Diabetic Retinopathy Grading: VGG16 Based on Transfer Learning and Synthetic Minority Over-Sampling Technique

By

Junqi Huang, Jess
(2130005028)

A Final Year Project thesis (STAT4004)
submitted in partial fulfillment of the requirements
for the degree of

Bachelor of Science (Honours) in Statistics

at

BNU-HKBU
UNITED INTERNATIONAL COLLEGE

December, 2024

DECLARATION

I hereby declare that all the work done in this Project is of my independent effort. I also certify that I have never submitted the idea and product of this Project for academic or employment credits.

Junqi Huang, Jess
(2130005028)

Date: December 30, 2024

Abstract

Diabetic Retinopathy(DR) is one of the leading causes of visual impairment, so early detection is essential to prevent serious complications. In this study, we will develop an effective DR severity grading neural network by utilizing a **transfer learning** technique with **VGG16** architecture. With the VGG16 model pre-trained on the ImageNet dataset and fine-tuned for the task of diabetic retinal image grading, the model was able to efficiently recognize five severity classes (no DR, mild DR, moderate DR, severe DR, and proliferative DR). Key Features of the process include assessing image quality by brightness and contrast, employing data balancing strategies **Borderline-SMOTE**, and comprehensively **fine-tuning** pre-trained VGG16 to improve model generalization.

The experimental results show that the model is robust with **91.07%** training accuracy and **85.26%** validation accuracy. In terms of category-specific performance, quantitative analysis by metrics such as Precision, Recall and F1-score shows that the model exhibits excellent grading ability for severe categories such as proliferative DR (F1-score: **0.99**) and severe DR (F1-score: **0.97**). However, for the normal, mild and moderate stages, the model still has some challenges with lesion detection, exhibiting confusion with each other. The visualization of training process further validates the model's ability to minimize misgrading at advanced stages, while revealing some room for improvement in early stage detection.

This study emphasizes the effectiveness of transfer learning in DR grading and demonstrates the potential of combining advanced data preprocessing techniques with deep learning models.

Keywords

Diabetic Retinopathy, Borderline-SMOTE, Transfer Learning, VGG16

Contents

Chapter 1

Introduction of Diabetic Retinopathy

1.1 Background

With rising living standards and an aging population, diabetes has become a common condition for an increasing number of people. In addition to its impact on overall health, diabetes can cause damage to several body parts, including the eyes, brain, kidneys, heart arteries, blood vessels, and skin. In this article, we will focus on diabetic retinopathy (DR), also known as "sugar mesh," which is a condition in which diabetic patients develop microvascular tumors, hemorrhages, hard exudates, cotton-wool spots, microvascular abnormalities, venous bead-like changes in the retina, and other abnormalities in the retina. DR refers to microvascular tumors, hemorrhages, hard exudates, cotton wool spots, intraretinal microvascular abnormalities, venous bead-like changes, and neovascularization, vitreous hemorrhage, and fibrous proliferation in the retina of diabetic patients.

Studies have shown that the longer the duration of diabetes, the higher the incidence of DR. The prevalence of DR is about 30% in diabetic patients with less than five years of diabetes and up to 90% in patients with more than ten years of diabetes. In addition, almost all patients with type 1 diabetes and 60% of patients with type 2 diabetes will develop varying degrees of retinopathy

when the disease reaches a duration of 20 years or more. Notably, the pathogenesis of DR is irreversible. Many diabetic patients do not show obvious symptoms during the initial examination, but by the time they seek medical attention for vision loss, the fundus is already severe. Therefore, it is particularly important to accurately help patients determine the condition of the retina in the fundus, not only to detect the onset of the disease at an early stage, but also to assist the doctor in treating the patient in a timely manner.

Traditional methods of diagnosing fundus lesions rely on ophthalmologists to manually analyze fundus images, which is subjective, time-consuming, and difficult to detect early lesions. This not only affects the consistency of diagnosis, but also fails to meet the demand for mass screening. In contrast, automated image grading technology has significant advantages. First, it can quickly process a large number of images, significantly improving screening efficiency; second, it reduces diagnostic discrepancies caused by human factors through standardized algorithms and enhances the consistency of diagnostic results; in addition, the automated system can more accurately identify early-stage lesions, thus improving the accuracy of early diagnosis. These advantages make automated image grading an important solution to address the limitations of traditional methods.

1.2 Objectives and Significance

Traditional methods of diagnosing fundus lesions mainly rely on manual analysis of fundus images by ophthalmologists, but have limitations such as high subjectivity, time-consuming and laborious, as well as difficulty in detecting early lesions. This not only affects the consistency of diagnosis, but also fails to meet the demand for mass screening. In contrast, automated image grading technology has significant advantages. First, it can quickly process a large number of images, significantly improving screening efficiency; second, it reduces diagnostic discrepancies caused by human factors through standardized

algorithms, enhancing the consistency of diagnostic results; in addition, the automated system can more accurately identify early-stage lesions, thus improving the accuracy of early diagnosis. These advantages make automated image grading an important solution to deal with the limitations of traditional methods. Therefore, the main goal of this study was to develop an effective image grading method for Diabetic Retinopathy (DR). By using advanced image processing techniques, it aims to improve the accuracy and efficiency of DR grading and overcome the limitations of traditional manual grading methods.

The study is significant in the field of medical diagnosis and can provide a reliable tool for early detection and monitoring of DR, which is crucial for preventing vision loss in diabetic patients. In addition, the study enriches the field of image processing by demonstrating the application of automated grading systems. These advances can be extended to other medical imaging tasks, ultimately enhancing diagnostic accuracy and improving healthcare outcomes.

1.3 Research Content and Methods

The aim of this study is to develop a transfer learning-based image grading method for diabetic retinopathy (DR), which mainly uses the VGG16 model for image grading. First, for the collected fundus image data, preprocessing was performed to ensure data quality. Specifically, lower quality images were screened out by the brightness and contrast of the images to enhance the effectiveness and accuracy of the subsequent model training. Subsequently, considering the possible imbalance in the number of samples of different DR classes, an improved method based on SMOTE (Synthetic Minority Over-sampling Technique) technique was used to over-sample the data to generate synthetic samples to balance the data distribution of each level, preventing the model from over-fitting to most classes during the training process. overfitting phenomenon for most classes during the training process.

In terms of model construction, transfer learning is performed based on the

pre-trained VGG16 model, and the grading accuracy and efficiency are improved by fine-tuning the model parameters to better adapt to the features of DR images. In order to comprehensively assess the effectiveness of the constructed model, the confusion matrix is mainly used as the evaluation metric, and the strengths and weaknesses of the model in different categories are further explored by analyzing the grading performance of the model on each DR level. Specifically, the study will focus on analyzing whether the model's grading effect at some DR levels is better than others, so as to identify the model's strengths and weaknesses. Based on the evaluation results, the model will be further optimized and adjusted, improving its overall grading performance and robustness.

Chapter 2

Literature Review

2.1 Overview of Diabetic Retinopathy

Diabetic retinopathy is one of the most common microvascular complications of diabetes mellitus, and its incidence increases with age and prolonged duration of diabetes.People with a history of diabetes mellitus for more than 10 years or more than half have DR, and DR is an important cause of blindness in adults.DR is classified into two categories according to the fundus alterations, the proliferative type and the nonproliferative type: the nonproliferative type is the early stage of the disease and is confined to the intraretinal area, which manifests itself as microvascular aneurysms and hemorrhages, The nonproliferative type is an early stage of the disease and is limited to the inner retina, showing microvascular tumors, hemorrhages, hard and soft exudates, and retinal arterial and venous lesions.(3)

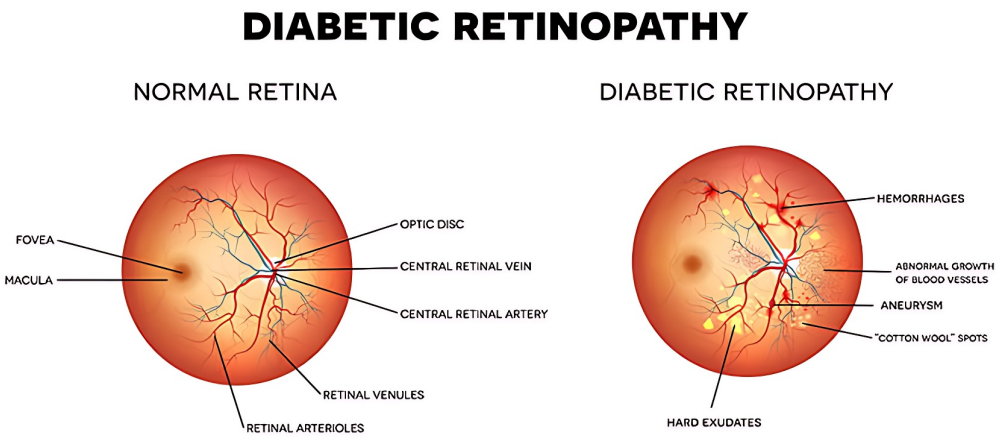


Figure 2.1: Normal Retina and DR Fundus

The pathogenesis of DR is complex, and long-term chronic hyperglycemia is the basis of its pathogenesis. Continuous hyperglycemic environment will cause the blood-retinal barrier to be destroyed at the early stage of DR, which is manifested as: relaxation of the tight junctions between microvascular endothelial cells, increased permeability, thickening of the capillary basement membrane, increased stiffness of the microvessels, disappearance of pericytes surrounding capillaries, and formation of balloon-like changes in the cavities of the capillary wall; overproliferation of endothelial cells, resulting in occlusion of the capillaries, small hemorrhages, and lipid deposition (rigid exudation), and ultimately the complete loss of the retinal microvascular cellular structure and the emergence of capillary acellularization. Overproliferation of endothelial cells leads to capillary occlusion, small hemorrhages, and lipid deposition (sclerotic exudation), culminating in the complete loss of retinal microvascular cytoarchitecture and anaplasticity of the capillaries.

2.2 Diabetic Retinopathy Image Severity Scale

For widespread everyday clinical use, the ETDRS severity scale has been further simplified into the International Clinical Diabetic Retinopathy (ICDR) severity scale. Five levels of retinopathy severity—No DR, Mild DR, Moderate DR, Severe DR, and Proliferate DR—are essentially extracted from the 14 ETDRS severity levels by the ICDR severity scale. The ICDR severity scale is by far the most widely used grading system in clinical practice worldwide due to its practicality and simplicity of use.(7)

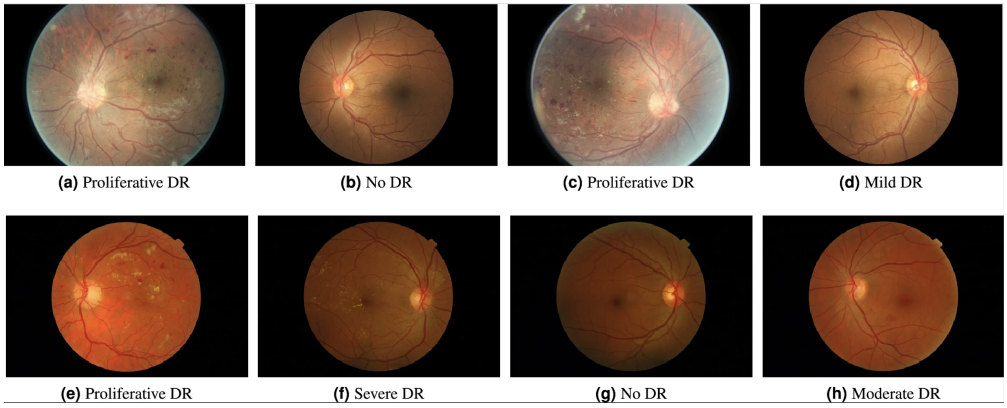


Figure 2.2: ICDR Severity Scale

The ICDR severity scale was developed using the ETDRS and WESDR data and categorisation systems following an international consensus workshop in 2002. In 2001, a preliminary planning conference was held in conjunction with the annual conference of the American Academy of Ophthalmology (AAO) and involved five different countries. Then, in 2002, at the International Congress of Ophthalmology in Sydney, 14 representatives from 11 different countries reached a consensus and established the ICDR using a modified Delphi method.

2.3 Machine Learning and Deep Learning in Image Grading

The advancement of machine learning (ML) and deep learning (DL) techniques has significantly enhanced the capabilities of automated diabetic retinopathy (DR) image grading systems. These approaches utilize sophisticated algorithms to analyze retinal images, enabling early detection and precise grading of DR severity levels

2.3.1 Traditional Machine Learning Approaches

Early efforts in DR image grading predominantly employed traditional ML algorithms, which relied on handcrafted features extracted from retinal images. Commonly used algorithms included Support Vector Machines (SVM), Random Forests, and K-Nearest Neighbors (KNN). These methods typically involved preprocessing steps such as image enhancement and feature extraction to identify key indicators of DR, such as microaneurysms, hemorrhages, and exudates. While these traditional ML approaches demonstrated reasonable accuracy in grading tasks, they often faced challenges related to scalability and adaptability. The necessity for extensive domain knowledge in feature engineering limited their applicability, and variations in image quality could adversely affect their performance.

2.3.2 Deep Learning Innovations

The emergence of deep learning, particularly Convolutional Neural Networks (CNNs), marked a significant breakthrough in DR image grading. Unlike traditional ML methods, CNNs automatically learn hierarchical feature representations directly from raw images, eliminating the need for manual feature extraction. Architectures such as AlexNet, VGG16, ResNet, and Inception have

been extensively applied to DR grading tasks with notable success. These deep learning models excel in capturing complex patterns and subtle features within retinal images, leading to higher grading accuracy and robustness compared to traditional methods.

2.3.3 Transfer Learning

Transfer learning has further enhanced the performance of DL models in DR grading. By leveraging pre-trained networks on large-scale datasets like ImageNet, models can be fine-tuned on specific DR datasets, improving their generalization capabilities even with limited labeled data. This approach not only reduces training time but also capitalizes on the rich feature representations learned from diverse image data, resulting in superior grading performance.

2.4 Current Research Status and Analysis

The use of machine learning algorithms to anticipate and identify DR is gaining traction among researchers and medical professionals. Machine learning has been used to develop a number of techniques for DR prediction. With an emphasis on the diabetic retinopathy level 1 (DR1), segmentation and indexing methods in the Domain of Retinal Ophthalmology (MESSIDOR), and the Kaggle diabetic retinopathy dataset, this section examines important studies that employed machine learning for DR prediction.

Qomariah et al.(9) created a novel method for DR detection that combines support vector machines (SVMs) and convolutional neural networks (CNNs). They used 77 and 70 eye photos from the 12th and 13th bases, respectively, in the Messidor database to test their methods. For base 12, the highest accuracy of 95.83% was obtained by combining resnet50, transfer learning, and SVM; for base 13, the highest accuracy of 95.24% was obtained by combining

Inception v3 and VGGNet type 19. According to the study, DR categorisation accuracy can be increased by combining CNN transfer learning and SVM features.

Doshi et al.(8) led a study that focused on deep convolutional neural networks specifically designed for DR detection and grading from colour fundus images, continuing the tradition of investigating the possibilities of CNNs. They examined three distinct CNN models as part of their diversified strategy, and the combined model produced a noteworthy kappa-squared result of 0.3996. The usefulness of deep learning methods in improving DR detection accuracy was further shown by this experiment.

A method based on MobileNetv2 that makes use of transfer learning and fine-tuning techniques was proposed by Patel and Chaware. The Kaggle dataset on diabetic retinopathy was used to assess the effectiveness of their method. In their experiment, they used 2,929 retinal fundus images to train the model and 733 images to validate it. Following fine-tuning, the network training accuracy significantly improved from 70% to 91%, and the validation accuracy rose from 50% to 81%(10).

Chapter 3

Fundus Image Preprocessing

3.1 Overview

The dataset used in this study was first put together for a 2015 Kaggle competition. It is important to note that this dataset differs greatly from other typical Kaggle datasets. Every picture in the collection was taken by a distinct person, with a range of cameras, and is displayed in various sizes. With 35,126 images in all, the collection offers a wealth of analytical resources.The dataset was initially intended to be used to predict five classes of DR levels, each representing a different stage of retinopathy, which we can analyze through the Table 3.1.

Table 3.1: Different Stages of DR

Stages of DR	**Descriptions**	**Number of Images**
Normal (No DR)	Without any abnormalities.	25,810
Mild NPDR	Presence of microaneurysms only.	2,443
Moderate NPDR	Microaneurysms are present but in smaller amounts as compared to severe NPDR.	5,292
Severe NPDR	Venous beading in two or more regions. Prominent intraretinal microvascular abnormality (IRMA) in one or more regions.	873
Proliferative DR	Vitreous/pre-retinal hemorrhage. Neovascularization.	708

With Table 3.1, we can find a noteworthy sample characteristic, the No DR stage has a sample size of 25,180, which is about 73% of the total sample size, while the Proliferate DR, which has the smallest sample size, has only 708, which is only 2% of the total sample size. Thus, we have a serious data imbalance problem in this dataset. We can also visualize the data sample imbalance problem more visually with Figure 3.1:

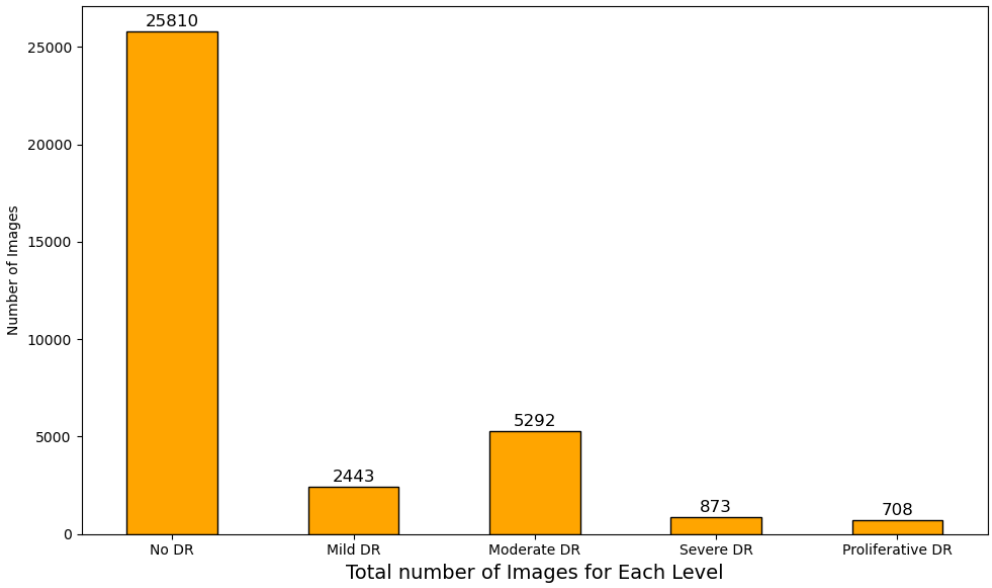


Figure 3.1: Distribution of Each DR Level

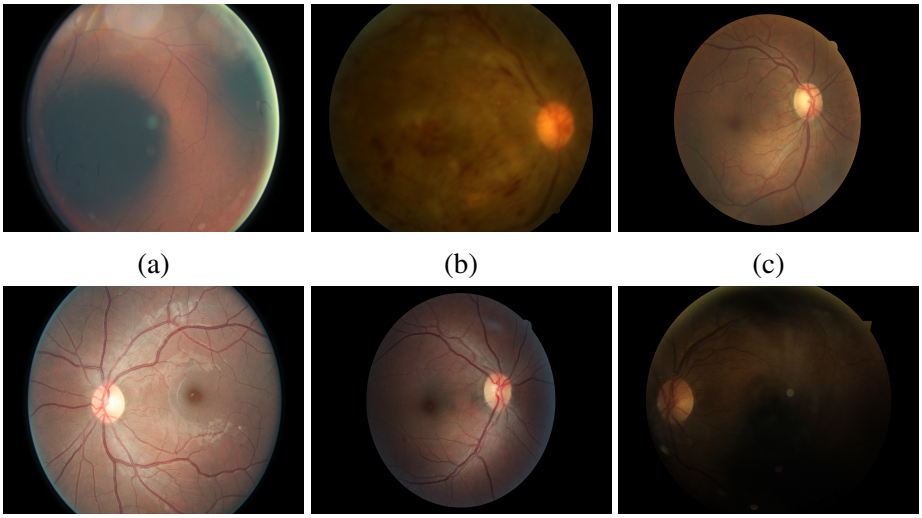
For implementing an effective deep learning neural network model, an understanding of the dataset is crucial. The presence of anomalous images in the dataset as well as imbalance in the data categories will severely contribute to the performance of the model. Therefore in order to train a stable and accurate model, data preprocessing is crucial. In this thesis, the main preprocessing of the data will be done by using the improved technique based on SMOTE, which will be specified in data preprocessing.

3.2 Image Quality Assessment

A high-quality image dataset is crucial for the training of deep learning models, which can enhance the performance of the models and better accomplish specific tasks. Therefore, in this section, we will preprocess the Kaggle dataset, mainly screening some of the lower quality images present and synthesizing a few column-specific samples to solve the data imbalance problem through the Smote technique.

3.2.1 Quality Issue

Figure 3.2 shows some fundus images in the dataset, and it can be seen that there are some abnormal images in the dataset, such as images a,b,f. These images are particularly blurred and dimly lit, which will affect the training effect of the model if such images are not removed.



(a) (b) (c)

(d) (e) (f)

Figure 3.2: Fundus of Kaggle Dataset

We filter out high quality images that meet the criteria by analyzing the

brightness and contrast of the original dataset. We first convert the color map to a gray image and then calculate the contrast and brightness of the image.

3.2.2 Quality Evaluation Metrics

To filter images that meet the requirements, we use the following two metrics:

1. **Brightness:** The brightness of the image is calculated as the mean value of the grayscale image, using the following formula:

$$\text{Brightness} = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} I(i,j) \tag{3.1}$$

   where:

   - H and W represent the height and width of the image, respectively.
   - $I(i,j)$ represents the grayscale intensity at position (i,j).

   The brightness of the image must satisfy the following condition:

$$50 \leq \text{Brightness} \leq 100$$

2. **Contrast:** The contrast of the image is calculated as the standard deviation of the grayscale image, using the following formula:

$$\text{Contrast} = \sqrt{\frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \left(I(i,j) - \mu\right)^2} \tag{3.2}$$

   where:

   - μ represents the mean grayscale intensity of the image:

$$\mu = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} I(i,j)$$

The contrast of the image must satisfy the following condition:

$$\text{Contrast} \geq 40$$

3. **Image Resizing:** All images that meet the brightness and contrast criteria are resized to a fixed dimension:

$$\text{IMAGE\_SIZE} = (512, 512)$$

Based on the image evaluation metrics, we retained only those images whose brightness and contrast simultaneously met the above thresholds.

3.3 Synthetic Minority Over-sampling Technique (SMOTE)

From the table, we can see that there is data imbalance in the data. Therefore, we invoke the SMOTE algorithm. SMOTE is a synthetic sampling technique that starts with minority class samples, finds neighboring samples, and synthesizes new minority class samples so that the number of minority class samples is consistent with the number of majority class samples.

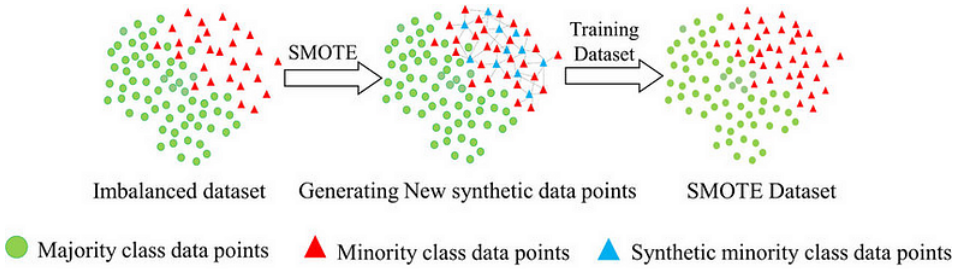


Figure 3.3: Overview of SMOTE

Prior to the introduction of SMOTE synthetic sampling technique, oversampling techniques basically increased the number of samples by replicating the samples (e.g., random oversampling technique). However, simple sample replication only increases the number of samples, but does not improve the quality of samples, the data is still unbalanced, so the classifier can only learn the same features over and over again, and the improvement of the grading performance is very limited. SMOTE, on the other hand, through the method of synthesizing new samples, the algorithm can learn from more new samples to be more conducive to the grading of the minority class, so SMOTE was very popular when introduced. is very hot, and so far became the classic algorithm for oversampling.

3.3.1 Principle of SMOTE

SMOTE achieves sample synthesis by finding neighboring samples using k-nearest neighbors for all minority class samples, followed by linear random interpolation. Among them, the position of interpolation is random, and the number of interpolated values for each sample point is equal (the excess is randomly deleted). The specific interpolation process is shown below:

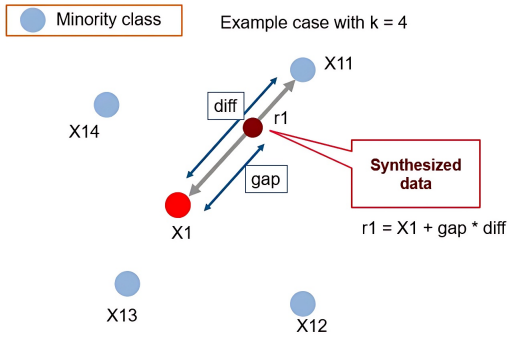


Figure 3.4: Process of SMOTE

The figure sets $k = 4$ in k-nearest neighbor, X_1 is the minority class sample and finds X_{11}, X_{12}, X_{13}, X_{14}, which are the four nearest neighbor sample

points.

$$r_1 = X_1 + gap \cdot d \tag{3.3}$$

In the interpolation between X_1 and X_{11}, d is the distance between the two sample points, and the newly generated sample point r_1 is in the connecting straight line, and gap is the random distance between X_1 and X_{11}.

3.3.2 Limitations of SMOTE

In medical image grading, SMOTE (Synthetic Minority Over-sampling Technique) is widely used to address data imbalance by generating synthetic samples. However, due to the complexity and variability of medical images, SMOTE exhibits several shortcomings that can limit its effectiveness, especially in handling the nuanced features present in medical datasets.

- **Boundary Sensitivity:** SMOTE does not consider the spatial or feature boundaries between classes, leading to the potential generation of synthetic samples in ambiguous regions, which can confuse the classifier.
- **Vulnerability to Noise:** SMOTE amplifies the influence of noisy minority samples by treating all minority samples equally, which can degrade model performance.
- **Inapplicability to Feature-Specific Patterns:** For medical images, specific patterns (e.g., lesions or anatomical structures) are critical. SMOTE may generate synthetic samples lacking these essential characteristics, reducing the clinical relevance of the data.

3.4 Borderline-SMOTE

In response to the shortcomings of the SMOTE algorithm, we introduced the Borderline-SMOTE improvement algorithm. The algorithm divides the few

sample points into two categories: safe points, dangerous points, and only interpolates the dangerous points, because we generally believe that the dangerous points play a greater role in grading, and highlighting the dangerous points is more conducive for the model to capture the important features of the category that are different from the other categories, and the model focuses on such samples to efficiently improve the grading performance. In this paper, we will also use the borderline-smote algorithm to upsample a few category samples.

3.4.1 Procedure of Borderline-SMOTE

Let the training set be T, where P and N represent the minority and majority classes, respectively:

$$P = \{p_1, p_2, \ldots, p_{pnum}\}, \quad N = \{n_1, n_2, \ldots, n_{nnum}\}. \tag{3.4}$$

Here:

- $pnum$ and $nnum$ are the numbers of samples in the minority and majority classes, respectively.
- p_i and n_j are feature vectors in the d-dimensional space: $\mathbb{R}^d$.

We define the **Danger Set** (D) as:

$$D = \{p_i \in P \mid \frac{k}{2} \leq m'(p_i) < k\}, \tag{3.5}$$

where $m'(p_i)$ represents the number of majority class neighbors of p_i among its k-nearest neighbors. Samples in D are borderline minority samples that are more likely to be misclassified.

Procedure:

1. **Identify Borderline and Danger Samples:** For each $p_i \in P$, find its

k-nearest neighbors from the entire training set T. Let:

$$m'(p_i) = \text{count(neighbors from } N), \quad m'(p_i) \leq k. \tag{3.6}$$

Classify p_i based on $m'(p_i)$ as follows:

- *Safe Sample:* If $m'(p_i) < \frac{k}{2}$, p_i is in a predominantly minority region.
- *Borderline (Danger) Sample:* If $\frac{k}{2} \leq m'(p_i) < k$, p_i belongs to the Danger Set D.
- *Noisy Sample:* If $m'(p_i) = k$, p_i is surrounded entirely by majority samples and is excluded.

2. **Generate Synthetic Samples for the Danger Set:** For each $p_i \in D$, select a subset of its k-nearest neighbors from the minority class P:

$$\{p_{i_1}, p_{i_2}, \ldots, p_{i_k}\}. \tag{3.7}$$

Randomly choose s neighbors ($1 \leq s \leq k$), and generate synthetic samples as:

$$\text{synthetic}_j = p_i + r_j \cdot (p_{\text{neighbor}} - p_i), \tag{3.8}$$

where $r_j \in [0, 1]$ is a random scalar, and neighbor $\in \{p_{i_1}, p_{i_2}, \ldots, p_{i_k}\}$.

3. **Balancing the Dataset:** Repeat the synthetic sample generation process for all samples in D. Ensure that the total number of minority samples after augmentation satisfies:

$$\text{new minority samples} = nnum - pnum. \tag{3.9}$$

If the number of generated samples exceeds the required count, a random selection of the required number is made.

3.4.2 Simulation of Borderline-SMOTE

The following simulated data set, circle, which contains two classes, makes it simple to understand our methods. The data set's initial distribution is depicted in (a), where majority examples are represented by circle points and minority examples by plus signs. First, we use borderline-SMOTE to identify the minority class's borderline cases, which are shown in (b) as solid squares. These liminal instances of the minority class are then used to create new synthetic examples. The hollow squares in (c) represent the synthetic cases. The statistics make it clear that, in contrast to SMOTE, our approaches merely over-sample or reinforce the minority class's borderline and surrounding points.

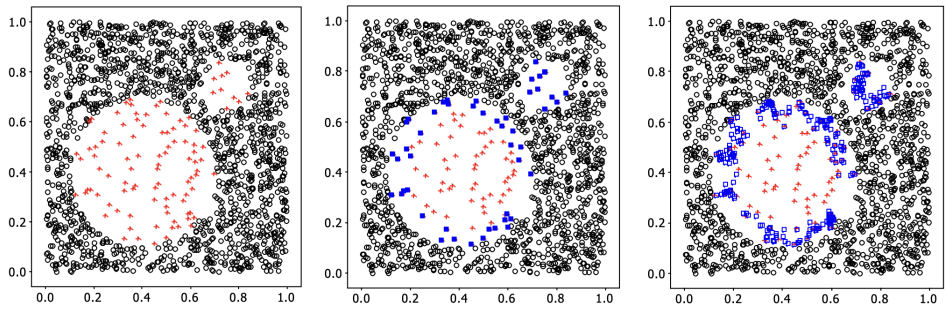


Figure 3.5: Simulation of Borderline-SMOTE

The minority borderline samples $p'_i \in$ DANGER are strengthened by generating synthetic data along the line between p'_i and its nearest minority neighbors:

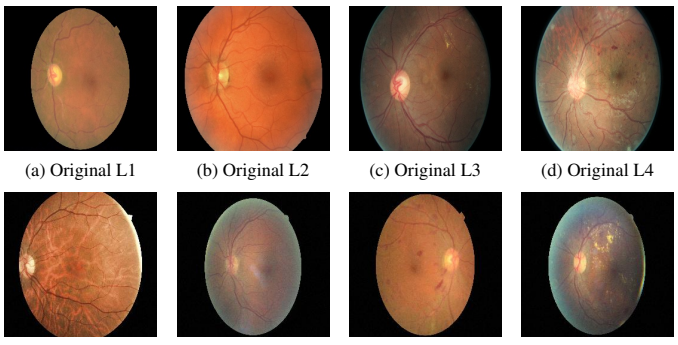
$$\text{synthetic}_j = p'_i + r_j \cdot (p_{\text{neighbor}} - p'_i). \tag{3.10}$$

This approach ensures that the new samples remain closer to the decision boundary, enhancing the classifier's ability to distinguish between classes.

3.4.3 Synthetic Fundus Image by Borderline-SMOTE

The following key parameters were selected when applying the Borderline-SMOTE technique:

- **Sampling Strategy (`sampling_strategy`):** For minority classes (1, 2, 3, 4), the target sample count was set to 3500, ensuring that the sample sizes of all classes were approximately balanced.
- **Neighbor Parameters (`k_neighbors` and `m_neighbors`):**
  - $k = 16$: Used to determine the k-nearest neighbors for each minority class sample, capturing the local boundary characteristics of the samples.
  - $m = 10$: Used to define the neighborhood range of majority class samples, ensuring that the synthetic samples are concentrated near the actual decision boundary.
- **Boundary Type (`kind=borderline-2`):** The Borderline-2 mode was chosen to focus on generating high-quality samples at the decision boundary, rather than extending to the entire minority class distribution.



(a) Original L1 (b) Original L2 (c) Original L3 (d) Original L4

(e) Synthetic L1 (f) Synthetic L2 (g) Synthetic L3 (h) Synthetic L4

Figure 3.6: Comparison between Original and Synthetic Images

Using the above parameters, Borderline-SMOTE generates new minority class samples while avoiding the redundancy of traditional SMOTE, which tends to generate samples in non-boundary regions. We can see that the fundus image synthesized using the borderline-smote technique retains the semantic information of the image well for each DR level, providing a complete dataset for the following model training.

Chapter 4

VGG16 Applied in DR Grading

4.1 Classical VGG16 Architecture

Since it was proposed by the Visual Geometry Group (VGG16) of Oxford University in 2014, the VGG16 model has become one of the classic architectures in the field of deep learning due to its concise and deep structure that performs well in image recognition tasks. In this section, we will introduce in detail the overall architectural design of VGG16, the specific configurations of each layer, and its application advantages in transfer learning. First, we will parse the layer structure of VGG16, including the specific settings of convolutional, pooling, and fully connected layers; subsequently, we will explore the impact of VGG16 on the performance of the model in terms of the number of parameters, computational complexity, and its design philosophy.

4.1.1 Layered Structure of VGG16

The VGG16 model consists of 16 network layers with learnable parameters, including 13 convolutional layers and 3 fully connected layers. Its core feature is the use of a series of consecutive 3×3 convolutional kernels, together with a 2×2 maximum pooling layer, to achieve layer-by-layer extraction and

abstraction of image features.

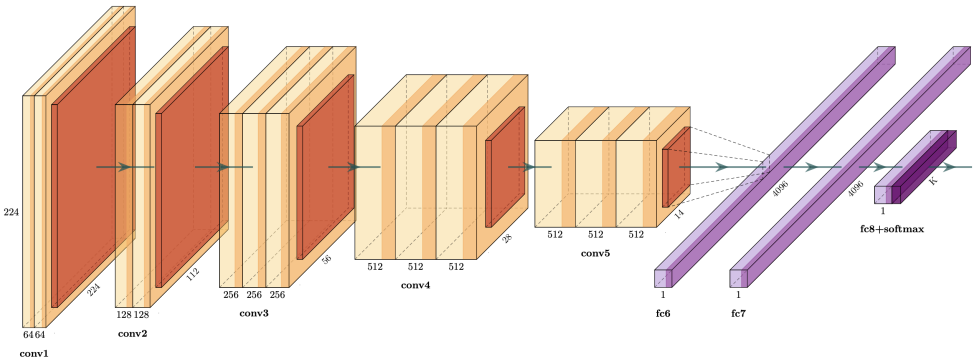


Figure 4.1: Classical VGG16 Structure

First Two Convolutional Blocks The input image will be resized to $224 \times 224 \times 3$ RGB images. The image will passe through two consecutive 3×3 convolutional layers, followed by a 2×2 max pooling layer with stride of 2.

Middle Three Convolutional Blocks Each of these blocks contains three 3×3 convolutional layers. The number of channels progressively increases to 256 and 512, respectively. After each convolutional block, a max pooling operation with a 2×2 window and stride of 2 is applied, halving the spatial dimensions each time. This results in feature maps of size $7 \times 7 \times 512$ after the final convolutional block.

Pooling Layers Following each group of convolutional layers, a 2×2 max pooling layer with a stride of 2 is employed for downsampling. This process reduces the spatial dimensions while preserving the most salient features extracted by the convolutional layers.

Fully Connected Layers After convolutional and pooling operations, VGG16 includes three fully connected layers. The first two fully connected layers each contain 4096 neurons, and the final layer has 1000 neurons corresponding to the 1000 classes in the ImageNet dataset.

Activation Functions All convolutional and fully connected layers are followed by the Rectified Linear Unit (ReLU) activation function. The use of ReLU introduces non-linearity into the model, enhancing its ability to learn complex patterns and improve overall expressiveness.

4.1.2 Design Philosophy and Parameter Optimization

The design of VGG16 demonstrates the potential of deep networks in image recognition tasks. Its core philosophy focuses on enhancing the model's representational power and generalization ability by increasing network depth and using small-sized convolution kernels.

- **Small-sized Convolution Kernels:** VGG16 employs 3×3 small convolution kernels, which, compared to larger kernels (e.g., 5×5 or 7×7), reduce the number of parameters. By stacking multiple small kernels, it effectively simulates a larger receptive field, capturing features at different scales.

- **Deep Network Architecture:** The network significantly increases its feature extraction capability by using a deep structure with 16 layers, enabling it to identify complex image patterns. However, the deeper structure introduces higher computational demands and potential vanishing gradient issues. By incorporating ReLU activation functions and efficient initialization strategies, VGG16 successfully mitigates these challenges.

- **Number of Parameters and Computational Complexity:** VGG16 contains approximately 138 million parameters, primarily concentrated in the fully connected layers. While this poses challenges in terms of storage and computation, techniques such as hardware acceleration (e.g., GPU parallelization) and optimization methods (e.g., batch normalization) effectively enhance the efficiency of training and inference.

- **Advantages in Transfer Learning:** Pretrained on large-scale datasets like ImageNet, VGG16 learns rich and generalizable features. These features demonstrate strong transferability across different image recognition tasks. By freezing the initial convolutional layers and fine-tuning only the later layers or fully connected layers, VGG16 achieves efficient and accurate grading on smaller datasets, significantly reducing training time and improving generalization performance.

In summary, the deep and refined architectural design of VGG16 not only achieves outstanding performance in image grading tasks but also provides a robust foundation for transfer learning, making it a preferred model for numerous computer vision applications.

4.2 VGG16 Architecture for DR grading

In the DR grading task, there are a large number of subtle local features (e.g., hard and soft exudates) in the fundus image, which puts a higher demand on the accuracy and interpretability of the model. In addition, medical image datasets are usually small and unbalanced, so considering the need to efficiently improve the performance of the model's performance, I will optimize the model using transfer learning based and modified VGG16 architecture.

4.2.1 Transfer Learning Strategy

Transfer learning is the reuse of a pre-trained model on a new problem Because it can train deep neural networks with relatively minimal data, it is well-liked in the field of deep learning. Since most real-world situations lack the millions of labelled data points needed to train such sophisticated models, this is extremely helpful in the data science profession.

There are numerous benefits of using transfer learning. It provides a notable reduction in computing time, first and foremost. Transfer learning enables us to leverage the knowledge gained from earlier training procedures rather of creating a whole new model from scratch. Furthermore, it enhances the learning process by expanding the range of data gleaned from earlier models.Finally, when working with a tiny new training dataset, transfer learning is quite helpful.Transfer learning lessens the negative impact of scarce data on model performance by applying the thorough feature representation discovered from bigger, previously trained datasets (11).

4.2.2 Fine-Tuning

Fine-tuning is a strategic approach in training a CNN, which involves using a pre-existing set of weights alongside new data. Essentially, the weights from the pre-trained CNN model are harnessed to initialize a target CNN model with an identical architecture. Subsequently, the target CNN is supervised and trained on the new target data (1).

There are two ways to go about fine-tuning. The first method, known as comprehensive fine-tuning, calls for adjusting each network layer in the CNN model. When there is a significant lack of correlation between the source and target domains, this approach works especially well. To maintain proper model performance in these situations, it becomes imperative to fine-tune every layer. The second strategy entails layer-by-layer fine-tuning of the CNN model that has already been trained. This approach saves computing power and may enhance model generalisation by providing the freedom to modify particular layers that are more pertinent to the novel task.

For our study, we've chosen comprehensive fine-tuning approach. This decision stems from the desire to harness the pre-trained model's knowledge. There are very subtle differences between different fundus images, and insufficient training parameters may result in the model failing to accurately learn

important features, such as exudates and tiny hemorrhages.

4.2.3 VGG16 Based on Transfer Learning

Due to the presence of a large number of subtle local features (e.g., hard and soft exudates) in fundus images, this places higher demands on the accuracy and interpretability of the model. In addition, medical image datasets are usually small and unbalanced, so considering the need to efficiently improve the performance of the model's performance, I will optimize the model using transfer learning based and modified VGG16 architecture.

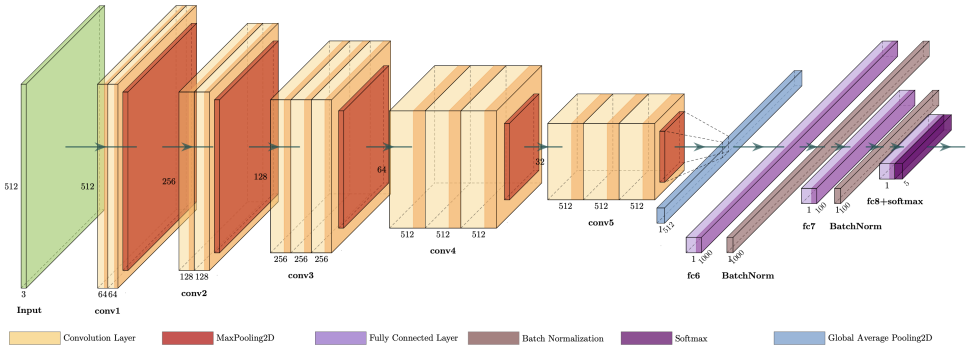


Figure 4.2: VGG16 Structure for DR grading

In adapting the VGG16 network architecture for the diabetic retinopathy (DR) grading task, I considered the characteristics of medical images, task requirements, and potential challenges to make the following targeted optimization adjustments:

1. **Resize Input to** 512×512**:** Medical images often contain critical localized lesion features (e.g., microvascular tumors and exudates) that are small and unevenly distributed, and the standard 224×224 input size may not be able to retain this detailed information completely. By increasing the input size, more spatial information can be captured and the model's ability to perceive lesion features can be improved. This tuning direction

stems from an in-depth analysis of the fundus image feature distribution and task requirements to ensure that the model is able to perform more efficient feature extraction based on the increased resolution.

2. **Batch Normalization Layer:** Medical image data usually have small sample sizes and large differences in distribution, which can easily lead to gradient instability and overfitting problems during model training. Batch normalization not only mitigates the gradient vanishing problem by standardizing the distribution of each batch of data, but also accelerates the convergence speed of the model and improves the robustness of the network. Aiming at the problem of data scarcity and inconsistent collection conditions in the DR hierarchical task, this adjustment helps the stability and generalization ability of the model on small-scale datasets.

3. **Global Average Pooling:** The fully connected layer occupies a large number of parameters in standard VGG16, which can easily lead to overfitting for small sample datasets common in medical tasks. Global average pooling significantly reduces the number of parameters by spatially globalizing the feature map, while enhancing the interpretability of the model so that its output can be more intuitively associated with specific lesion regions.

After designing the modified structure of VGG16, the following process is to carry out transfer learning strategy to fine tune parameters by using the preprocessed fundus image dataset. The exact experimental setup and details will be explained in next chapter.

Chapter 5

Experiments Outcome

5.1 Experimental Setup

The Table 5.1 shows the platform for experimental training and its resource allocation. Due to the limited computational resources, after several attempts with different sample sizes, and considering the model performance and training efficiency, the up-sampling is performed through the borderline-smote technique, and the final number of samples for each DR level is selected to be 3500.

Table 5.1: Experiment Platform and Environment Configuration

No.	Component	Configuration
1	Operating System	Ubuntu 20.04
2	GPU	NVIDIA GeForce RTX 4090 (24GB)
3	CPU	Intel Xeon Platinum 8352V
4	Deep Learning Framework	Tensorflow 2.13.1
5	Python Version	Python 3.8.20

The hyperparameters of the VGG16 model are presented in the Table 5.2.

Using transfer learning and a carefully chosen set of hyperparameters, a modified VGG16 model was created and trained for diabetic retinopathy (DR) grading in this study. These hyperparameters were adjusted to ensure the model's applicability in environments with limited resources by striking a balance between grading performance and computational efficiency.

Table 5.2: Hyperparameters Configuration for VGG16-based Model

Hyperparameter	**Value**
Input Image Size	512×512
Batch Size	32
Learning Rate	1×10^{-4}
Optimizer	SGD (momentum=0.9)
Loss Function	Categorical Crossentropy
Number of Epochs	80
Early Stopping Patience	15 epochs
Learning Rate Scheduler	ReduceLROnPlateau (factor=0.5, patience=5, min_lr=1×10^{-7})
Pretrained Weights	ImageNet
Base Model Layers Trainable	All Layers

In addtion to the hyperparameters from Table 5.2, numerous data augmentation techniques, such as rotation, width/height shifting, zoom, brightness adjustment, and horizontal flipping were used because of the training dataset's small size and variability. Moreover, to avoid overfitting and reduce unnecessary computation, early stopping was employed with a patience of 15 epochs. This ensured that training ceased once the validation loss stopped improving for a significant duration. Learning Rate and Scheduler is also defined. The initial learning rate was set to 1×10^{-4}, ensuring gradual and stable updates to the model weights. To dynamically adjust the learning rate, a ReduceLROnPlateau scheduler was implemented, reducing the learning rate by a factor

of 0.5 if the validation loss plateaued for 5 epochs. This prevents premature convergence and optimizes learning in later training stages.

5.2 Experimental Results and Analytics

At the beginning of the training phase, I randomly shuffled 17,500 image samples, of which 14,000 images were used to train the model and 3,500 images were used for model validation, due to the limited computational resources, ten-fold cross-validation was not done here, but due to the more than sufficient amount of data, the final output of the model is also of strong interpretive and practical significance

5.2.1 Training Dynamics and Model Generalization

The training and validation results provide crucial insights into the performance of the VGG16 model for diabetic retinopathy (DR) grading. In Figure 5.1, the loss curve reveals a steady decline in training loss, which stabilizes after 60 epochs, indicating that the model has effectively converged. The validation loss, however, exhibits significant fluctuations during the early epochs, particularly between 10 and 30. This pronounced oscillation can be attributed to the initially high learning rate, which causes the model to take large steps in the optimization process.

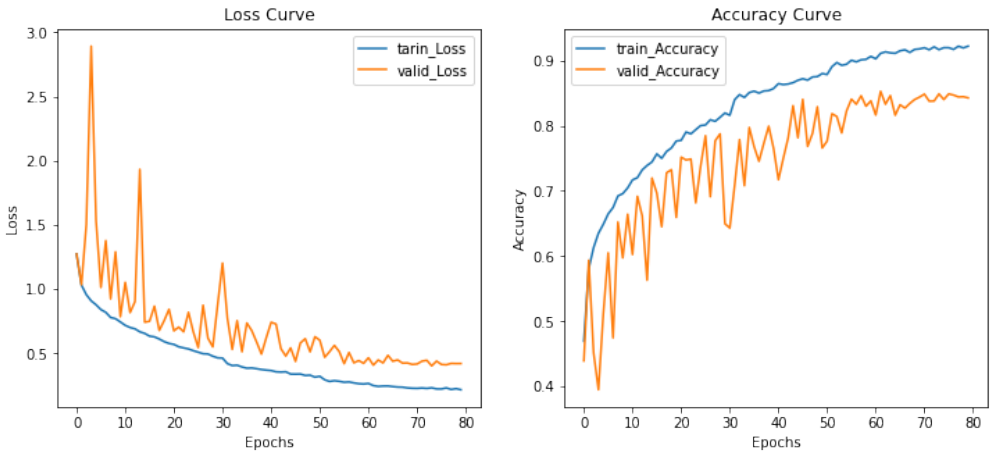


Figure 5.1: Loss and Accuracy for Each Epoch

As training progresses, the learning rate is gradually reduced due to the implemented reduced learning rate scheduler, allowing the model to stabilize and achieve a more consistent decline in validation loss. This mechanism proves effective in enabling the model to better generalize the complex and subtle features of DR, particularly those distinguishing overlapping severity levels.

The accuracy curves further highlight the model's training dynamics. The training accuracy improves consistently, surpassing 90% by the final epochs. Validation accuracy, while lower at 85.26%, follows a similar trend with smaller fluctuations as the learning rate decreases. This stabilization reflects the model's ability to adapt to unseen data, even when faced with the challenges of inter-class feature overlap, such as distinguishing between Moderate and Severe DR.

Table 5.3 summarizes the final training and validation metrics, emphasizing the system's effectiveness. The training accuracy of 91.07% and validation accuracy of 85.26% confirm that the model captures key patterns in the training set while maintaining acceptable generalization.

Table 5.3: grading Accuracy of the System

	Training	Validation
Accuracy	91.07%	85.26%
Loss	0.2496	0.4087

The associated loss values, 0.2496 for training and 0.4087 for validation, further indicate successful optimization, with the remaining performance gap primarily reflecting the inherent difficulty of distinguishing subtle retinal features in specific DR stages. For instance, while Proliferative DR and Severe DR often exhibit pronounced abnormalities like neovascularization, differentiating between Mild and Moderate DR requires identifying more subtle textural changes, which can challenge even well-optimized models.

5.2.2 Performance Across DR Severity Levels

In order to further explore the correlation between the model grading situation and the real-world meaning, I also introduced the confusion matrix and some grading model performance assessment metrics.

As can be seen from Table 5.4, there are some differences in the model's grading performance metrics (Precision, Recall and F1-score) on different DR classes. Among them, Proliferative DR (proliferative DR) has the best performance, with Precision and Recall close to 1.0 and F1-score reaching 0.99. This indicates that the model has almost no error in classifying this grade, probably because the lesion features of proliferative DR (e.g., neovascularization and vitreous hemorrhage) are distinctive and easy to be recognized. For Mild NPDR (mild nonproliferative DR) and Moderate NPDR (moderate nonproliferative DR), on the other hand, the F1-score was 0.77 and 0.73, respectively, which is a relatively low performance. This is mainly attributed to the fact that the differences in features between these two classes are more subtle, e.g., areas of

microangiomas and mild exudates are more difficult to differentiate, adding to the complexity of grading.

Table 5.4: grading performance of each class in the system

Class	Precision	Recall	F1-score
Normal (No DR)	0.73	0.76	0.74
Mild NPDR	0.73	0.81	0.77
Moderate NPDR	0.82	0.67	0.73
Severe NPDR	0.96	0.98	0.97
Proliferative DR	0.99	0.99	0.99

The confusion matrix in Figure 5.2 further reveals the specific grading behavior of the model. For the Normal (no DR) class, while the majority of samples were correctly classified (532), 107 samples were still misclassified as Mild NPDR, suggesting that there is some confusion in the model's identification of early lesions.

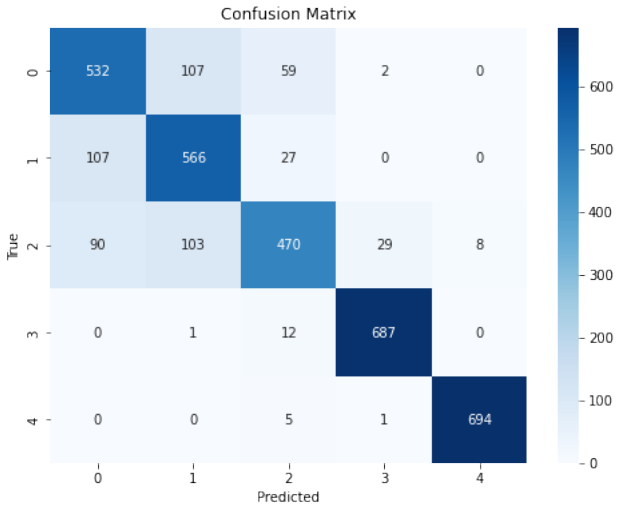


Figure 5.2: Confusion Matrix of the System

In addition, the grading error of Moderate NPDR was more significant, with 90 samples misclassified as Normal and 103 samples misclassified as Mild NPDR.This error may be due to the fact that the lesion features of Moderate NPDR overlapped with some of the features of Normal and Mild NPDR at the same time, e.g., some moderate exudates or hemorrhagic spots were not accurately captured. This error may be due to the fact that the lesion characteristics of Moderate NPDR also partially overlap with those of Normal and Mild NPDR, such as some moderate exudates or hemorrhages that are not accurately captured. In contrast, Severe NPDR (severe non-proliferative DR) and Proliferative DR had very few misgradings and showed high grading reliability, consistent with their high Precision and Recall in Table 5.4.

Chapter 6

Research Discussion

6.1 Conclusions

This study proposes an improved VGG16 architecture for diabetic retinopathy (DR) grading based on transfer learning. By deploying a pre-trained VGG16 model on ImageNet, followed by fine-tuning for the specific task of diabetic retinal image grading, including adjusting the convolutional layer size accordingly to the fundamentals of VGG16 after resizing the inputs to 512×512, introducing a batch normalization layer to stabilize the training process and replacing the fully-connected with a global average pooling layer to reduce the model complexity. layer to reduce the model complexity. In terms of data preprocessing, we first screened the images that might have an impact on the model training by contrast and brightness, and then utilized Borderline-SMOTE, a modified algorithm of SMOTE, to synthesize a small number of samples in order to solve the data imbalance problem. In order to preserve the spatial structure of the fundus image as much as possible during the sample synthesis process, under the condition of limited computational cost, we chose to extract the features in the images, including optic disc, macula, exudates, microvessels, and so on, by pre-training the convolutional layer of the VGG16 model. Afterwards, these features are utilized for sample synthesis,

and the final synthesized image achieves excellent results and has certain reference value.

6.2 Contributions

The results once again demonstrate the power of transfer learning in medical image tasks, showing the good adaptability of the pre-trained model in scenarios with limited labeled data. On this basis, considering resource-constrained scenarios such as clinical environments or edge computing, we effectively improve the robustness of the model and reduce the risk of overfitting on small medical datasets by adapting the VGG16 architecture (e.g., introducing global average pooling and batch normalization). My proposed method efficiently identifies and classifies micro-pathological features in fundus images with limited computational resources, providing a possible solution for DR screening automation.

6.3 Limitations

Despite the promising results, this study has several limitations:

1. **Dataset Size and Diversity:** Although the dataset used for training and validation was effective, it may not fully represent the diversity of real-world retinal images. Variations in image quality, acquisition conditions, and demographic factors could affect the model's generalizability.

2. **Explainability of Predictions:** While the model achieved high accuracy, its decision-making process remains challenging to interpret. Future work should incorporate advanced attention mechanisms or visualization techniques to enhance clinical explainability.

3. **Computational Constraints:** Although the model performs well under limited computational resources, further reductions in complexity could make it more accessible for deployment in low-resource environments.

4. **Comparison with State-of-the-Art Models:** This study did not extensively benchmark the proposed model against other advanced architectures, such as EfficientNet or Vision Transformers, which may provide different trade-offs between performance and computational cost.

6.4 Future Work

Based on the findings and limitations of this study, several directions for future research are proposed. Firstly, expanding the dataset with more diverse and heterogeneous retinal images is essential to improve the model's robustness across various imaging conditions and patient demographics. This will help enhance the model's generalization ability and applicability in real-world settings. Secondly, integrating explainable artificial intelligence (XAI) techniques, such as Grad-CAM or attention mechanisms, will enable better visualization of the regions contributing to the model's predictions, thereby increasing its trustworthiness in clinical applications. Additionally, exploring model compression and optimization techniques, such as pruning, quantization, or adopting lightweight architectures like MobileNet, could further reduce computational requirements while maintaining performance. Moreover, future research should focus on benchmarking the modified VGG16 against other state-of-the-art models, such as EfficientNet or Vision Transformers, to provide a comprehensive comparison of performance and efficiency. Finally, validating the model in real clinical environments is critical to assess its usability, scalability, and integration potential within existing healthcare workflows. These directions will pave the way for more effective and practical applications of automated DR grading systems in medical practice.

Bibliography

[1] H. Mustafa, M. Chrayah, N. Ourdani, and N. Aknin, "Rapid detection of diabetic retinopathy in retinal images: A new approach using transfer learning and synthetic minority oversampling technique," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 1, pp. 1091–1101, Feb. 2024. doi: 10.11591/ijece.v14i1.

[2] Y. M. Yan et al., "Identifying diabetic retinopathy based on deep transfer learning" *OPTICAL INSTRUMENTS*, vol. 42, no. 5, pp. 33–42, Oct. 2020. doi: 10.3969/j.issn.1005-5630.2020.05.006.

[3] X. L. Yi and M. X. Yu, "Pathogenesis of diabetic retinopathy," *Fudan University Journal of Medical Sciences*, vol. 37, no. 5, pp. 604-607, Sep. 2010.

[4] J. L. Wan, J. B. Hu, W. D. Jin, and P. Tang, "Screening and grading of fundus images of diabetic retinopathy based on visual attention," *Chinese Journal of Experimental Ophthalmology*, vol. 37, no. 8, pp. 630-637, Aug. 2019.

[5] L. Dai et al., "A deep learning system for detecting diabetic retinopathy across the disease spectrum," *Nature Communications*, vol. 12, no. 3242, pp. 1–12, May. 2021. doi: 10.1038/s41467-021-23458-5.

[6] Y. W. Ju, "Automated Diabetic Retinopathy Grading and Segmentation Based on Semi-Supervised Learning," *Modeling and Simulation*, vol. 13, no. 3, pp. 3828–3841, May. 2024. doi: 10.12677/mos.2024.133349.

[7] Z. W. Yang, T. Tan, Y. Shao, T. Y. Wong, and X. Li, ”Classification of diabetic retinopathy: Past, present and future,” *Frontiers in Endocrinology*, vol. 13, Art. no. 1079217, pp. 1-18, Dec. 2022. doi: 10.3389/fendo.2022.1079217.

[8] D. Doshi, A. Shenoy, D. Sidhpura and P. Gharpure, ”Diabetic retinopathy detection using deep convolutional neural networks,” *2016 International Conference on Computing, Analytics and Security Trends (CAST)*, Pune, India, 2016, pp. 261-266, doi: 10.1109/CAST.2016.7914977.

[9] D. U. N. Qomariah, H. Tjandrasa and C. Fatichah, ”Classification of Diabetic Retinopathy and Normal Retinal Images using CNN and SVM,” *2019 12th International Conference on Information & Communication Technology and System (ICTS)*, Surabaya, Indonesia, 2019, pp. 152-157, doi: 10.1109/ICTS.2019.8850940.

[10] R. Patel and A. Chaware, ”Transfer Learning with Fine-Tuned MobileNetV2 for Diabetic Retinopathy,” *2020 International Conference for Emerging Technology (INCET)*, Belgaum, India, 2020, pp. 1-4, doi: 10.1109/INCET49848.2020.9154014.

[11] F. Zhuang et al., ”A Comprehensive Survey on Transfer Learning,” in *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43-76, Jan. 2021, doi: 10.1109/JPROC.2020.3004555.