Improving Diabetic Retinopathy Classification using Transfer Learning and Optimized Deep Learning Models

BY

Dai, Cheng Xiao

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF COMPUTER SCIENCE (HONOURS)

Faculty of Information and Communication Technology

(Kampar Campus)

JUNE 2024

UNIVERSITI TUNKU ABDUL RAHMAN

REPORT STATUS DECLARATION FORM

Title: **Improving Diabetic Retinopathy Classification using Transfer Learning and Optimized Deep Learning Models**

Academic Session: JUNE 2024

I DAI, CHENG XIAO

(CAPITAL LETTER)

declare that I allow this Final Year Project Report to be kept in Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.

Verified by,

Dai Cheng Xiao

(Author's signature)

Sayed Ahmad Zikri

(Supervisor's signature)

Address:
No. 1159, Jalan Seksyen 1/2,
Taman Bandar Barat,
31900, Kampar, Perak

Sayed Ahmad Zikri Bin Sayed Aluwee
Supervisor's name

Date: 09/13/2024

Date: 09/13/2024

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

UNIVERSITI TUNKU ABDUL RAHMAN

Date: 09/13/2024

SUBMISSION OF FINAL YEAR PROJECT /DISSERTATION/THESIS

It is hereby certified that ***Dai, Cheng Xiao*** (ID No: ***2100469***) has completed this final year project/ dissertation/ thesis* entitled "*Improving Diabetic Retinopathy Classification using Transfer Learning and Optimized Deep Learning Models*" under the supervision of *Dr. Sayed Ahmad Zikri Bin Sayed Aluwee* (Supervisor) from the Department of *Computer Science*, Faculty/Institute* of *Information and Communication Technology*, and *Dr. Ng Hui Fuang* (Co-Supervisor)* from the Department of *Computer Science*, Faculty/Institute* of *Information and Communication Technology*.

I understand that University will upload softcopy of my final year project / dissertation/ thesis* in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

Dai Cheng Xiao

(*Dai, Cheng Xiao*)

*Delete whichever not applicable

DECLARATION OF ORIGINALITY

I declare that this report entitled "**Improving Diabetic Retinopathy Classification using Transfer Learning and Optimized Deep Learning Models**" is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature : Dai Cheng Xiao

Name : Dai, Cheng Xiao

Date : 09/13/2024

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my supervisor Dr. Sayed Ahmad Zikri Bin Sayed Aluwee and my moderator Dr. Ng Hui Fuang . I am thankful to them for their guidance and help and for giving me the opportunity to work in the field of Medical Artificial Intelligence, especially in Diabetic Retinopathy Image Classification. This is a great opportunity for me to explore the field of Artificial Intelligence and Deep Learning, which will help me in my future career development.

I would also like to express my sincere thanks to my parents and my best friends Lee Jian Shen, Paun Tek You, Foo Shao Lun, and Sheik Abdul Rahuman. Appreciate the fact that they always go out of their way to help me, both in study and in life. All of my achievements would not have been possible without their support and encouragement.

ABSTRACT

Diabetic retinopathy (DR) is an eye disease closely related to diabetes that may lead to severe vision loss or even blindness if not diagnosed and treated in time. With the increasing number of diabetic patients, DR has gradually become a global public health problem. Against this background, it is essential to develop a method that can accurately and efficiently diagnose DR. With the continuous advancement of Artificial Intelligence (AI) technology, deep learning and transfer learning have become powerful tools for modern medical research, especially in medical image analysis. This research project aims to combine transfer learning and deep learning techniques to solve this challenge. The main focus is to address the problem of insufficient labelled image datasets in the medical field through transfer learning, exploring how to use limited labelled data to train high-performance models effectively. It will also delve into the use of optimized deep learning models to improve the classification accuracy of DR. It is expected to provide a more accurate and efficient tool for the early diagnosis and treatment of DR, thus helping to decrease vision loss and related complications caused by DR.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

DR	Diabetic Retinopathy
AI	Artificial Intelligence
IDF	International Diabetes Federation
DM	Diabetes Mellitus
APTOS	ASIA PACIFIC TELE-OPHTHALMOLOGY SOCIETY
DenseNet	Dense Connection Network
CNN	Convolutional Neural Network
QWK	Quadratic Kappa
TTA	Test-Time Augmentation
CLAHE	Contrast Limited Adaptive Histogram Equalization
GAN	Generative Adversarial Network
MIL-VT	Multiple Instance Learning Enhanced Vision Transformer
ViT	Vision Transformer
DDPM	Denoising Diffusion Probabilistic Model
LayerNorm	Layer Normalisation
MLP	Multilayer Perceptron Machine
LMF Loss	Large Margin aware Focal Loss
LDAM Loss	Label-Distribution-aware Margin Loss
TP	True Positives
TN	True Negatives
FP	False Positives
FN	False Negatives

Chapter 1

Project Background

1.1 Introduction

Over the last decade, the prevalence of diabetes mellitus is rising in Malaysia and globally due to changes in modern lifestyles and an ageing population. According to the International Diabetes Federation (IDF), the number of people with diabetes mellitus (DM) worldwide is projected to be around 463 million in 2019 and is predicted to increase to 700 million by 2045. In 2020, it was estimated that around 103.12 million humans would be impacted by DR globally, a figure that is anticipated to escalate to approximately 160.50 million by the year 2045 [1]. This disease places a huge health burden on patients and a heavy burden on the global public health system. Damage to the retinal blood vessels brought on by high blood sugar levels causes DR [2]. The clinical manifestations of DR can progress from initial microvascular abnormalities to more severe retinal haemorrhages and leakage. Without prompt and appropriate treatment, DR could lead to severe vision loss possibly even blindness. Further deterioration of DR lesions may be successfully controlled by timely diagnosis and treatment, thus protecting patients' vision and quality of life. Therefore, developing more efficient and accurate diagnostic methods for DR is an essential direction for current research.

In the contemporary era, AI has emerged as a forefront domain of technological progress. It represents a formidable learning technology capable of handling vast datasets through neural networks that emulate the human brain. Deep Learning, a subset of AI, has demonstrated its remarkable prowess in analysis and prediction across various domains. One notable application is in medical image analysis, where the goal is to enhance diagnostic precision and streamline the diagnostic process. Transfer learning is another branch of the AI field. The core strength of transfer learning is the capability to solving new and relevant problems using pre-trained models, thereby significantly decreasing data requirements and computational resources. Especially in medical image analysis, better models can be trained through some small amount of labelled data, which was essential to alleviating the problem of scarce labelled data.

1.2 Problem Statement

• DR manual diagnosis is less accurate.

Typically, the diagnosis of this deadly disease relies on highly trained specialists to examine colour fundus images. However, this manual diagnostic approach by clinicians could be more convenient and error-prone [3]. At first, it requires extensive experience and expertise to accurately identify and resolve small changes and abnormalities in the fundus images. Secondly, this method also relies heavily on the judgement and interpretation of individual doctors, which can lead to some variation and inconsistency between diagnostic results. More importantly, manual diagnosis is usually time-consuming, as each fundus image requires meticulous examination. In addition, long working hours may increase the risk of errors due to human fatigue and concentration limitations. As a result, this method may lead to misdiagnosis or omission, which may affect the patient's treatment and recovery.

• Lack of high-quality labelled medical image datasets.

Accurately training a deep learning model for DR classification necessitates a substantial quantity of images, presenting a significant constraint within the field of diabetic retinopathy [4]. Initially, the assembly and annotation of such a dataset demand considerable time and resources. Every image necessitates meticulous labelling by a seasoned medical specialist, it is a time-intensive and expensive process. Furthermore, developing a well-rounded and varied dataset presents a formidable challenge due to the complex and intricate cases. This dataset should encompass a broad spectrum of cases, from the mild to the extreme, to guarantee the model's proficiency in discerning and categorising different forms of DR. Consequently, the scarcity of premium quality datasets might culminate in subpar model efficacy, hindering the precise detection and categorisation of DR.

• Deep learning techniques still have optimisation space.

Deep learning technology is in its developmental stage and still has a lot of space for improvement. While new deep learning architectures are emerging, existing deep learning models still rely on large amounts of labelled data for training and require careful hyperparameters tuning to improve model performance.

1.3 Motivation

The motivation of this research study is to improve the existing deep learning models applied transfer learning techniques and some model optimization techniques so that the improved model can more accurately identify and classify the different stages of DR and estimate its severity, as well as overcome the problem of insufficient existing DR datasets and reduce the time cost of diagnosis for doctors. In addition, the use of the latest transfer learning and deep learning prediction techniques to assist doctors in detecting and classifying DR is not yet widespread in Malaysia compared to other developed countries such as the United Kingdom and the United States that are using the latest transfer learning and deep learning prediction techniques, which is also the motivation for this project. This research project aims to combine deep learning and transfer learning techniques to optimize existing prediction models and improve their accuracy.

1.4 Research Objectives

- **Research and implement transfer learning techniques to improve DR classification performance by using pre-trained models.**

Many related works have used deep learning models to train DR classification models from scratch, which is the most common approach, but this approach tends to rely on large amounts of labelled data. Therefore, this project plans to adopt an innovative approach to decrease the reliance on labelled data by applying transfer learning techniques and using pre-trained models. This not only decreases the need for large amounts of labelled data but also helps to enhance the performance of the model.

- **Explore and implement various optimisation techniques to further enhance the performance of DR classification models.**

This project will explore and implement various deep learning model optimisation techniques to find suitable model architectures and model optimization methods for DR classification to further boost the performance of models.

- **Combining transfer learning and deep learning techniques for DR classification related task.**

The project will introduce transfer learning training methods to improve the performance of the existing model. So that the results classified by pertinent deep learning models can be the most reliable second opinion for ophthalmologists, assisting them to improve the efficiency of DR diagnosis and decrease the misdiagnosis rate. It will also help to detect DR as early as possible for more patients or potential patients, so that they can be treated as early as possible to prevent their disease from deteriorating.

1.5 Project Scope and Direction

The purpose of this project is to research how to enhance the accuracy for existing DR classification model and decrease the model's dependence on large amounts of labelled data by using transfer learning techniques and optimizing deep learning models. In this way, its classification results can be used as a second opinion for doctors, thus improving diagnostic accuracy and reducing diagnostic time.

i. The transfer learning techniques will be able to apply to DR detection and classification problems.
ii. The deep learning techniques will be able to apply to DR detection and classification.
iii. The model optimization techniques will be able to optimize DR classification models

1.6 Contributions

The value of this research project lies in its ability to utilize the transfer learning technique and optimize the existing deep learning prediction models, which enables the prediction models to be trained more efficiently with limited annotated data, thus improving the performance and accuracy of the models. Not only can it significantly save doctors' time, but it can also maintain high accuracy of the results, thus improving the efficiency of diagnosis.

1.7 Report Organization

In Chapter 1 of the report, the background, objectives, scope, and motivation of the project were described in detail. In the Chapter 2 section of this project report, reviewed some of the latest articles related to the classification or detection of DR lesions using deep learning or migration learning methods. In Chapter 3 and Chapter 4, the methodological design of this study is listed, as well as the project timeline and the hardware and software used. In the Chapter 5 the results of the experiments were compared and analysed, some examples of classification samples and visualisation examples were also output. Finally, Chapter 6 is conclusion and recommendation part.

Chapter 2

Literature Review

2.1 Related works on the classification of DR and fundus image generation

2.1.1 Preliminary Study of Diabetic Retinopathy Classification from Fundus Images Using Deep Learning Model

This paper [5] proposed a classification method using deep learning approach for five classification of DR fundus images task. This paper used Dense Connection Network (DenseNet) model to train DR fundus images data from Kaggle (ASIA PACIFIC TELE-OPHTHALMOLOGY SOCIETY) APTOS 2019 Blindness Detection competition. DenseNet is a variant deep learning architecture of a convolutional neural network (CNN), which is characterised by the fact that each layer in this network is connected to all previous layers.

Besides, this paper also explored different data preprocessing and data augmentation methods, in addition to the commonly used image dark-area cropping and image resizing, this paper tried seven different image augmentation methods, and tested the Quadratic Kappa Value of the dataset after preprocessing of these methods, and finally the combination as shown in the Figure 2.1.1.1 below accomplished the best quadratic kappa (QWK) value of 0.9308.

Proposed Data Enhancement Methods	Highest QWK Value
rotation_range = 360, horizontal_flip = True, vertical_flip = True, width_shift_range = 0.2, height_shift_range = 0.2	0.9308

Figure 2.1.1.1 Proposed data enhancement methods

Since this paper is tested on a test set that is different from the source of the training set, on top of the test-time augmentation (TTA) method employed to optimized model performance on test datasets. Finally, this model was test on the Messidor-2 dataset with 65% accuracy, which successfully tested the model's generalisation capability. The results of the classification test are reported as shown in Figure 2.1.1.1 below.

	precision	recall	f1-score	support
0	0.72	0.91	0.80	1017
1	0.19	0.09	0.12	270
2	0.55	0.45	0.49	347
3	0.44	0.20	0.28	75
4	0.37	0.20	0.26	35
accuracy			0.65	1744
macro avg	0.45	0.37	0.39	1744
weighted avg	0.58	0.65	0.60	1744

Figure 2.1.1.2 Test results of presented model on messidor-2 dataset

From the results, it was found that the paper still had only moderate accuracy, and in this paper, as shown in figure 2.1.1.2 below, it was also displayed that the model still had overfitting problems, and there was still more room for improvement. To further enhanced the classification accuracy of the proposed model, more deep learning architectures should be tried, and transfer learning methods should be applied to overcome the lack of dataset.

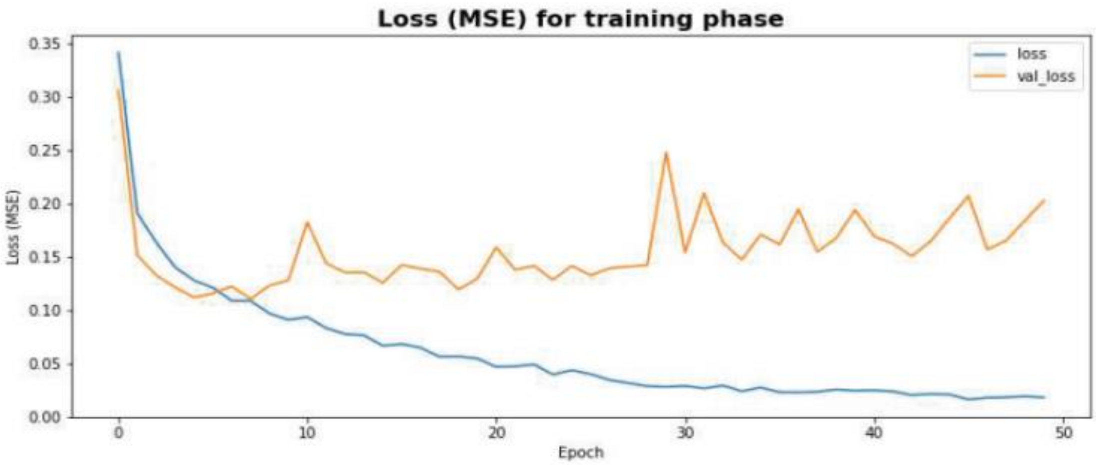


Figure 2.1.1.3 The loss of presented model during training phase

2.1.2 EDLDR: An Ensemble Deep Learning Technique for Detection and Classification of Diabetic Retinopathy

In this journal paper [6], an integrated deep learning technique is proposed under the name EDLDR, to solve the problem of detecting and classifying DR. This article integrates two optimised deep learning models as a way to boost the accuracy of deep learning methods in DR classification filed. These two models were the improved DenseNet101 model and the ResNeXt model, respectively. The training flow of the integrated modelling approach presented by this article is shown in Figure 2.1.2.1 below.

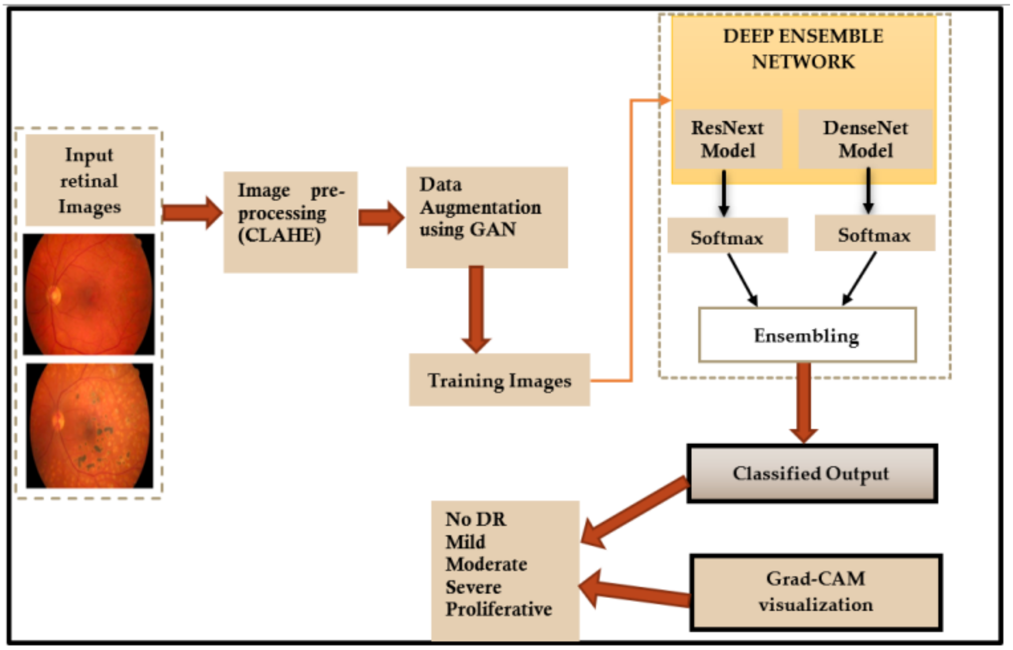


Figure 2.1.2.1 Flowchart of the method proposed in this article

The input fundus image data were pre-processed by Contrast Limited Adaptive Histogram Equalization (CLAHE) method at first, subsequently as the authors found that the high categories in the dataset are very unbalanced and the proliferative types of images were obviously less than the other categories of images, so in this paper the Generative Adversarial Network (GAN) based data enhancement technique is used for data enhancement as a way of supplementing the dataset to make up for the imbalance in the categories. Afterwards, the pre-processed data is put into the integrated model introduced by this journal paper for training, within the integrated model presented in this journal paper, the ResNeXt model is used as a part of the integrated model, which is improved on top of the existing ResNet model by using the split-transform-merge strategy.

Finally, the authors most compared the results of this paper with other existing methods as shown in Figure 2.1.2.2 and Figure 2.1.2.3. In this article, the binary classification performance of the model is tested on the binary classification dataset DIARETDB1 is and the final binary classification accuracy was 96.98%. Meanwhile, it is tested on the five classification dataset APTOS2019 and finally the DR five classification accuracy of the integrated model presented from this journal paper was 86.08% .

Method	Precision	Recall	Accuracy
DRISTI (VGG16 + Capsule) [25]	0.96	0.96	96.24
EfficientNet-B3 [26]	0.95	0.96	96.07
Resnet50 + Capsule [25]	0.94	0.93	95.54
EDLDR (Proposed Method)	0.97	0.97	96.98

Figure 2.1.2.2 Binary classification performance of integrated models

Method	Precision	Recall	Accuracy
DRISTI (VGG16 + Capsule) [25]	0.91	0.88	82.06
EfficientNet-B3 [26]	0.59	0.66	84.86
Resnet50 + Capsule [25]	0.59	0.69	76.80
EDLDR (Proposed Method)	0.76	0.82	86.08

Figure 2.1.2.3 Five classification performance of integrated models

To conclude, the integrated model that was proposed in this article integrated multiple deep learning models to improve performance, and the used of GAN-based data augmentation techniques overcame the problem of insufficient data. These were innovative approaches that were worthwhile references for work in the field of DR recognition and classification.

2.1.3 MIL-VT: Multiple Instance Learning Enhanced Vision Transformer for Fundus Image Classification

This article [7] presented a model named Multiple Instance Learning Enhanced Vision Transformer (MIL-VT) and applied it to the DR image classification missions. The methodology outlined in this paper combines Multiple Instance Learning (MIL) and Vision Transformer (ViT) techniques. The proposed model's architecture in this article is shown below as Figure 2.1.3.1.

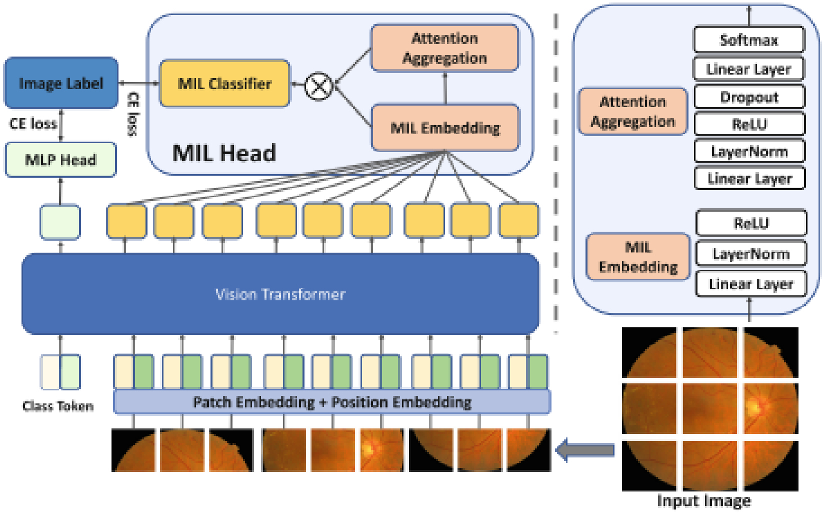


Figure 2.1.3.1 MIL-VT model architecture

Vit differs from traditional deep learning models using a CNN architecture in that his implementation relies heavily on self-attention mechanisms, this is because it is critical if that model is to capture long-distance dependencies in the image. Also, with the MIL module added, the model is able to more adequately capture the features of the various parts of the fundus image., and the improved Vit model based on this foundation greatly improves the efficiency of feature extraction as well as exploitation of relatively complex medical images. In this article, the training strategy for the model was to pre-trained with a huge dataset of fundus images and then applied it to the DR classification task by fine-tuning it on this relevant dataset, which is a traditional transfer learning strategy that overcomes the lack of medical datasets very well.

The authors then performed model ablation tests on two relevant datasets, which are APTOS2019 and RFMiD2020. and compared the results with the current the SOTA results for each dataset as shown in Figures 2.1.3.2 and 2.1.3.3 below, it can be found that the model presented in this article in ablation and comparison tests achieved the best performance, with Accuracy and Kappa metrics of 97.9% and 92.0%, respectively, tested on APTOS2019 dataset.

Model	Combination		DR Grading			
	Pre-train	MIL	AUC	Acc	F1	Kappa
VT (ImageNet)			96.7	82.3	81.5	89.0
VT (Fundus)	✓		97.5	84.6	83.8	91.1
MIL-VT (Fundus)	✓	✓	**97.9**	**85.5**	**85.3**	**92.0**

Figure 2.1.3.2 Ablation test performance of MIL-VT (APTOS2019)

Method	AUC	Acc	F1	Kappa
ResNet34 (ImageNet)	96.5	82.9	82.4	88.8
ResNet34 (Fundus)	97.0	85.0	84.7	90.2
DLI [13]	–	82.5	80.3	89.5
CANet [10]	–	83.2	81.3	90.0
GREEN-ResNet50 [11]	–	84.4	83.6	90.8
GREEN-SE-ResNext50 [11]	–	**85.7**	85.2	91.2
MIL-VT (Proposed)	**97.9**	85.5	**85.3**	**92.0**

Figure 2.1.3.3 Comparative experimental performance performance of MIL-VT (APTOS2019)

In conclusion, the innovation of this article was to explore the application of ViT in DR classification tasks with an additional MIL head module, which improves the efficiency of feature extraction.

2.1.4 Diabetic Retinopathy Prediction Based on Deep Learning and Deformable Registration

This article [8] proposed a method in order to improving diagnostic accuracy for DR lesion by integrating learning with a voting system. The proposed method integrated a total of four deep learning models based on CNN architecture, namely DenseNet-121, Xception, Inception-v3, and ResNet-50. This article adopted a migration learning approach by first pre-training them on the ImageNet dataset, and then subsequently in this task, each of these four deep learning models were trained and fine-tuned. Ultimately, in the testing phase of this experiment, the outputs for the four models were integrated into a voting system as a way to enhance the reliability of the models. Specific architecture for the integrated voting model introduced in this journal paper is shown in Figure 2.1.4.1 below.

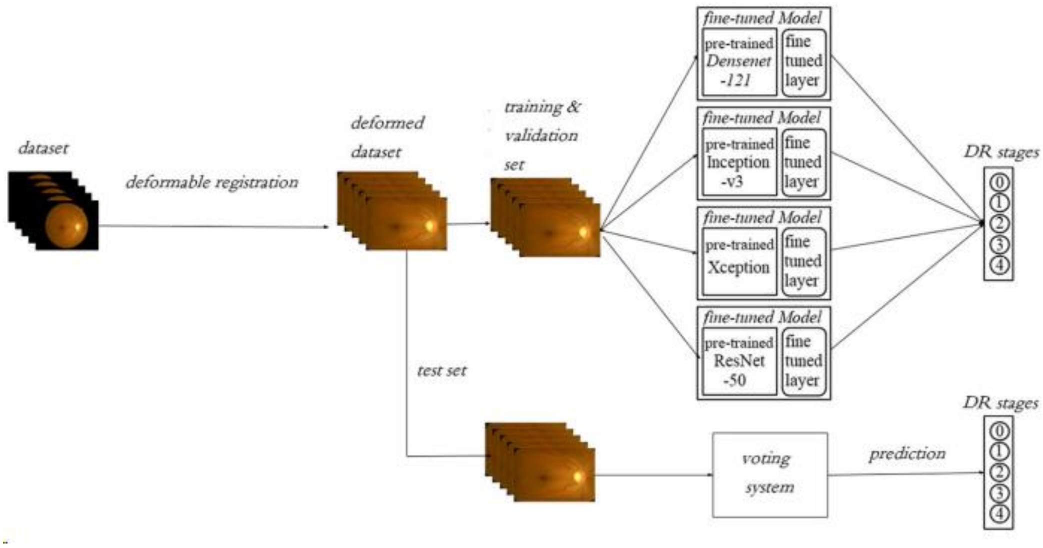


Figure 2.1.4.1 The architecture of voting model

In addition to that, this article also did some explorations in data preprocessing. The authors used the deformable alignment technique to preprocess the dataset images, this method allows the classified key feature regions to be more prominent in the overall input image, effectively decreasing irrelevant and redundant background interference.

The final experiment of this paper was fine-tuned and tested on the APTOS2019 dataset, and the confounding evidence of the test results is shown in Figure 2.1.4.2 below. The results showed that the presented method achieved an accuracy of 85.28% in classifying DR lesions.

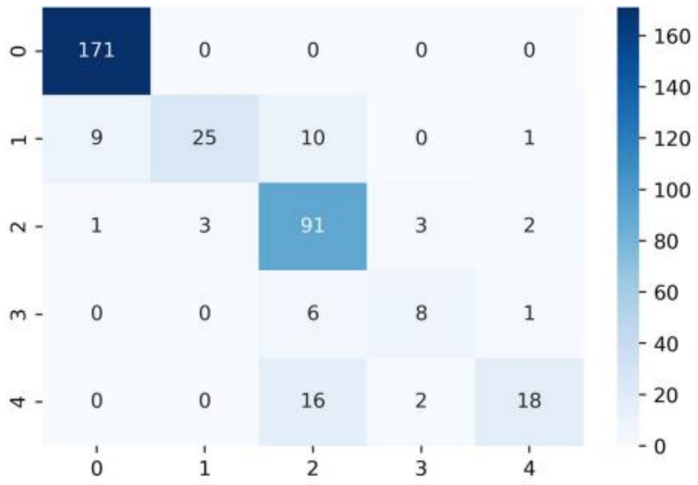


Figure 2.1.4.2 Confusion matrix for the presented voting model

In summarisation, the innovation of this article was to significantly improve the accuracy of lesion detection by integrating multiple deep learning models and an optimised image pre-processing step. For noisy images, deformable alignment may be affected, which in turn affects the DR classification performance. However, for noisy images, deformable alignment may be affected, which in turn affects the classification performance. Nevertheless, this article still has some limitations. For example, for noisy images, deformable alignment may be affected, and this situation will affect the classification performance at the same time.

2.1.5 Simple methods for the lesion detection and severity grading of diabetic retinopathy by image processing and transfer learning

This article [9] described how to overcome the problem of lack of relevant datasets using transfer learning strategy. Since the medical image datasets relevant to this article were very niche, the model training method proposed in this article adopted a migration learning approach, i.e., the authors used deep learning models pre-trained on ImageNet or other large-scale datasets, in order to efficiently utilised the limited datasets for the DR detection and classification tasks.

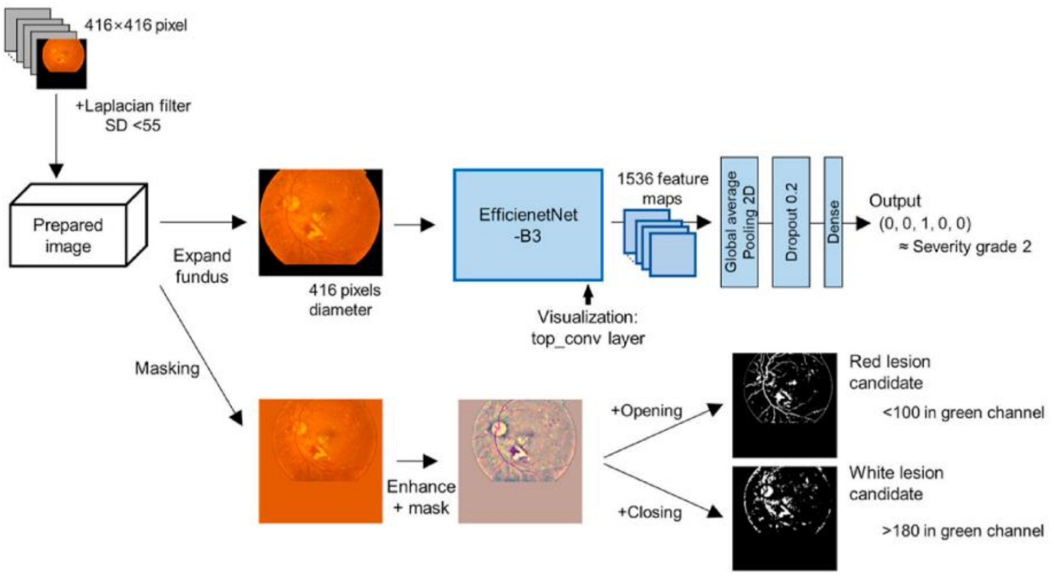


Figure 2.1.5.1 Specific training process

As shown in Figure 2.1.5.1 above, the image is pre-processed before training phase, and in this article image pre-processing methods such as resizing, using filters to enhance image details and enhance image contrast, and also used de-noising as a way of ensuring that the inputs to the subsequent model training are high quality images so as to facilitate the pre-training in absorbing useful information about the image features, and to improve the accuracy of the fine-tuning phase. In this article, the EfficientNet-B3 pre-training model is used, from, this pre-training model has a smaller number of parameters compared to the ResNet50 and VGG16 pre-training, but it scored high in pre-training phase. The final accuracy of the proposed model in this article is 84.86% in the DR lesion five classification task.

At the same time, this article has made some explorations on the interpretability aspect of the model, as shown in Figure 2.1.5.2 below, in which Grad CAM++, Faster Score CAM and Smooth Grad are used to provide visual interpretation for the proposed model. From the output, it can be seen that the model focuses on the relevant regions of the optic disc and the lesion during the classification process and also captures the blood vessels. Visualisation will better help doctors understand the decision-making process of the model.

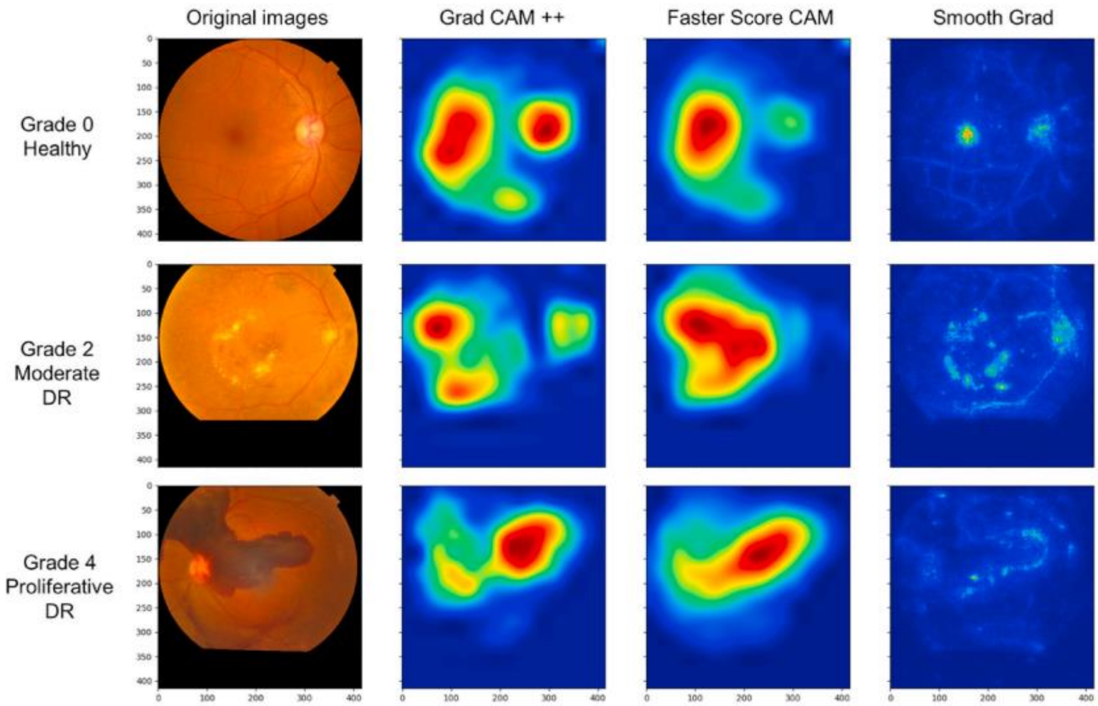


Figure 2.1.5.2 Visualisation interest areas of the model

2.1.6 A feasibility study on the adoption of a generative denoising diffusion model for the synthesis of fundus photographs using a small dataset

Paper [10] focused on the feasibility of using denoising diffusion probabilistic model (DDPM) to synthesise fundus photographs under limited dataset conditions. With the development of deep learning techniques, generative models, especially GAN and DDPM, played a key role in overcoming the imbalance of medical image datasets and the field of medical image generation. The aim of this study was to validate the feasibility of DDPM to generate high-quality fundus images on a small fundus dataset.

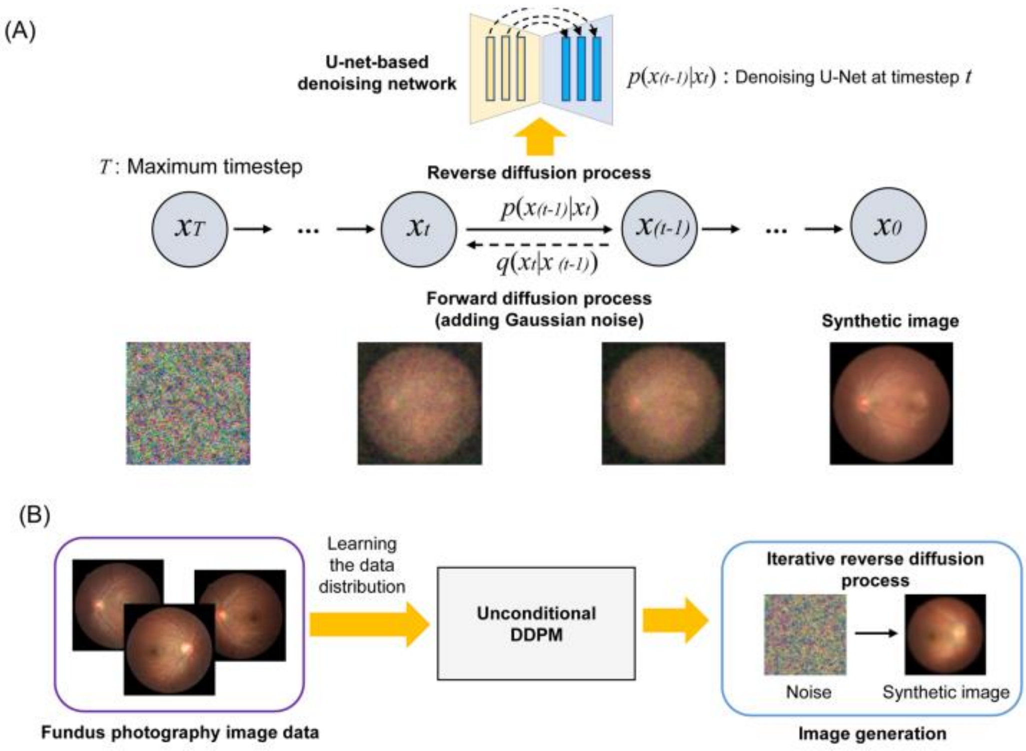


Figure 2.1.6.1 DDPM architecture and flowchart of the training process

This paper utilized 1,000 healthy retinal images to train the DDPM. The specific model structure and training process of which DDPM as shown in Figure 2.1.6.1 above. These images were sourced from a publicly available fundus photography database. In terms of model construction, the DDPM was based on the U-Net architecture, incorporating a series of diffusion steps to introduce random noise into the images and learning to reverse this process to synthesize the target images. The experimental results demonstrated that the DDPM was capable of effectively generating fundus images at a resolution of 128x128 pixels but attempts to synthesize higher resolution images

(256x256 pixels) were unsuccessful due to computational resource limitations. In comparisons of image quality, although the DDPM successfully avoided mode collapse and grid artifacts, showing its potential in fundus image synthesis, its performance was still limited due to the small size of the dataset used.

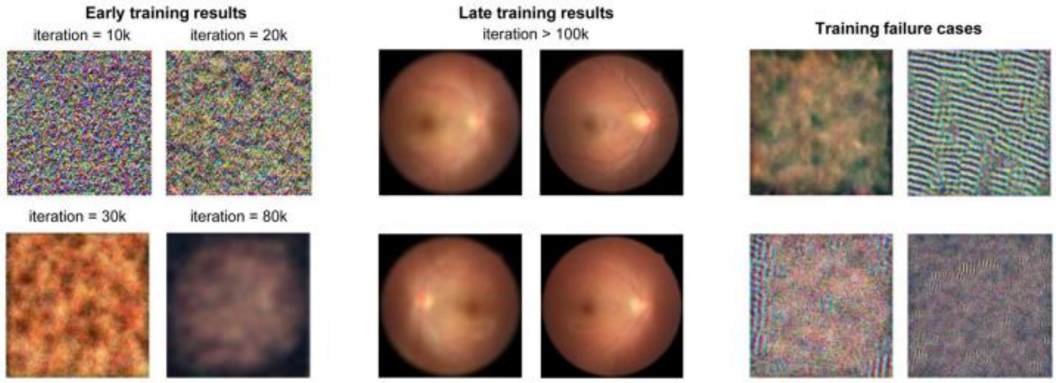


Figure 2.1.6.2 Comparison of training results of DDPM with different number of iterations

By observing Figure 2.1.6.2 above we can find that DDPM successfully generated healthy fundus images with a resolution of 128 × 128 pixels using 1,000 training samples, requiring about 1 million training iterations. This took approximately 250 hours to complete. Moreover, such results depended on the batch size and the computational power of the device. If the training iterations were too few or stopped early, confusing and erroneous images could be generated. The article also pointed out that training sometimes failed due to incorrect tuning of the learning rate and certain hyperparameters. Therefore, this study showed that DDPM was promising in the field of medical image generation, but more explorations and improvements were needed to enhance the training speed and the quality of images generated by DDPM.

2.2 Shortcomings and limitations of Related works

At first, in terms of datasets, the number of datasets on DR lesion classification is still very limited and there are different levels of image quality, which largely limits the development of related research, and the generalisability of the existing datasets is also relatively poor, as a dataset is often only fundus images collected in a small area, which is difficult to substitute for all areas of the globe, as diabetic patients in each area may experience different degrees and characteristics of fundus lesions. In order to solve these problems, different data enhancement methods or GAN image generation methods have been proposed to overcome the lack of datasets, but the quality and usability of the relevant images generated by these methods are still questionable.

In addition, there were many studies that had proposed the use of migration learning methods to overcome dataset limitations, but most of the pre-training models or pre-training weights that were being applied had been trained on the ImageNet dataset. The problem was that ImageNet did not contain medical images or images that were relevant to this task, thus questioning the effectiveness of merely fine-tuning certain layers of the pre-training model, which did not seem to be significant. Therefore, this study argued that the most traditional migration science approach should be used, which was to train all the layers of the model as a way of adapting the model to the DR lesion classification task. Alternatively, the deep learning model was pre-trained using a dataset relevant to the task.

Finally, there was also the fact that the model proposed in the study still suffered from the overfitting problem, which needed to be further adapted to the existing deep learning model, e.g., by using an early-stop training strategy, to solve the overfitting problem.

Chapter 3

Proposed Method/Approach

3.1 Research Project Implementation Workflow

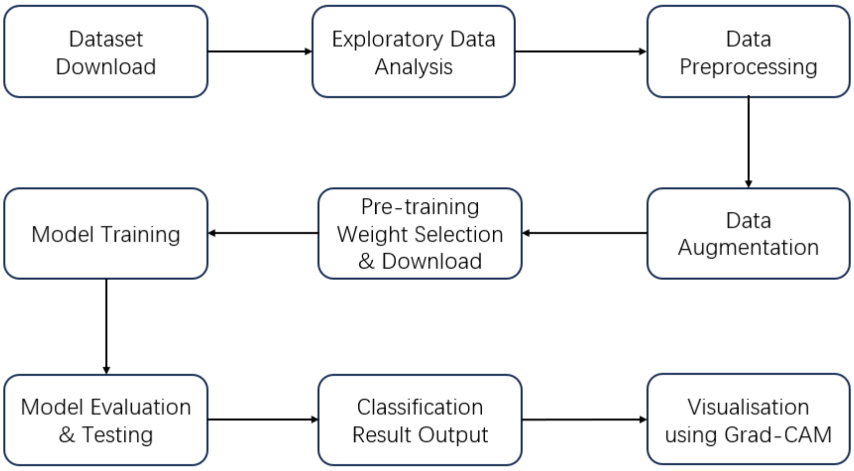


Figure 3.1.1 Project Implementation Workflow

Dataset Download

As shown in Figure 3.1.1 above. The dataset used in this study is from the official dataset of APTOS 2019 Blindness Detection competition on Kaggle platform. After downloaded the dataset from the Kaggle platform, the data need to be uploaded to Google Cloud Drive to proceed with data preprocessing, data augmentation, and model training phase.

Exploratory Data Analysia

Exploratory data Analysis is a good way to understand a dataset, so this study used data exploratory analyses aimed at understanding dataset composition and dataset structure. Additionally, it sought to understand the composition of data labels and the distribution of different numbers of labels. Moreover, it is also necessary to observe the specific fundus lesion data images in the dataset because different types of fundus images may have different lesion features, and these to these lesion features largely affect the classification performance of the model. Observing the image quality and specific data samples in the dataset also provides targeted directions and ideas for the next pre-processing step.

Data Preprocessing

Data preprocessing is also a key step in the implementation of this study, as the quality of the images and the image formats, as well as the structure of the dataset, varied greatly due to the different sources of the dataset. It had to be processed into a uniform format and size as well as structure during pre-processing in order to facilitate later model training and model testing. At the same time, the dataset needed to be partitioned proportionally into training, testing, and validation sets, which was also a key step before model training. In addition to this, special image processing methods such as grey scale threshold cropping, Gaussian blurring, and contrast enhancement were tried to reduce image noise and optimize image quality, and to crop out irrelevant parts of the image so that the model could be trained to focus on the main content.

Data Augmentation

In this research, a series of image enhancement methods were attempted to enhance fundus image data, which included random horizontal flip, random vertical flip, random rotation, and random affine. This research was conducted using the transforms module of the PyTorch library to achieve dynamic augmentation of funds images. The purpose of performing image augmentation was to overcome the effects of data imbalance. Through data augmentation techniques, it was possible to expand the categories with less sample size and increase the diversity of input data. This facilitated the model to better learn the characteristics of different categories of lesions and improved the final classification accuracy.

Pre-training Weight Selection & Download

Since this study aimed to improve the accuracy of past related work through transfer learning, it required the use of pre-training weights from existing deep learning models. Most of these pre-training weights were derived from the model being trained on a large dataset such as ImageNet. Using a pre-trained model as a starting point significantly decreased the time it took for the model to learn from scratch. This is because the pre-trained model had already learned certain features on a large-scale dataset. Also, training a complex model directly from scratch might have led to overfitting when facing a task such as medical image classification, where the amount of data is insufficient. The use of pre-training weights helped the model to learn more features on very little data, thus improving the performance of the model. Currently, pre-training weights for most models are available for download from the model's official GitHub page.

Model Training

After all the preliminary preparations had been completed, we were able to enter the model training phase, which provided GPU support with the help of relevant platforms, so that the model could learn the features of the input data more quickly, and ultimately train it into a model that could be directly applied to the DR classification task. During the training process, appropriate parameters needed to be adjusted according to different task types to prevent the model from overfitting or underfitting, which would affect the final evaluation.

Model Evaluation & Testing

For the model evaluation phase, the accuracy and loss values during model training and model validation were used to generate line plots to facilitate observation of model learning performance. Also, evaluation metrics such as the confusion matrix, classification report were applied in this phase. Finally, it was the testing of the trained model on untrained data, but also on unfamiliar datasets with different sources of relevance, which gave a good indication of the generalisation performance of the model.

Visualisation using Grad-CAM

The Grad-CAM technique was a method used to explain the predictive decisions of deep learning models. Grad-CAM highlighted the image regions of interest to the network by calculating the gradient weights. Therefore, in this study, the Grad-CAM technique was utilized to visualize the decision-making process of the proposed model. The visualization results of Grad-CAM helped ophthalmologists and related practitioners to analyze the region of attention of the model for a certain category, in order to better understand the decision-making process of the model. Meanwhile, the region of interest of the model was also used to analyze whether the network had learned the correct features or information.

3.2 Model Architecture Diagram

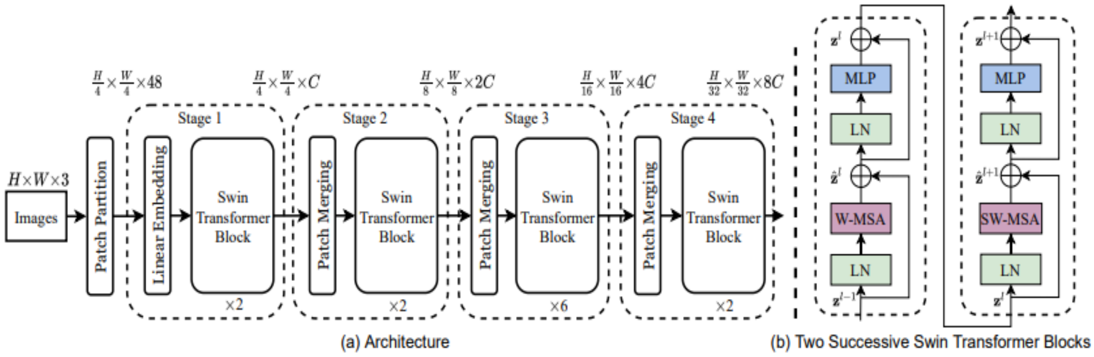


Figure 3.2.1 Architecture of the Swin Transformer model [11]

As shown in Figure 3.3.1 above, the Swin Transformer model adopts a hierarchical design, which contains a total of four stages, where each stage does a Patch Embedding at input through the Patch Merging module, which slices the image into individual patches and embeds them into the Embedding. Its primary purpose is to decrease the resolution of the input feature image. Simultaneously, Swin Transformer Block is mainly composed of modules such as Layer Normalisation (LayerNorm), Multilayer Perceptron Machine (MLP), Window Attention and Shifted Window Attention, the output of each stage in Swin Transformer model will be used as the input of the next stage, this allows the model can gradually learn from the low-level features to fine feature representation. Inside each stage, it consists of multiple Blocks in series, each of which is feature normalised by LayerNorm and then the features are nonlinearly transformed by MLP module. Window Attention mechanism enables the model to capture local features, while Shifted Window Attention enhances the global perception of the model by misaligning windows, avoiding segregation between different windows, and improving the interaction between features. This design makes Swin Transformer not only retain the advantages of Transformer's self-attention mechanism, but also introduce the idea of local windows, which effectively balances the extraction of global and local information, and thus is suitable for complex image recognition tasks.

3.3 Large Margin aware Focal Loss Function

Large Margin aware Focal Loss (LMF Loss) is a hybrid loss function for unbalanced medical image classification. It is composed of a linear combination of Focal Loss and Label-Distribution-aware Margin Loss (LDAM Loss) and is designed to better handle unbalanced datasets.

Focal Loss function is improved by Cross Entropy Loss function, where the Cross Entropy Loss function gives equal weight to all classes to learn, while the Focal Loss function adjusts the number of samples and the difficulty of the samples by introducing two modifiers, α and γ, so that the model focuses on learning a small number of classes. $p_t$is the predicted probability score. The specific equation is as follows:

$$FL(p_t) = -\alpha(1 - p_t)^{\gamma}\log(p_t) \tag{1}$$

The LDAM loss function emphasises the introduction of stronger regularisation to the minority classes than to the majority classes to reduce their generalisation errors and thus maintain the model's ability to learn from the majority classes and emphasise the minority classes. LDAM loss function focuses on the smallest margin per class and achieving uniform labelling test errors for every class, rather than encouraging the large margins of most class training samples from decision boundaries. In LDAM Loss function, x represents the sample, y the label of the sample, f the model, and z_y represents the model's output for the sample, n_j denotes the number of samples per class, and c is a fixed constant. The specific marginal value equation and the LDAM loss function equations are shown as follows:

$$\Delta_j = \frac{c}{n_j^{1/4}} \tag{2}$$

$$u = e^{z_y - \Delta_y} \tag{3}$$

$$L_{LDAM}\big((x, y), f\big) = -\log\frac{u}{u + \sum_{j \neq y} e^{z_j}} \tag{4}$$

This results in larger margins x for classes with smaller sample numbers and the model tends to reserve a larger decision boundary space for these classes when classifying them, thus effectively helping to reduce misclassification of these classes.

Focal Loss function devises a strategy to train the model to pay more attention to samples that are difficult to classify accurately, which usually includes samples with a small number of categories, while LDAM Loss function assigns weights based on the distribution of categories in the dataset. The LMF Loss function is a combination of the Focal Loss function and the LDAM Loss function, weighted by two hyperparameters, to provide better results. The LMF Loss function combines the Focal Loss function and the LDAM Loss function by weighting them with two hyperparameters so as to achieve better results. The equation is as follows:

$$L_{LMF} = \alpha L_{LDAM} + \beta L_{FL} \tag{5}$$

In the above LMF Loss function equation, α and β are constants and are considered adjustable hyperparameters. Furthermore, in [12] the authors found through iterative trials that assigning the same weights to both components can achieve better results.

Chapter 4

Experiment/Simulation

4.1 Experiment Requirement

4.1.1 Hardware

The hardware device used in this research project is a laptop, which is one of the most basic tools for the project, as a platform for the other software and to provide enough memory to store the datasets of the project. The hardware configuration used in this project is shown in Figure 4.1.1.1 below.

Table 4.1.1.1 Computer specification

Description	Specifications
Model	HP EliteBook 845
Processor	AMD Ryzen 9 PRO 6950HS with Radeon Graphics
Operating System	Windows 11 Professional Workstation Edition 64-bit version
Graphic	AMD Radeon(TM) 680 Graphics
Memory	64GB DDR4 4800MHz RAM
Storage	1TB

4.1.2 Software

The software platforms used in this research project are shown below：

i. Google Colaboratory & Kaggle Cloud GPU
ii. Jupyter Notebook
iii. PyCharm Community Edition
iv. Google Drive
v. Baidu Netdisk

They respectively provide GPUs support for training deep learning model, as well as data preprocessing, data analysis, a compilation platform for model training and cloud storage space for quickly get datasets.

4.2 Issues and Challenges During Implementation

Google Cloud Drive upload speed is too slow

In the preliminary work, the dataset needed to be uploaded locally to Google Cloud Drive as it was trained using the Google Colab platform, and then the dataset was extracted in Colab by hooking directly to Google Cloud Drive. However, due to unforeseen reasons, the data upload speed of Google Cloud Drive was very slow, so it often took a lot of time in this step. At the same time, Google Colab read data from the mounted Google Cloud Drive very slowly because the data was read directly from the Drive and transferred over the network, not from the local file disk. If there were more data files, then many network requests were sent, resulting in very slow data loading.

Google Colab and Kaggle Limited free resources

The free arithmetic resources that Google Colab offers to the public user come with usage limits. Each user can use up to 12 hours of GPU time and 36 hours of CPU at a time, with a cool-down period after the limit is reached before it can be used again. Additionally, storage is limited to 100 GB per user. If the GPU limit is reached during training and breakpoints are not saved, everything will be cleared out. Meanwhile the Kaggel platform also offers only time-limited GPU training arithmetic.

APTOS2019 datasets have a large number of missing labels

The APTOS2019 dataset does not have a test set label, so it is not possible to test with the test set for this dataset, and it may be necessary to find other relevant datasets to test with. Unless the training set in the original dataset is repartitioned into a new training set, validation set and test set.

4.3 Project Timeline

Figure 4.3.1 Project I timeline

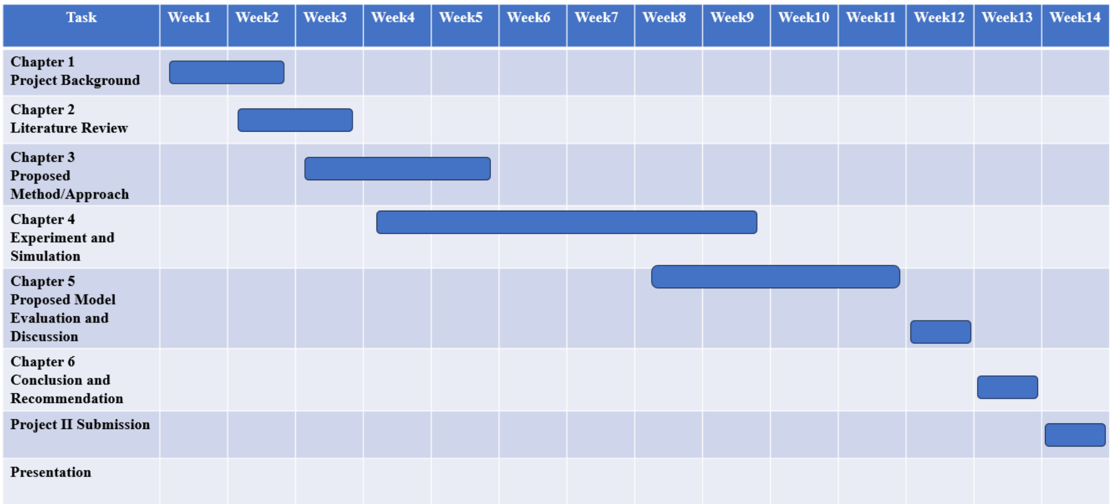


Figure 4.3.2 Project II timeline

Chapter 5

Proposed Model Evaluation and Discussion

5.1 Brief description of model evaluation metric

- **Accuracy**

Accuracy is the most intuitive performance metric for evaluating a model, representing the proportion of correctly predicted samples out of the total number of samples. Specifically, accuracy includes both True Positives (TP) and True Negatives (TN) as a proportion of all samples, which consist of TP, false positives (FP), TN and false negatives (FN). The specific equation is as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

- **Precision, Recall, and Specificity**

Precision indicates the proportion of samples predicted by the model to be in the positive category that are actually in the positive category. It is an indicator for assessing the accuracy of the model in predicting positive classes. The specific equation is as follows:

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

Recall indicates the proportion of all samples that are actually positively classified that are correctly predicted as positively classified by the model. The importance of this metric lies in assessing the model's ability to capture positively classified samples. The specific equation is as follows:

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

Specificity is the proportion of all samples that are actually negatively classified that are correctly predicted to be negative, and it measures the accuracy of the model in predicting negatively classified samples. The specific equation is as follows:

$$Specificity = \frac{TN}{TN + FP} \tag{9}$$

- **QWK value**

QWK is used to assess the consistency between the predicted results and the actual classifications of a model, making it particularly suitable for multi-class problems. Compared to simple accuracy, QWK considers the weights between classes, making it more sensitive to the severity of classification errors. This is especially important in the field of medical DL models, where classification errors of different categories have different impacts on clinical decisions. The QWK coefficient reflects these differences by assigning different weights to different classification errors. For example, misclassifying a severe disease as a mild one has a greater impact than misclassifying it as a moderate disease. QWK can reflect this difference through weighting. The specific equations are shown as follows:

$$W_{i,j} = \frac{(i-j)^2}{(N-1)^2} \tag{10}$$

$$E_{i,j} = \frac{N_i \cdot N_j}{N_t} \tag{11}$$

$$QWK = 1 - \frac{\sum_{i,j} W_{i,j} O_{i,j}}{\sum_{i,j} W_{i,j} E_{i,j}} \tag{12}$$

In above three equations, W represents the weight matrix, where $W_{i,j}$ denotes the weight between class i and class j . E represents the expected matrix, where $E_{i,j}$ denotes the probability of class i being predicted as class j under complete randomness. Here, N_i is the number of samples in the actual class i, N_j is the number of samples predicted as class j, and N_t is the total number of samples. O represents the confusion matrix, where $O_{i,j}$ denotes the number of samples with the actual class i that were predicted as class j.

- **Confusion Matrix**

Confusion Matrix is a two-dimensional chart that clearly demonstrates the accuracy of a model's predictions by comparing the actual categories with the categories predicted by the model. In this matrix, the rows represent the real categories, and the columns represent the predicted categories, thus allowing us to clearly observe the model's performance in each category identification.

5.2 Results comparison and analysis

5.2.1 Swin Transformer

By observing the accuracy and loss curves of the model in the process of training and validation as shown in figures 5.2.1.1 and 5.2.1.2 below, it was found that, no matter in the process of training or validation, the accuracy rate increased with the increase of epoch, and the validation loss had been decreasing, tending to stabilize after 150 epochs. There were small fluctuations at times, but overall, the trend generally stabilized, and no obvious signs of overfitting were observed.

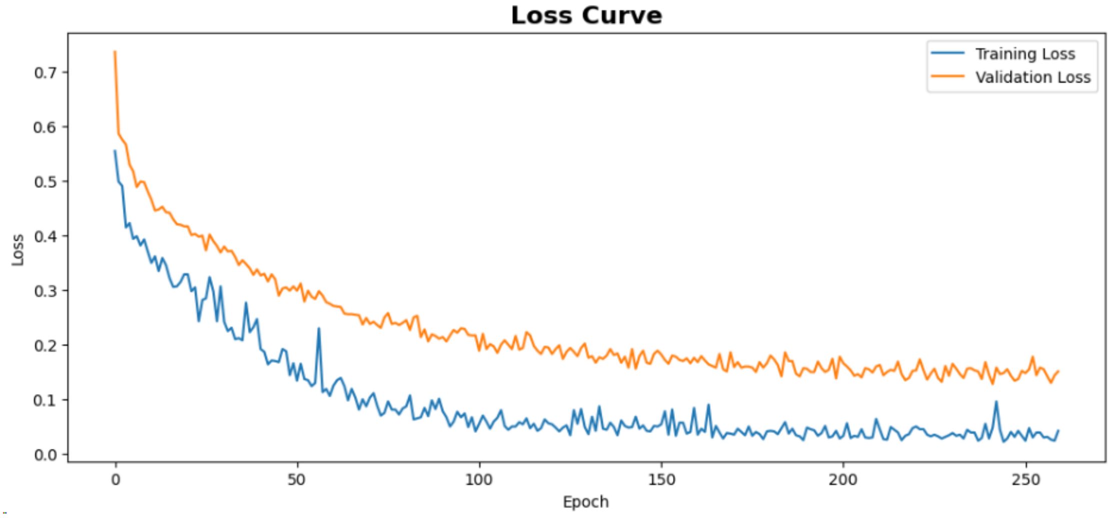


Figure 5.2.1.1 Loss Curve (Swin Transformer)

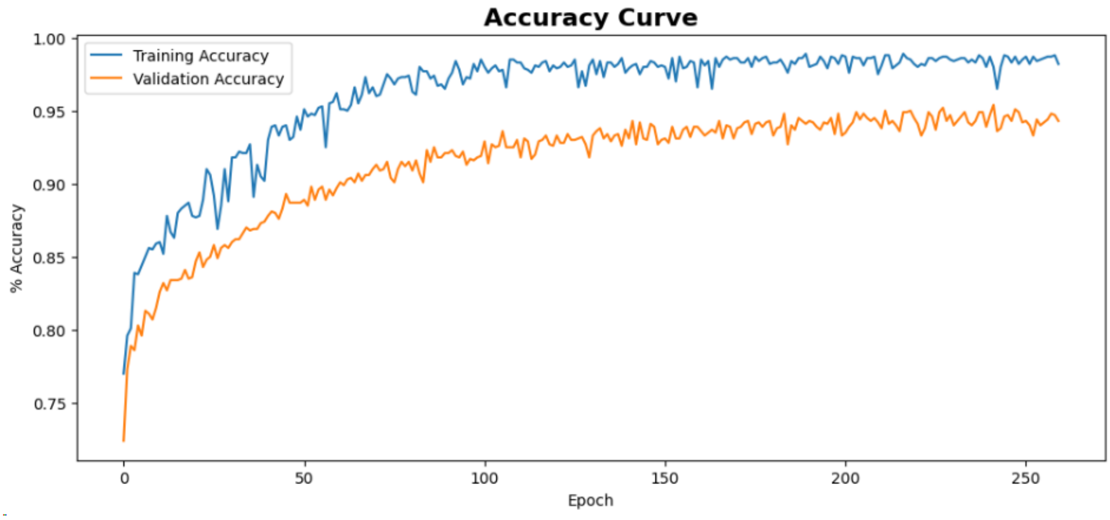


Figure 5.2.1.2 Accuracy Curve (Swin Transformer)

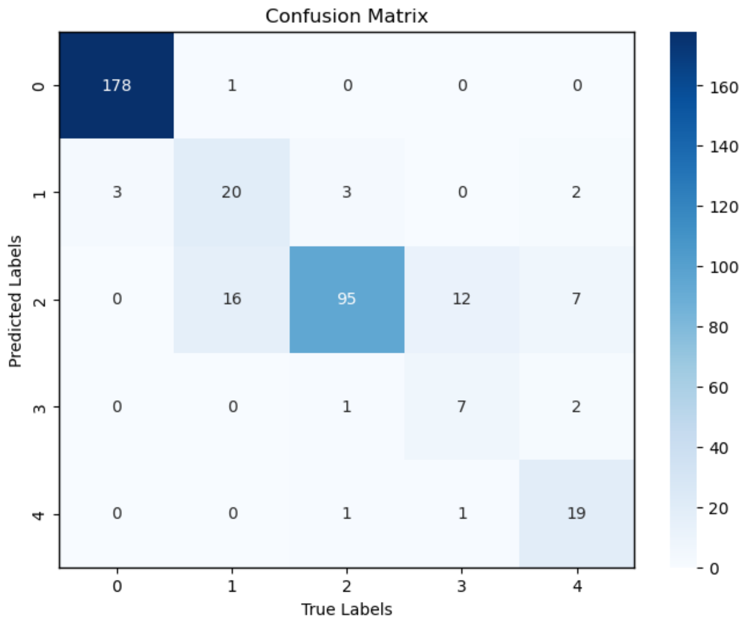


Figure 5.2.1.3 Confusion Matrix (Swin Transformer)

As shown in Figure 5.2.1.3 above, upon observation of the test confusion matrix, it was found that the Swin Transformer achieved more accurate classification for the majority of the health and lesion categories. However, the classification accuracy for the first and third categories was lower compared to the other lesion categories. The results indicated that the model had incorrectly judged 16 test samples of the first category as the second category, which may have been due to an imbalance in data distribution. Given that the number of samples in the first and third categories was less than in the other categories, this led to the model not accurately learning the characteristics of the lesions in the first and third categories, resulting in misclassification during testing. In addition, table 5.2.1.1 shows other evaluation metrics, including Precision, Precision, etc.

Table 5.2.1.1 Model evaluation metric (Swin Transformer)

Class No.	Precision	Recall	Specificity
0	0.994	0.983	0.995
1	0.714	0.541	0.976
2	0.731	0.95	0.869
3	0.7	0.35	0.991
4	0.905	0.633	0.994
Testing Accuracy：86.68%		**Testing QWK: 0.923**	

5.2.2 Swin Transformer with LMF Loss

The training process for the Swin Transformer and LMF Loss methods is illustrated in Figure 5.2.2.1 and Figure 5.2.2.2 below. Where the accuracy and loss values, although slightly fluctuating at the beginning of training, which may be related to the adjustment of the learning rate, still maintain a good upward and downward trend. After 150 epochs, the image shows that the model has gradually converged, and the final training accuracy is stable at about 0.95, and the validation accuracy is stable at about 0.92. The training lasted for a total of 300 epochs, and there was no overfitting phenomenon, the model converged normally and showed good generalization performance.

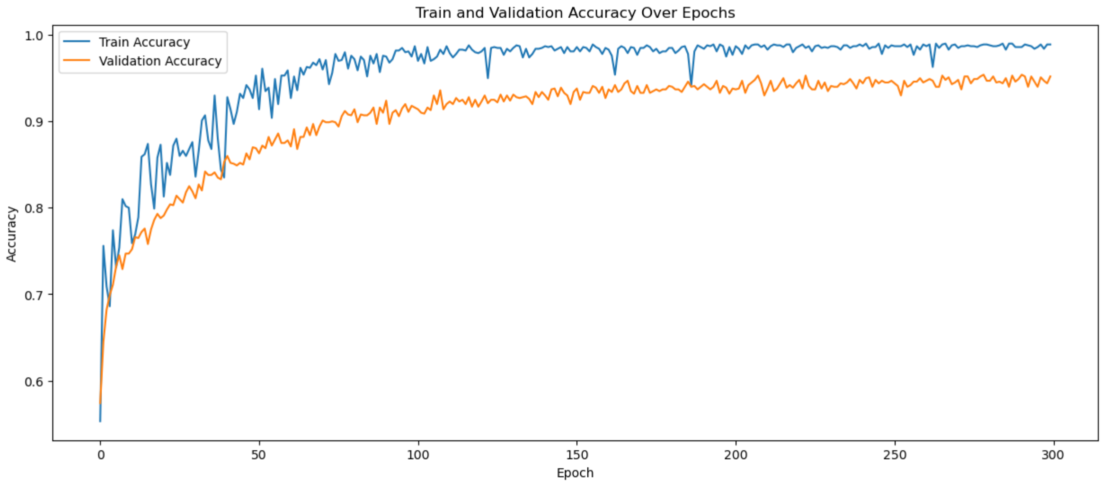


Figure 5.2.2.1 Accuracy Curve (Swin Transformer with LMF Loss)

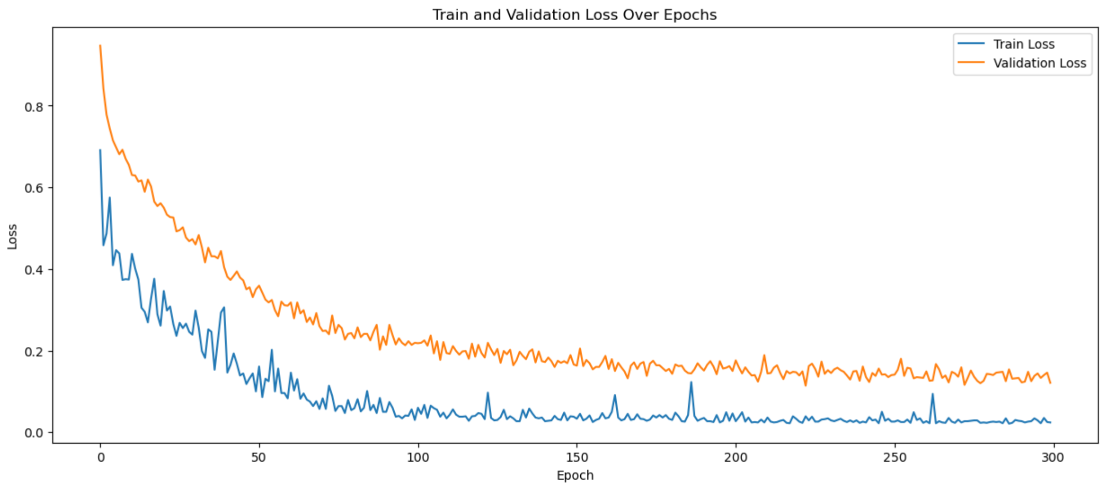


Figure 5.2.2.2 Loss Curve (Swin Transformer with LMF Loss)

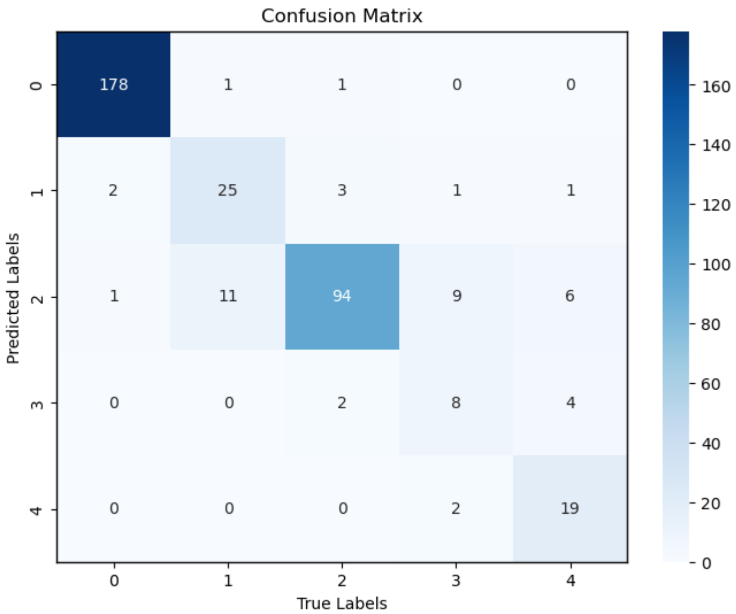


Figure 5.2.2.3 Confusion Matrix (Swin Transformer with LMF Loss)

The figure 5.2.2.3 displays the test confusion matrix of the model using the Swin Transformer with the LMF Loss method. Compared to figure 5.2.1.3, which shows the test confusion matrix of the model using the Swin Transformer with the traditional cross-entropy loss function, the accuracy of previously lower-performing categories such as Class 1 and Class 3 has significantly improved while maintaining stability in other categories. This indicates that the LMF Loss function is more effective in handling imbalanced medical image datasets. Additionally, the subsequent table 5.2.2.1 presents other model evaluation metrics, where the accuracy has improved to 88.04%, and the QWK value has also shown an improvement, demonstrating the efficacy of this method.

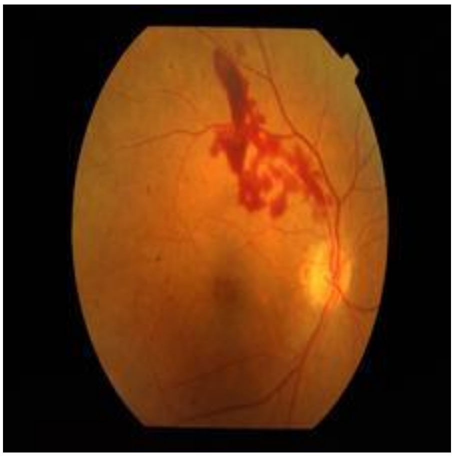
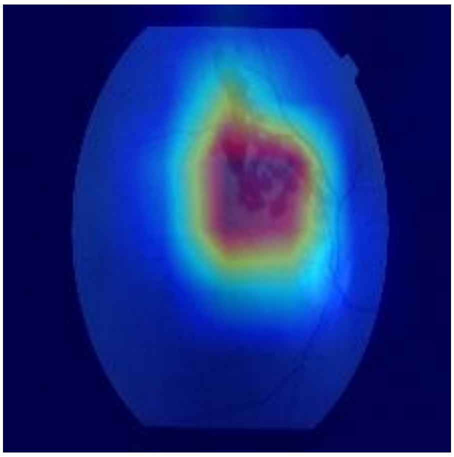
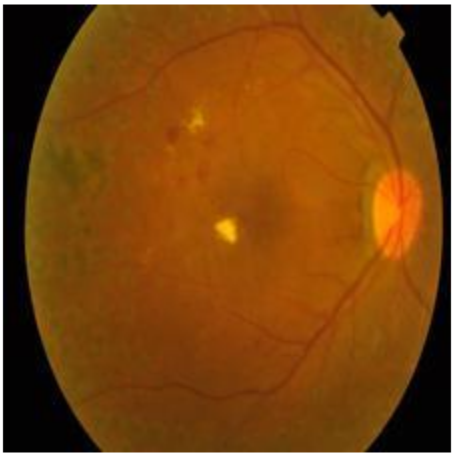
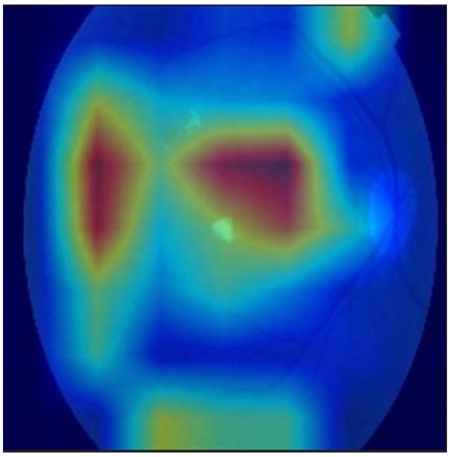
Table 5.2.2.1 Model evaluation metric (Swin Transformer with LMF Loss)

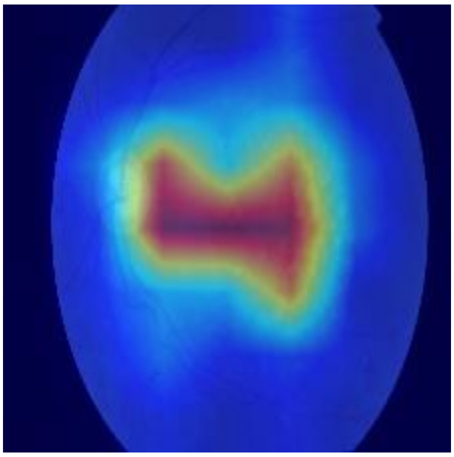
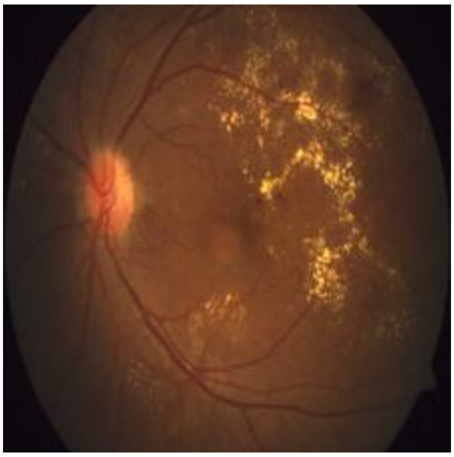
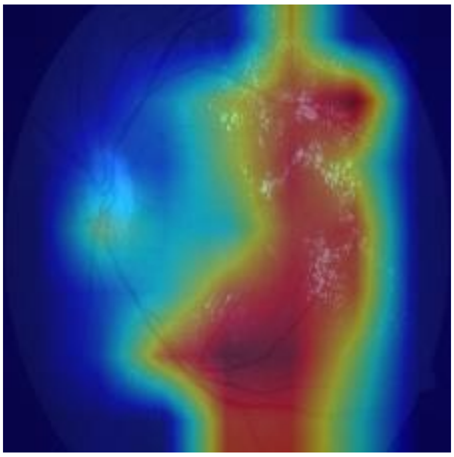
Class No.	**Precision**	**Recall**	**Specificity**
0	0.989	0.983	0.989
1	0.781	0.676	0.979
2	0.777	0.94	0.899
3	0.571	0.4	0.983
4	0.905	0.633	0.994
Testing Accuracy: 88.04%		**Testing QWK: 0.932**	

5.3 Visualisation of the model decision-making process

To better understand the decision-making process of the model, Grad-CAM was used in this study to visualize the areas of focus of the model. In the Grad-CAM visualized images, the closer to the red region, the more it was the area of focus for the model. Table 5.3.1 below shows a comparison of the output from four sets of visualizations using Grad-CAM, from which it can be seen that the areas of major focus of the model were generally those of lesions, such as hemorrhages, micro-aneurysms, proliferative vessels, and scar tissue.

Table 5.3.1 Grad-CAM Visualisation Output

Original Image	Grad-CAM
	
	

Chapter 6

Conclusion and Recommendation

6.1 Conclusion

In summary, this study proposed a method for multi-class classification of DR using deep learning and transfer learning techniques, utilizing the Swin Transformer and LMF Loss. The model was trained on the APTOS2019 dataset, achieving an accuracy of 88.04% and a QWK score of 0.932 on an independent test set, surpassing previous related studies. Transfer learning was employed to overcome the limitations of a small dataset. Without the use of pre-trained weights on large datasets, training the model from scratch could have made it difficult for the model to converge and might have led to overfitting issues. Additionally, this study implemented diverse data preprocessing, data augmentation techniques, and a novel loss function to address the imbalance in the dataset, successfully improving the classification accuracy and QWK score for classes with fewer samples, as well as the overall model performance. Finally, Grad-CAM was used to visualize the model's decision-making process, enabling ophthalmologists and professionals to better understand the model's focus areas and decision processes when further improving or applying the model in real-world scenarios.

6.2 Recommendation

In future work, more diverse deep learning models can be explored for application in DR classification tasks to achieve high-accuracy automated classification. Additionally, the use of transfer learning can be further optimized, as current pre-trained weights are not specifically trained on medical image datasets or fundus image datasets. Therefore, there is potential for improvement in the effectiveness of these pre-trained weights. In the future, it may be beneficial to collect and organize large-scale medical image datasets or fundus image datasets for pre-training deep learning models, followed by fine-tuning on smaller datasets. Furthermore, to address the imbalance in dataset sample sizes, existing image generation models could be optimized to automatically generate fundus images, thereby expanding the dataset and balancing the sample sizes across different classes. Since current image generation models often face challenges such as slow training speeds, high computational requirements, and large data demands, future work could involve the use of more powerful computing devices for training and optimizing models to improve training stability and speed. Finally, on the algorithmic level, more loss functions or feature fusion modules can be explored and integrated into existing or newer deep learning models to improve classification accuracy and model generalization.

REFERENCES

[1] Z. L. Teo et al., “Global Prevalence of Diabetic Retinopathy and Projection of Burden through 2045: Systematic Review and Meta-analysis,” Ophthalmology, vol. 128, no. 11, May 2021, doi: https://doi.org/10.1016/j.ophtha.2021.04.027.

[2] R. Helen, R. Sindhuja, V. Mounika, and M. Sharmila, “Detection and Classification of Diabetic Retinopathy by Fuzzy Based Decision-Making System,” Dec. 2023, doi: https://doi.org/10.1109/icacrs58579.2023.10404303.

[3] S. Qummar et al., “A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection,” IEEE Access, vol. 7, pp. 150530–150539, 2019, doi: https://doi.org/10.1109/access.2019.2947484.

[4] I. Kandel and M. Castelli, “Transfer Learning with Convolutional Neural Networks for Diabetic Retinopathy Image Classification. A Review,” Applied Sciences, vol. 10, no. 6, p. 2021, Mar. 2020, doi: https://doi.org/10.3390/app10062021.

[5] Hoe Yean Sam, A. Zikri, Nur, C. M. Goh, Chai Meei Tyng, and B. Murat, “Preliminary Study of Diabetic Retinopathy Classification from Fundus Images Using Deep Learning Model,” Sep. 2022, doi: https://doi.org/10.1109/aidas56890.2022.9918681.

[6] S. S. Mondal, N. Mandal, K. K. Singh, A. Singh, and I. Izonin, “EDLDR: An Ensemble Deep Learning Technique for Detection and Classification of Diabetic Retinopathy,” Diagnostics, vol. 13, no. 1, p. 124, Dec. 2022, doi: https://doi.org/10.3390/diagnostics13010124.

[7] S. Yu et al., “MIL-VT: Multiple Instance Learning Enhanced Vision Transformer for Fundus Image Classification,” Medical Image Computing and Computer Assisted Intervention – MICCAI 2021, pp. 45–54, 2021, doi: https://doi.org/10.1007/978-3-030-87237-3_5.

[8] M. Oulhadj et al., “Diabetic retinopathy prediction based on deep learning and deformable registration,” Multimedia tools and applications, vol. 81, no. 20, pp. 28709–28727, Mar. 2022, doi: https://doi.org/10.1007/s11042-022-12968-z.

[9] A. Sugeno, Y. Ishikawa, T. Ohshima, and R. Muramatsu, "Simple methods for the lesion detection and severity grading of diabetic retinopathy by image processing and transfer learning," Computers in Biology and Medicine, vol. 137, 104795, 2021. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2021.104795

[10] Hong Kyu Kim, Ik Hee Ryu, Joon Yul Choi, and Tae Keun Yoo, “A feasibility study on the adoption of a generative denoising diffusion model for the synthesis of fundus photographs using a small dataset,” Deleted Journal, vol. 6, no. 4, Apr. 2024, doi: https://doi.org/10.1007/s42452-024-05871-9.

[11] Z. Liu *et al.*, “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows,” *arXiv:2103.14030 [cs]*, Mar. 2021, Available: https://arxiv.org/abs/2103.14030

[12] A. A. Sadi et al., “LMFLOSS: A Hybrid Loss For Imbalanced Medical Image Classification,” arXiv (Cornell University), Jan. 2022, doi: https://doi.org/10.48550/arxiv.2212.12741.

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 2, Year 4	**Study week no.:** 2
Student Name & ID: Dai, Cheng Xiao, 21ACB00469	
Supervisor: Dr Sayed Ahmad Zikri Bin Sayed Aluwee	
Project Title: Improving Diabetic Retinopathy Classification using Transfer Learning and Optimized Deep Learning Models	

1. WORK DONE

Review the code and report for preliminary work (FYP1).

2. WORK TO BE DONE

Review more papers in related fields.

3. PROBLEMS ENCOUNTERED

No problems were encountered.

4. SELF EVALUATION OF THE PROGRESS

Great start, more time should be utilized for the research.

Sayed

Supervisor's signature

Dai Cheng Xiao

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT
(Project II)

Trimester, Year: Trimester 2, Year 4	**Study week no.:** 4
Student Name & ID: Dai, Cheng Xiao, 21ACB00469	
Supervisor: Dr Sayed Ahmad Zikri Bin Sayed Aluwee	
Project Title: Improving Diabetic Retinopathy Classification using Transfer Learning and Optimized Deep Learning Models	

1. WORK DONE Studied Swin Transformer and ConvNeXt model architectures
2. WORK TO BE DONE Learning diffusion models applied to medical image generation
3. PROBLEMS ENCOUNTERED No problems were encountered.
4. SELF EVALUATION OF THE PROGRESS It should be speeded up.

Sayed

Supervisor's signature

Dai Cheng Xiao

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 2, Year 4	**Study week no.:** 6
Student Name & ID: Dai, Cheng Xiao, 21ACB00469	
Supervisor: Dr Sayed Ahmad Zikri Bin Sayed Aluwee	
Project Title: Improving Diabetic Retinopathy Classification using Transfer Learning and Optimized Deep Learning Models	

1. WORK DONE Several image generation techniques were experimented. Explored different image preprocessing and image augmentation techniques.
2. WORK TO BE DONE Start training and improving the Swin Transformer model.
3. PROBLEMS ENCOUNTERED No problems were encountered.
4. SELF EVALUATION OF THE PROGRESS Don't overcomplicate things and get sidetracked from the objectives of the research.

Sayed
Supervisor's signature

Dai Cheng Xiao
Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 2, Year 4	**Study week no.:** 8
Student Name & ID: Dai, Cheng Xiao, 21ACB00469	
Supervisor: Dr Sayed Ahmad Zikri Bin Sayed Aluwee	
Project Title: Improving Diabetic Retinopathy Classification using Transfer Learning and Optimized Deep Learning Models	

1. WORK DONE

Improved Swin Transformer model architecture and loss function.

2. WORK TO BE DONE

Perform comparison and ablation experiments.

3. PROBLEMS ENCOUNTERED

GPU resource constraints that resulted in lower efficiency.

4. SELF EVALUATION OF THE PROGRESS

Reasonable use of GPU resources and rationalization of time.

Sayed

Supervisor's signature

Dai Cheng Xiao

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 2, Year 4	**Study week no.:** 10
Student Name & ID: Dai, Cheng Xiao, 21ACB00469	
Supervisor: Dr Sayed Ahmad Zikri Bin Sayed Aluwee	
Project Title: Improving Diabetic Retinopathy Classification using Transfer Learning and Optimized Deep Learning Models	

1. WORK DONE

A few improvements and experiments were completed.
Started writing the FYP2 report.

2. WORK TO BE DONE

Conduct more experiments and correctly reproduce LMFLoss

3. PROBLEMS ENCOUNTERED

No problems were encountered.

4. SELF EVALUATION OF THE PROGRESS

Need to read the paper more carefully and replicate the relevant code.

Sayed

Supervisor's signature

Dai Cheng Xiao

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT
(Project II)

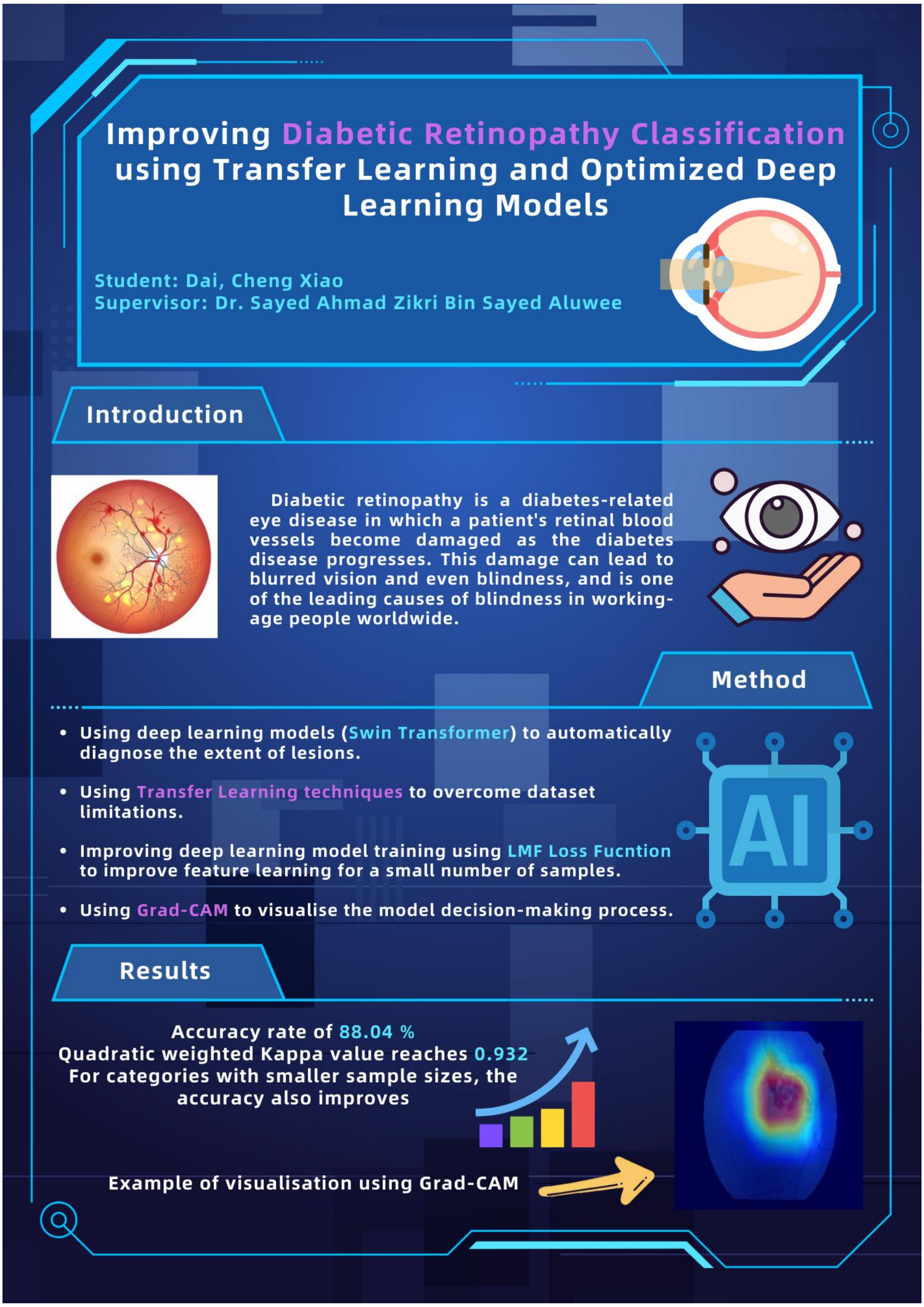
Trimester, Year: Trimester 2, Year 4	**Study week no.:** 12
Student Name & ID: Dai, Cheng Xiao, 21ACB00469	
Supervisor: Dr Sayed Ahmad Zikri Bin Sayed Aluwee	
Project Title: Improving Diabetic Retinopathy Classification using Transfer Learning and Optimized Deep Learning Models	

1. WORK DONE Final experiments were carried out. Completed FYP2 report.
2. WORK TO BE DONE Check code, report format and content syntax.
3. PROBLEMS ENCOUNTERED No problems were encountered.
4. SELF EVALUATION OF THE PROGRESS It's almost done. It needs further checking.

Sayed
Supervisor's signature

Dai Cheng Xiao
Student's signature

POSTER



Improving Diabetic Retinopathy Classification using Transfer Learning and Optimized Deep Learning Models
Student: Dai, Cheng Xiao
Supervisor: Dr. Sayed Ahmad Zikri Bin Sayed Aluwee
Introduction
Diabetic retinopathy is a diabetes-related eye disease in which a patient's retinal blood vessels become damaged as the diabetes disease progresses. This damage can lead to blurred vision and even blindness, and is one of the leading causes of blindness in working-age people worldwide.
Method
Using deep learning models (Swin Transformer) to automatically diagnose the extent of lesions.
Using Transfer Learning techniques to overcome dataset limitations.
Improving deep learning model training using LMF Loss Fucntion to improve feature learning for a small number of samples.
Using Grad-CAM to visualise the model decision-making process.
AI
Results
Accuracy rate of 88.04 %
Quadratic weighted Kappa value reaches 0.932
For categories with smaller sample sizes, the accuracy also improves
Example of visualisation using Grad-CAM

PLAGIARISM CHECK RESULT

FYP2.pdf

ORIGINALITY REPORT

12%	7%	8%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	"Image Processing, Electronics and Computers", IOS Press, 2024 Publication	1%
2	www.mdpi.com Internet Source	1%
3	arxiv.org Internet Source	1%
4	Submitted to University of Edinburgh Student Paper	1%
5	ouci.dntb.gov.ua Internet Source	<1%
6	"Medical Image Computing and Computer Assisted Intervention – MICCAI 2022", Springer Science and Business Media LLC, 2022 Publication	<1%
7	mdpi-res.com Internet Source	<1%
8	www.nature.com Internet Source	<1%

9	Hong Kyu Kim, Ik Hee Ryu, Joon Yul Choi, Tae Keun Yoo. "A feasibility study on the adoption of a generative denoising diffusion model for the synthesis of fundus photographs using a small dataset", Discover Applied Sciences, 2024 Publication	<1 %
10	Submitted to Universiti Tunku Abdul Rahman Student Paper	<1 %
11	Hanping Hu, Shuai Wang, Chenjie Zhang, Yue Pan. "Garbage Image Classification Algorithm Based on Swin Transformer*", 2023 WRC Symposium on Advanced Robotics and Automation (WRC SARA), 2023 Publication	<1 %
12	Submitted to Coventry University Student Paper	<1 %
13	Cheng Wan, Rongrong Hua, Kunke Li, Xiangqian Hong, Dong Fang, Weihua Yang. "Automatic Diagnosis of Different Types of Retinal Vein Occlusion Based on Fundus Images", International Journal of Intelligent Systems, 2023 Publication	<1 %
14	repository.nida.ac.th Internet Source	<1 %
	redc.revistas-csic.com	

15	Internet Source	<1%
16	www.unboundmedicine.com Internet Source	<1%
17	www.frontiersin.org Internet Source	<1%
18	www.researchgate.net Internet Source	<1%
19	"Machine Learning in Medical Imaging", Springer Science and Business Media LLC, 2024 Publication	<1%
20	issuu.com Internet Source	<1%
21	Chrysovalantis Voutouri, Demetris Englezos, Constantinos Zamboglou, Iosif Strouthos et al. "A convolutional attention model for predicting response to chemo-immunotherapy from ultrasound elastography in mouse tumor models", Springer Science and Business Media LLC, 2023 Publication	<1%
22	"Medical Image Computing and Computer Assisted Intervention – MICCAI 2021",	<1%

Springer Science and Business Media LLC, 2021
Publication

23	C.H. Chen. "Signal and Image Processing for Remote Sensing", CRC Press, 2024 Publication	<1 %
24	"Artificial Neural Networks in Pattern Recognition", Springer Science and Business Media LLC, 2018 Publication	<1 %
25	miccai2021.org Internet Source	<1 %
26	pubmed.ncbi.nlm.nih.gov Internet Source	<1 %
27	A. Rosline Mary, P. Kavitha. "Diabetic retinopathy disease detection using shapley additive ensembled densenet-121 resnet-50 model", Multimedia Tools and Applications, 2024 Publication	<1 %
28	par.nsf.gov Internet Source	<1 %
29	www.brownstoneresearch.com Internet Source	<1 %
30	ar.scribd.com Internet Source	<1 %

31	ntrs.nasa.gov Internet Source	<1 %
32	Tengping Jiang, Bisheng Yang, Yongjun Wang, Lei Dai, Bo Qiu, Shan Liu, Shiwei Li, Qinyu Zhang, Xin Jin, Wenjun Zeng. "RailSeg: Learning Local-Global Feature Aggregation with Contextual Information for Railway Point Cloud Semantic Segmentation", IEEE Transactions on Geoscience and Remote Sensing, 2023 Publication	<1 %
33	Zhigang Ding, Fucheng Fu, Jishi Zheng, Haiyan Yang, Fumin Zou, Kong Linghua. "Intelligent Wood Inspection Approach Utilizing Enhanced Swin Transformer", IEEE Access, 2024 Publication	<1 %
34	dot.ca.gov Internet Source	<1 %
35	journals.plos.org Internet Source	<1 %
36	learn.microsoft.com Internet Source	<1 %
37	www.ijabe.org Internet Source	<1 %

38	G. Sai Chaitanya Kumar, R. Kiran Kumar, K. Parish Venkata Kumar, N. Raghavendra Sai, Madamachi Brahmaiah. "Deep Residual Convolutional Neural Network: An Efficient Technique for Intrusion detection system", Expert Systems with Applications, 2023 Publication	<1 %
39	Ri-Gui Zhou, Shuaidong Chang, Yaochong Li. "A neural network architecture for information extraction in Chinese drug package insert", IEEE Access, 2020 Publication	<1 %
40	Saad Ather, Aamir Wali, Tayyaba Gul Malik, Khawaja Muhammad Fahd, Sana Fatima. "A novel vessel extraction technique for a three-way classification of diabetic retinopathy using cascaded classifier", Multimedia Tools and Applications, 2024 Publication	<1 %
41	espace.curtin.edu.au Internet Source	<1 %
42	pdffox.com Internet Source	<1 %
43	Ayaka Sugeno, Yasuyuki Ishikawa, Toshio Ohshima, Rieko Muramatsu. "Simple methods for the lesion detection and severity grading of diabetic retinopathy by image processing	<1 %

and transfer learning", Computers in Biology and Medicine, 2021
Publication

44	Da-zhou Li, Xin Xu, Jia-heng Pan, Wei Gao, Shi-rui Zhang. "Image2InChI: Automated Molecular Optical Image Recognition", Journal of Chemical Information and Modeling, 2024 Publication	<1 %
45	Hoe Yean Sam, Sayed Ahmad Zikri B Sayed Aluwee, Nur Syadhila Bt Che Lah, C.M. Goh, Chai Meei Tyng, Haizul Ikhwan B Murat. "Preliminary Study of Diabetic Retinopathy Classification from Fundus Images Using Deep Learning Model", 2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS), 2022 Publication	<1 %
46	Muhammad Zaheer Sajid, Muhammad Fareed Hamid, Ayman Youssef, Javeria Yasmin et al. "DR-NASNet: Automated System to Detect and Classify Diabetic Retinopathy Severity Using Improved Pretrained NASNet Model", Diagnostics, 2023 Publication	<1 %
47	Submitted to University of Southampton Student Paper	<1 %
48	Yangyang Zhu, Zheling Meng, Hao Wu, Xiao Fan, Wenhao lv, Jie Tian, Kun Wang, Fang Nie.	<1 %

"Deep learning radiomics of multimodal ultrasound for classifying metastaticcervicallymphadenopathy into primary cancer sites: a feasibility study", Ultraschall in der Medizin - European Journal of Ultrasound, 2023
Publication

49	link.springer.com Internet Source	<1 %
50	scholarworks.iupui.edu Internet Source	<1 %
51	tohoku.repo.nii.ac.jp Internet Source	<1 %
52	"Advances in Information and Communication", Springer Science and Business Media LLC, 2024 Publication	<1 %
53	Jingbo Hu, Huan Wang, Le Wang, Ye Lu. "Graph Adversarial Transfer Learning for Diabetic Retinopathy Classification", IEEE Access, 2022 Publication	<1 %

Exclude quotes On Exclude matches Off
Exclude bibliography On

Universiti Tunku Abdul Rahman			
Form Title : Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)			
Form Number: FM-IAD-005	Rev No.: 0	Effective Date: 01/10/2013	Page No.: 1of 1

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Full Name(s) of Candidate(s)	Dai, Cheng Xiao
ID Number(s)	21ACB00469
Programme / Course	BACHELOR OF COMPUTER SCIENCE (HONOURS)
Title of Final Year Project	Improving Diabetic Retinopathy Classification using Transfer Learning and Optimized Deep Learning Models

Similarity	**Supervisor's Comments (Compulsory if parameters of originality exceeds the limits approved by UTAR)**
Overall similarity index: 12 % **Similarity by source** Internet Sources: 7 % Publications: 8 % Student Papers: 2 %	
Number of individual sources listed of more than 3% similarity: 0	
Parameters of originality required and limits approved by UTAR are as Follows: **(i) Overall similarity index is 20% and below, and** **(ii) Matching of individual sources listed must be less than 3% each, and** **(iii) Matching texts in continuous block must not exceed 8 words** ***Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.***	

Note Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.

Sayed Ahmad Zikri

______________________________ ______________________________

Signature of Supervisor Signature of Co-Supervisor

Name: Sayed Ahmad Zikri Bin Sayed Aluwee Name: ______________________

Date: 09/13/2024 Date: ______________________

UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY (KAMPAR CAMPUS)

CHECKLIST FOR FYP2 THESIS SUBMISSION

Student Id	21ACB00469
Student Name	Dai, Cheng Xiao
Supervisor Name	Dr Sayed Ahmad Zikri Bin Sayed Aluwee

TICK (√)	**DOCUMENT ITEMS** Your report must include all the items below. Put a tick on the left column after you have checked your report with respect to the corresponding item.
√	Title Page
√	Signed Report Status Declaration Form
√	Signed FYP Thesis Submission Form
√	Signed form of the Declaration of Originality
√	Acknowledgement
√	Abstract
√	Table of Contents
√	List of Figures (if applicable)
√	List of Tables (if applicable)
	List of Symbols (if applicable)
√	List of Abbreviations (if applicable)
√	Chapters / Content
√	Bibliography (or References)
√	All references in bibliography are cited in the thesis, especially in the chapter of literature review
	Appendices (if applicable)
√	Weekly Log
√	Poster
√	Signed Turnitin Report (Plagiarism Check Result - Form Number: FM-IAD-005)
√	I agree 5 marks will be deducted due to incorrect format, declare wrongly the ticked of these items, and/or any dispute happening for these items in this report.

*Include this form (checklist) in the thesis (Bind together as the last page)

I, the author, have checked and confirmed all the items listed in the table are included in my report.

Dai Cheng Xiao

(Signature of Student)

Date:09/13/2024