research Contributions

This research has established the baseline of the effectiveness of using the U-Net segmentation model to segment haemorrhage and cotton wool spots in UWF retinal fundus images. This baseline performance serves as a valuable reference point for assessing and comparing the performance of future research related to the segmentation of haemorrhage and cotton wool spots in UWF retinal fundus images. Furthermore, it can be used as a benchmark to evaluate and compare alternative methods for similar tasks. This research expands the knowledge on retinal image segmentation and provides a foundation for developing more precise segmentation.

This research has addressed the scarcity of publicly available annotated UWF retinal fundus images for research purposes which is a big challenge for researchers in this field. By leveraging publicly available annotated regular retinal fundus images, this research presents an approach to develop a segmentation model capable of segmenting UWF images. This research built a segmentation model by training with solely regular fundus dataset and

pre-trained weights as initial weights when training with UWF images. To assess the effectiveness of these models, a segmentation model trained solely using UWF images was compared to these models. The results show that the segmentation model trained solely using UWF images is the most effective, however, the transfer learning approach showed significant improvement compared to only using regular datasets. It suggests that the transfer learning approach involving building a custom pre-trained model is promising for future research dealing with limited data. Leveraging pre-training with a larger dataset and utilising the pre-trained weights as a starting point, reduces the training time and improves generalisation, primarily because it has been trained for a similar task.

This research has contributed to the accurate detection of RVO using the proposed algorithm for UWF retinal fundus images. The U-Net segmentation and subsequent classification of abnormalities provides a solution that not only identifies the presence of RVO but also localises the signs of RVO in the retinal fundus images. For healthcare providers, this research offers a tool for detection of RVO that enhances the interpretability and reliability of automated diagnosis processes, thereby assisting them in more informed decisions.

In short, this research contributes to the scientific research community by establishing a baseline in retinal image segmentation and expands the knowledge of retinal disease research. It contributes to the medical industry and healthcare providers by delivering an efficient tool to enhance diagnostic interpretability.

5.3 Limitations and Suggestions

This study has observed that segmenting the haemorrhages and cotton wool spots apart from other anatomical structures is complex and the process begins with image pre-

processing. The pre-processing steps should remove all irrelevant features except for the target signs as much as possible to avoid misidentification of the target signs. The pre-processing steps in this study could not remove all the irrelevant features resulting in some misidentification when other features have similar colour, size, and shape characteristics.

In this study, it was observed that the model could identify dot and blot haemorrhages well but not the other types of haemorrhages such as pre-retinal and flame-shaped haemorrhages because the training images used for segmentation of haemorrhages consist of more dot and blot haemorrhages. Hence, the model can be improved by using a dataset containing all the variations of haemorrhages so it can segment all the haemorrhages equally well.

In this study, a relatively small, annotated dataset was employed. However, the UWF dataset used in this research reflects real-world data from the local population and captures varying image quality. It is worth highlighting that the model's performance could potentially benefit from a larger annotated dataset sourced from multiple origins, as this would enable a more comprehensive representation of the diverse characteristics present in UWF images.

Additionally, due to computational constraints, the images underwent resizing to a smaller pixel resolution. It is worth considering that utilizing a larger image size or preserving the original dimensions could have provided the model with more information to work with, potentially enhancing its performance.

REFERENCES

- Abdelmaksoud, E., El-Sappagh, S., Barakat, S., Abuhmed, T., & Elmogy, M. (2021). Automatic diabetic retinopathy grading system based on detecting multiple retinal lesions. *IEEE Access*, 9, 15939–15960.
<https://doi.org/10.1109/ACCESS.2021.3052870>
- Ali, D. (2019). Diabetic exudate detection in color retinal images. *Internation Journal of Computers & Technology*, 19, 7510–7518. <https://doi.org/10.24297/ijct.v19i0.8377>
- Butt, M. M., Latif, G., Iskandar, D. N. F. A., Alghazo, J., & Khan, A. H. (2019). Multi-channel convolutions neural network based diabetic retinopathy detection from fundus images. *Procedia Computer Science*, 163, 283–291.
<https://doi.org/10.1016/j.procs.2019.12.110>
- Chen, Q., Lin, S., Liu, B.-S., Wang, Y., Wei, Q.-J., He, X.-X., Ding, F., Yang, G., Chen, Y.-X., Li, X.-R., & Hu, B.-J. (2021). Artificial intelligence can assist with diagnosing retinal vein occlusion. *International Journal of Ophthalmology*, 14(12), 1895–1902.
<https://doi.org/10.18240/ijo.2021.12.13>
- Choudhury, B., Then, P. H. H., & Raman, V. (2017). Automated detection of central retinal vein occlusion using convolutional neural network. In S. Suh & T. Anthony (Eds.), *Big Data and Visual Analytics* (pp. 1–21). Springer, Cham.
https://doi.org/10.1007/978-3-319-63917-8_1

ClinicalKey. (2021, February 2). *Retinal vein occlusion*. Clinical Overview.

https://www.clinicalkey.com/#!/content/clinical_overview/67-s2.0-9a44acc7-e161-4c9d-b84b-9484d39f7df5

Encyclopædia Britannica. (2020). *Medicine*.

<http://academic.eb.com./levels/collegiate/article/medicine/110314>

Flaxel, C. J., Adelman, R. A., Bailey, S. T., Fawzi, A., Lim, J. I., Vemulakonda, G. A., &

Ying, G. (2020). Retinal vein occlusions preferred practice pattern. *Ophthalmology*, 127(2), 288–320. <https://doi.org/10.1016/j.optha.2019.09.029>

Glycosmedia. (n.d.). *Diabetic retinopathy - Treatment of diabetic retinopathy*. Retrieved

April 2, 2022, from <http://www.glycosmedia.com/education/diabetic-retinopathy/diabetic-retinopathy-treatment-of-diabetic-retinopathy/>

Hagos, M. T., & Kant, S. (2019). Transfer learning-based detection of diabetic retinopathy

from small dataset. *ArXiv*, *abs/1905.07203*.

Ip, M., & Hendrick, A. (2019). Retinal vein occlusion review. *Asia-Pacific Journal of*

Ophthalmology. 7(1), 40-45. <https://doi.org/10.22608/APO.2017442>

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning.

Electronic Markets, 31(3), 685–695. <https://doi.org/10.1007/s12525-021-00475-2>

Jin, K., Zhou, M., Wang, S., Lou, L., Xu, Y., Ye, J., & Qian, D. (2018). Computer-aided

diagnosis based on enhancement of degraded fundus photographs. *Acta*

Ophthalmologica, 96(3), 320–326. <https://doi.org/10.1111/aos.13573>

- Kaur, J., & Kaur, P. (2022). UNICnv: An enhanced U-Net based InceptionV3 convolutional model for DR semantic segmentation in retinal fundus images. *Concurrency and Computation: Practice and Experience*, 34(21).
<https://doi.org/10.1002/cpe.7138>
- Khojasteh, P., Aliahmad, B., & Kumar, D. K. (2018). Fundus images analysis using deep features for detection of exudates, hemorrhages and microaneurysms. *BMC Ophthalmology*, 18(1), 288. <https://doi.org/10.1186/s12886-018-0954-4>
- Kipli, K., Hoque, M. E., Lim, L. T., Mahmood, M. H., Sahari, S. K., Sapawi, R., Rajae, N., & Joseph, A. (2018). A review on the extraction of quantitative retinal microvascular image feature. *Computational and Mathematical Methods in Medicine*, 2018, 1–21.
<https://doi.org/10.1155/2018/4019538>
- Kousai, S. (2019, November 1). *Understand the architecture of CNN*. Medium.
<https://towardsdatascience.com/understand-the-architecture-of-cnn-90a25e244c7>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
<https://doi.org/10.1038/nature14539>
- Lin, Y.-N., Hsieh, T.-Y., Yang, C.-Y., Shen, V. R., Juang, T. T.-Y., & Chen, W.-H. (2020). Deep petri nets of unsupervised and supervised learning. *Measurement and Control*, 53(7–8), 1267–1277. <https://doi.org/10.1177/0020294020923375>
- Lu, W., Tong, Y., Yu, Y., Xing, Y., Chen, C., & Shen, Y. (2018). Applications of artificial intelligence in ophthalmology: General overview. *Journal of Ophthalmology*, 2018,

- 1–15. <https://doi.org/10.1155/2018/5278196>
- Luo, X., Li, J., Chen, M., Yang, X., & Li, X. (2021). Ophthalmic disease detection via deep learning with a novel mixture loss function. *IEEE Journal of Biomedical and Health Informatics*, 25(9), 3332–3339. <https://doi.org/10.1109/JBHI.2021.3083605>
- Mehrotra, K. G., Mohan, C. K., & Huang, H. (2017). *Anomaly detection principles and algorithms*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-67526-8>
- Mehta, S. (2020). *Central retinal vein occlusion and branch retinal vein occlusion*. Merck Manual. <https://www.merckmanuals.com/professional/eye-disorders/retinal-disorders/central-retinal-vein-occlusion-and-branch-retinal-vein-occlusion>
- Minar, M. R., & Naher, J. (2018). *Recent advances in deep learning: An overview*. <https://doi.org/10.13140/RG.2.2.24831.10403>
- Nagasato, D., Tabuchi, H., Ohsugi, H., Masumoto, H., Enno, H., Ishitobi, N., Sonobe, T., Kameoka, M., Niki, M., & Mitamura, Y. (2019). Deep-learning classifier with ultrawide-field fundus ophthalmoscopy for detecting branch retinal vein occlusion. *International Journal of Ophthalmology*, 12(1), 94–99. <https://doi.org/10.18240/ijo.2019.01.15>
- Nicholson, L., Talks, S. J., Amoaku, W., Talks, K., & Sivaprasad, S. (2022a). Retinal vein occlusion (RVO) guideline: Executive summary. *Eye*, 36(5), 909–912. <https://doi.org/10.1038/s41433-022-02007-4>

- Nicholson, L., Talks, S. J., Amoaku, W., Talks, K., & Sivaprasad, S. (2022b). Retinal vein occlusion (RVO) guideline: Executive summary. *Eye*, 36(5), 909–912.
<https://doi.org/10.1038/s41433-022-02007-4>
- Pachade, S., Porwal, P., Thulkar, D., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., Giancardo, L., Quellec, G., & Mériaudeau, F. (2021). Retinal Fundus Multi-Disease Image Dataset (RFMiD): A dataset for multi-disease detection research. *Data*, 6(2), 14. <https://doi.org/10.3390/data6020014>
- Partovi, M., Rasta, S. H., & Javadzadeh, A. (2016). Automatic detection of retinal exudates in fundus images of diabetic retinopathy patients. *Journal of Analytical Research in Clinical Medicine*, 4(2), 104–109. <https://doi.org/10.15171/jarcm.2016.017>
- Patel, S. N., Shi, A., Wibbelsman, T. D., & Klufas, M. A. (2020). Ultra-widefield retinal imaging: An update on recent advances. *Therapeutic Advances in Ophthalmology*, 12, 251584141989949. <https://doi.org/10.1177/2515841419899495>
- Piyush Mishra. (2020). *Convolutional neural networks (CNN)*. OpenGenus IQ.
<https://iq.opengenus.org/convolutional-neural-networks/>
- Ponto, K. A., Elbaz, H., Peto, T., Laubert-Reh, D., Binder, H., Wild, P. S., Lackner, K., Pfeiffer, N., & Mirshahi, A. (2015). Prevalence and risk factors of retinal vein occlusion: The Gutenberg Health Study. *Journal of Thrombosis and Haemostasis*, 13(7), 1254–1263. <https://doi.org/10.1111/jth.12982>
- Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., &

Meriaudeau, F. (2018). Indian Diabetic Retinopathy Image Dataset (IDRiD): A database for diabetic retinopathy screening research. *Data*, 3(3), 25.

<https://doi.org/10.3390/data3030025>

Rasche, C. (2019). *Computer vision*.

https://www.researchgate.net/publication/336460083_Computer_Vision

Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, 29(9), 2352–2449.

https://doi.org/10.1162/neco_a_00990

Retina Gallery. (2014). *Central retinal vein occlusion*.

<https://retinagallery.com/displayimage.php?pid=7453>

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In: N. Navab, J. Hornegger, W. Wells, A. Frangi, (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. MICCAI 2015 (pp. 234-241) Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

Sahlsten, J., Jaskari, J., Kivinen, J., Turunen, L., Jaanio, E., Hietala, K., & Kaski, K. (2019). Deep learning fundus image analysis for diabetic retinopathy and macular edema grading. *Scientific Reports*, 9(1), 10750. <https://doi.org/10.1038/s41598-019-47181-w>

w

Salau, A., & Jain, S. (2019). Feature extraction: A survey of the types, techniques,

applications. *2019 International Conference on Signal Processing and Communication (ICSC)*, 158–164.

<https://doi.org/10.1109/ICSC45622.2019.8938371>

Sarker, I. H. (2021). Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2(6), 420.

<https://doi.org/10.1007/s42979-021-00815-1>

Sharif, N. A. M., Azhar, A. S. M., Harun, N. H., Bakar, J. A., Abdullah, A. A., & Chong, Y.

F. (2021). Green channel and top hat-based image enhancement for diabetic retinopathy screening. *Journal of Physics: Conference Series*, 1997(1), 012002.

<https://doi.org/10.1088/1742-6596/1997/1/012002>

Shetty, D. K., Talasila, A., Shanbhag, S., Patil, V., Hameed, B. M. Z., Naik, N., & Raju, A.

(2021). Current state of artificial intelligence applications in ophthalmology and their potential to influence clinical practice. *Cogent Engineering*, 8(1), 1920707.

<https://doi.org/10.1080/23311916.2021.1920707>

Sivaprasad, S., Amoaku, W. M., & Hykin, P. (2015). The Royal College of

Ophthalmologists guidelines on retinal vein occlusions: Executive summary. *Eye*, 29(12), 1633–1638. <https://doi.org/10.1038/eye.2015.164>

Son, J., Shin, J. Y., Chun, E. J., Jung, K.-H., Park, K. H., & Park, S. J. (2020). Predicting high coronary artery calcium score from retinal fundus images with deep learning algorithms. *Translational Vision Science & Technology*, 9(2), 28.

<https://doi.org/10.1167/tvst.9.2.28>

Srivignesh Rajan. (2020, July 15). *An introduction to artificial neural networks*. Medium.

<https://towardsdatascience.com/an-introduction-to-artificial-neural-networks-5d2e108ff2c3>

Sumit Saha. (2018, December 16). *A comprehensive guide to convolutional*

neural networks-the ELI5 way. Medium. <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>

Wan, C., Chen, Y., Li, H., Zheng, B., Chen, N., Yang, W., Wang, C., & Li, Y. (2021). EAD-

Net: A novel lesion segmentation method in diabetic retinopathy using neural networks. *Disease Markers*, 2021, 1–13. <https://doi.org/10.1155/2021/6482665>

Wang, W., & Yang, Y. (2019). Development of convolutional neural network and its

application in image classification: A survey. *Optical Engineering*, 58(04), 1. <https://doi.org/10.1117/1.OE.58.4.040901>

Wehle, H. (2017). Machine learning, deep learning, and AI: What's the difference?

Conference: Data Scientist Innovation Day.

Xu, K., Feng, D., & Mi, H. (2017). Deep convolutional neural network-based early

automated detection of diabetic retinopathy using fundus image. *Molecules (Basel, Switzerland)*, 22(12). <https://doi.org/10.3390/molecules22122054>

Zhang, W., Zhao, X., Chen, Y., Zhong, J., & Yi, Z. (2021). DeepUWF: An automated Ultra-

wide-field fundus screening system via deep learning. *IEEE Journal of Biomedical and Health Informatics*, 25(8), 2988–2996.

<https://doi.org/10.1109/JBHI.2020.3046771>

APPENDICES

Appendix A: Journal Publications

1. **Ong, I. S. Y.**, Lim, L. T., Mahmood, M. H., & Lee, N. K. (2023). U-Net segmentation of ultra-widefield retinal fundus images for retinal vein occlusion associated lesion recognition. *2023 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAJET)*, 106–111.
<https://doi.org/10.1109/IICAJET59451.2023.10291722>

Appendix B: Web app demonstration of segmentation and classification of RVO

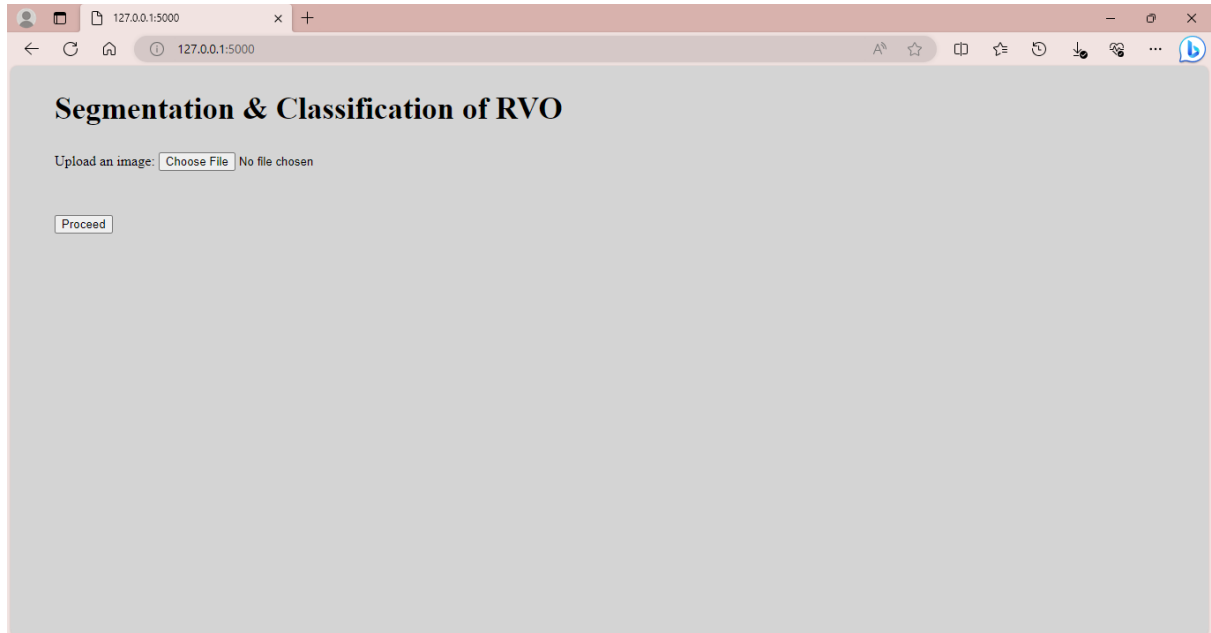


Figure 1: Sample of the web page that allows users to upload a UWF fundus image

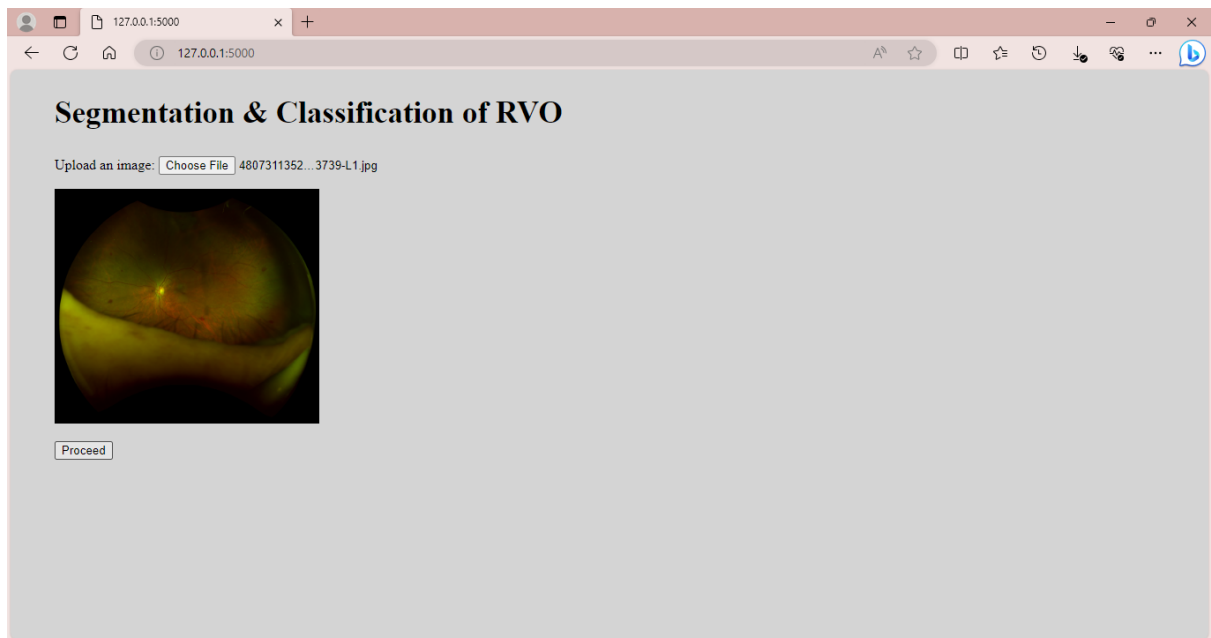


Figure 2: The user has chosen a UWF fundus image for screening

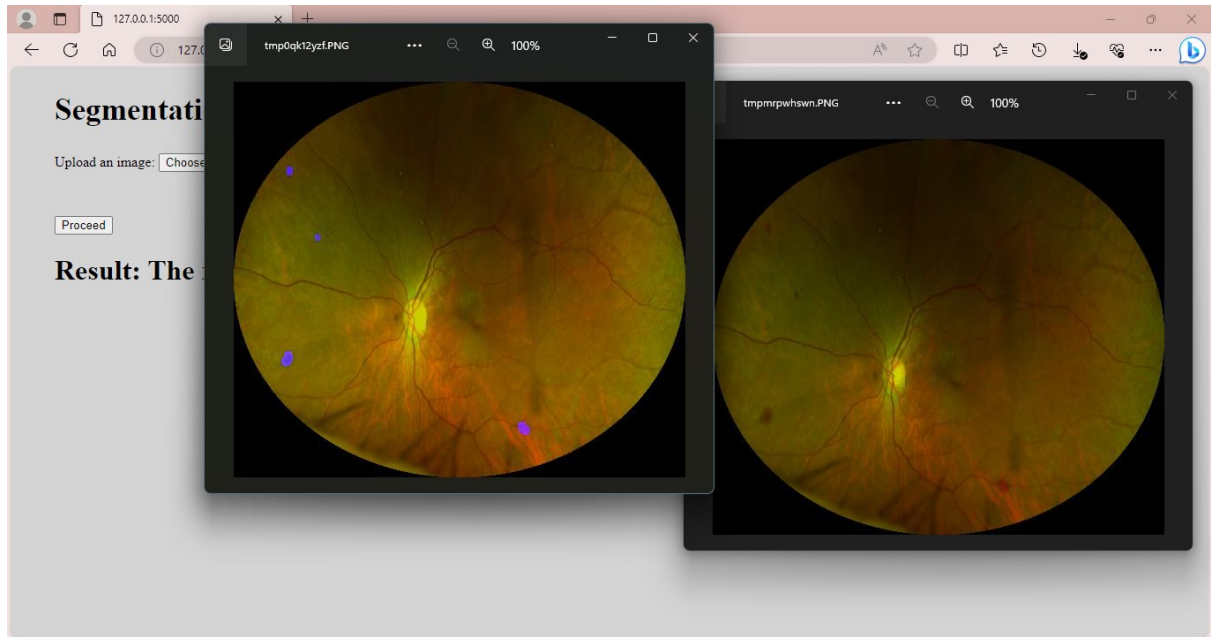


Figure 3: The predicted signs of RVO are shown on the web page

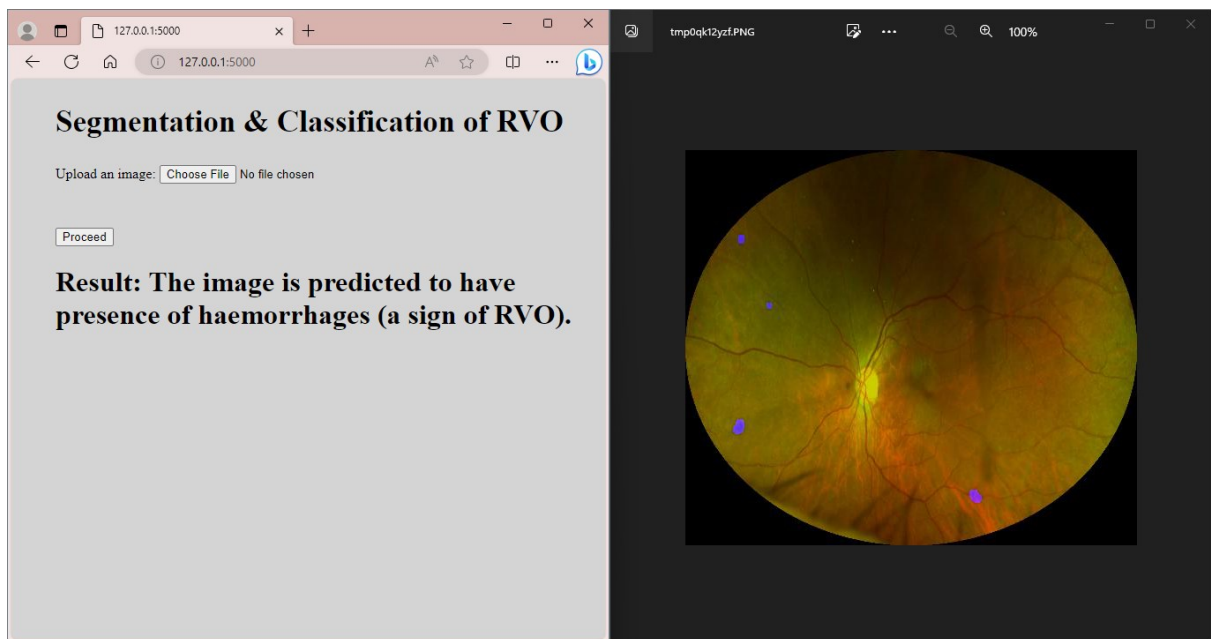


Figure 4: The predicted classification of RVO is shown on the web page