PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
Ministry of Higher Education and Scientific Research

UNIVERSITY OF ECHAHID HAMMA LAKHDAR - EL OUED

FACULTY OF EXACT SCIENCES
Computer Science Department

Thesis
submitted in partial fulfilment of the requirements for the degree of

ACADEMIC MASTER

Field: **Mathematics and Computer Science**
Option: **Computer Science**
Specialty: **Artificial Intelligence & Data Science**

Submitted by :

- **ABID Abdelbadie**
- **BEYAT Ahmed Taha**

Title

Building Robust Pretrained Deep Learning Models for Diabetic Retinopathy (DR) Classification on Local Datasets

Defended on May 12 th 2025

Examination Commitee :

M.	**GUIA Sana Sahar**	Professor	Chair
M.	**ZAOUI Sayeh**	MCA	Examinator
M.	**GHENABZIA Ahmed**	MAA	Supervisor
M.	**KHELAIFA Abdennacer**	MAA	Co-Supervisor

Academic year: 2024-2025

Acknowledgments

We would like to express our sincere gratitude to **Prof. Ahmed Ghanbazieh** for his kind supervision, valuable guidance, and constructive feedback, which greatly contributed to the successful completion of this work.

We also extend our heartfelt appreciation to **Dr. Khelaifa Abdennacer** and **Dr. Abid Mohamed Nadhir** for their insightful comments, encouragement, and valuable suggestions that helped refine and strengthen our research.

In addition, we express our deepest thanks to our **parents** for their continuous moral support and encouragement, which have been a great source of strength throughout our academic journey.

Furthermore, we are grateful to our family, friends, and colleagues who stood by us and offered their help and motivation, enabling us to overcome the challenges faced during this project.

Abstract

Diabetic Retinopathy is a top cause of bad eye sight all over the world, showing how key it is to find it early and right. Deep models, mainly CNN, show great promise in making DR checks automatic. Yet, their success rests on having big and mixed sets of data, which might not be easy to have in local health spots.

This work looks into making strong deep learning models for sorting DR using data from local places. We look at many CNN set-ups, use new ways to grow data, and use ways to adapt to different areas to make the model work better. We also use methods to make things clear, helping to trust AI more in eye checks.

Our results show the hard parts and good points of using local data. They make clear how transfer learning and deep data work help cut down biases and boost how well models work. The plan we suggest tries to close the space between AI-based disease finding and real-world doctor use. It makes sure it works well for many different types of patients.

Keywords: Diabetic Retinopathy, Deep Learning, Convolutional Neural Networks, Transfer Learning, Explainability, Domain Adaptation, Medical Image Analysis.

Résumé

La rétinopathie diabétique (RD) est l'une des principales causes de déficience visuelle dans le monde, nécessitant un diagnostic précoce et précis pour une intervention efficace. Les modèles d'apprentissage profond ont démontré un fort potentiel dans l'automatisation de la détection de la RD, mais leur robustesse dépend largement de la disponibilité de vastes ensembles de données diversifiés. Cette étude vise à développer des modèles d'apprentissage profond robustes pour la classification de la RD en utilisant des ensembles de données locaux. Nous explorons différentes architectures de réseaux de neurones convolutifs (CNN), des techniques d'augmentation des données et des stratégies d'adaptation au domaine pour améliorer la généralisation du modèle. De plus, nous employons des méthodes d'explicabilité pour garantir l'interprétabilité et la fiabilité clinique des résultats. Nos analyses mettent en évidence les défis et les avantages liés à l'utilisation de données locales, en soulignant le rôle du transfert d'apprentissage et des techniques avancées de prétraitement dans la réduction des biais et l'amélioration des performances du modèle. L'approche proposée vise à combler l'écart entre la détection de la RD basée sur l'IA et son application clinique, garantissant ainsi une fiabilité accrue sur des populations diverses.

Mots-clés : Rétinopathie diabétique, Apprentissage profond, Réseaux de neurones convolutifs (CNN), Diagnostic précoce, Données locales, Explicabilité de l'IA

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Definition
AI	Artificial Intelligence
CNN	Convolutional Neural Network
DR	Diabetic Retinopathy
ML	Machine Learning
RNN	Recurrent Neural Network
TL	Transfer Learning
PM	Pretrained model
APTOS	Asia Pacific Tele-Ophthalmology Society
ResNet	Residual Network
ReLU	Rectified Linear Unit
HF	Hugging Face
GAN	Generative Adversarial Networks
HPC	High-Performance Computing

General Introduction

Introduction

Diabetic Retinopathy (DR) is a leading cause of vision impairment and blindness among diabetic patients worldwide. It results from prolonged exposure to high blood sugar levels, which damage the blood vessels in the retina, leading to progressive vision loss if left untreated. Early detection and timely intervention are critical in preventing severe complications. Traditionally, DR diagnosis relies on manual examination of retinal fundus images by ophthalmologists, a process that is time-consuming, prone to human error, and limited by the availability of specialists [20].

With advancements in AI and deep learning, automated DR detection systems have emerged as a promising solution to enhance diagnostic accuracy and accessibility. Convolutional Neural Networks (CNNs) have demonstrated exceptional performance in image classification tasks, including medical image analysis [21]. However, the robustness and generalizability of these models largely depend on the quality and diversity of training datasets. Many existing deep learning models are trained on large-scale, publicly available datasets that may not adequately represent local populations and healthcare conditions, limiting their applicability in real-world clinical settings [22].

Problem Statement

Despite the success of DL in DR detection, significant challenges remain in developing robust models that perform reliably across different demographic groups and medical imaging conditions. Models trained on global datasets often suffer from domain shift

when applied to local datasets, leading to reduced accuracy and increased false positives or negatives. Additionally, variations in imaging equipment, illumination conditions, and labeling inconsistencies can impact model performance [23].

This study focuses on developing robust DL models for DR detection using local datasets. By leveraging domain adaptation techniques, advanced data augmentation strategies, and explainable AI methods, the research aims to enhance model reliability and interpretability for clinical deployment.

Hypothesis

This research is based on the hypothesis that DL models trained and fine-tuned on locally sourced datasets will exhibit improved accuracy and generalizability compared to models trained on generic, publicly available datasets. Additionally, implementing domain adaptation and explainability techniques will enhance the interpretability and trustworthiness of the models in clinical settings [24].

Thesis Structure

This thesis is structured as follows:

- **Chapter Two: Background** provides an overview of existing research on DL models for DR detection, highlighting key advancements, limitations, and gaps.
- **Chapter Three: Pretrained Model for DR Detection** outlines the data collection process, model architectures, training procedures, and evaluation metrics used in this study.
- **Chapter Four: Implementation and Results** presents the experimental results, model comparisons, and insights gained from the study.
- **Chapter Five: Conclusion and Future Work** summarizes the key findings, discusses the implications, and suggests directions for future research.

This introduction sets the foundation for understanding the research problem, objectives, and significance, guiding the subsequent chapters in addressing the challenges of developing robust DL models for diabetic retinopathy detection on local datasets.

Chapter 1:

Background

1.1. Introduction

The background chapter outlines fundamental concepts central to the research, such as healthcare, diabetes, and diabetic retinopathy. It also examines diagnostic imaging techniques for DR, discusses the application of machine learning in medical settings, and reviews progress in DL . Additionally, it summarizes prior studies that have informed the development of automated models for detecting DR.

1.2. Healthcare

Healthcare encompasses a wide range of services aimed at promoting, maintaining, and restoring health. These services include preventive, curative, rehabilitative, and palliative care, all designed to improve both individual and public health outcomes. The healthcare sector involves medical professionals, healthcare institutions, government bodies, and technological advancements that enhance service delivery [25].

1.2.1 Stages of Healthcare

Healthcare services can be broadly classified into three primary stages:

1.2.1.1 Primary Healthcare

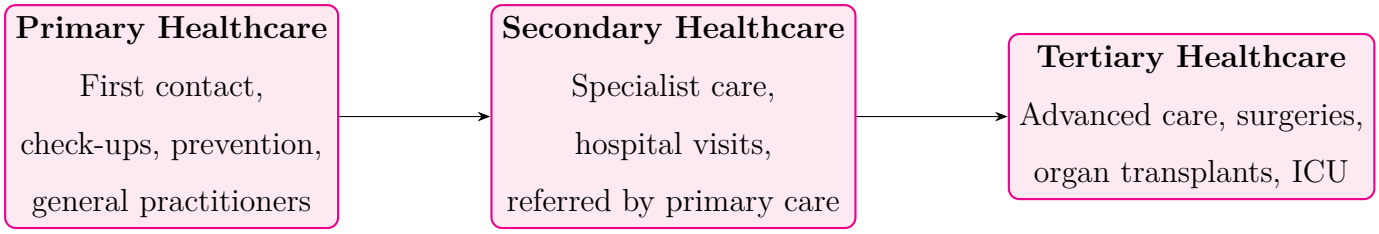
Primary healthcare serves as the first point of contact for individuals seeking medical services. It includes routine check-ups, vaccinations, early disease detection, and general wellness programs. Family physicians, general practitioners, and community health workers play a significant role in primary healthcare. Preventive care, health education, and lifestyle modifications are key components of this stage [26].

1.2.1.2 Secondary Healthcare

Secondary healthcare involves specialized medical services provided by healthcare professionals such as cardiologists, endocrinologists, and ophthalmologists. This stage includes hospital visits, diagnostic tests, and treatments for more complex health conditions. Patients are usually referred to secondary healthcare facilities by primary care providers when advanced medical intervention is required [27].

1.2.1.3 Tertiary Healthcare

Tertiary healthcare consists of highly specialized medical services offered in advanced healthcare institutions, including major hospitals and research centers. It includes organ transplants, cancer treatment, neurosurgery, and intensive care services. This stage often requires state-of-the-art medical equipment and expertise [28].



1.2.2 Role of Technology in Healthcare

Technological advancements have revolutionized healthcare, leading to better diagnosis, treatment, and patient management [29]. Some of the key technological contributions in healthcare include:

1.2.2.1 Medical Imaging and Diagnostics

Advancements in imaging technologies, such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and Optical Coherence Tomography (OCT), have significantly improved disease detection and monitoring. AI-powered imaging systems assist radiologists in identifying abnormalities with greater accuracy and efficiency [30, 31].

1.2.2.2 Telemedicine and Remote Monitoring

Telemedicine has expanded healthcare access by allowing patients to consult doctors remotely using video conferencing and mobile health applications [32]. Remote patient monitoring (RPM) enables continuous tracking of vital signs, such as blood pressure and glucose levels, reducing hospital visits and improving chronic disease management [33].

1.2.2.3 Electronic Health Records (EHR)

The adoption of electronic health records has streamlined patient data management, ensuring efficient record-keeping and improved coordination between healthcare providers.

EHR systems facilitate quick access to medical history, reducing errors and enhancing patient safety [34].

1.3. Diabetes

Diabetes is a chronic metabolic disorder characterized by high blood glucose levels due to the body's inability to produce or effectively utilize insulin. It is a major global health concern, affecting millions of individuals and leading to severe complications if left uncontrolled. The prevalence of diabetes has been increasing worldwide due to lifestyle changes, aging populations, and genetic predisposition [35].

1.3.1 Types of Diabetes

Diabetes is primarily classified into three main types:

1.3.1.1 Type 1 Diabetes

Type 1 diabetes, also known as insulin-dependent diabetes mellitus (IDDM), is an autoimmune condition where the immune system attacks the insulin-producing beta cells in the pancreas. This leads to an absolute deficiency of insulin, requiring lifelong insulin therapy. Type 1 diabetes is commonly diagnosed in children and young adults, and its exact cause remains unknown, though genetic and environmental factors play a role [36].

1.3.1.2 Type 2 Diabetes

Type 2 diabetes, or non-insulin-dependent diabetes mellitus (NIDDM), is the most common form of diabetes, accounting for over 90% of cases. It is characterized by insulin resistance, where the body's cells fail to respond effectively to insulin. Over time, insulin production declines, leading to hyperglycemia. Type 2 diabetes is strongly associated with obesity, physical inactivity, and genetic predisposition [37].

1.3.1.3 Gestational Diabetes

Gestational diabetes occurs during pregnancy due to hormonal changes that affect insulin sensitivity. While it often resolves after childbirth, it increases the risk of develop-

ing Type 2 diabetes later in life. Proper monitoring and management during pregnancy are crucial to prevent complications for both the mother and baby [38].

1.3.2 Complications of Diabetes

Uncontrolled diabetes can lead to severe long-term complications affecting multiple organ systems:

- **Cardiovascular Diseases:** Diabetes significantly increases the risk of heart disease, hypertension, and stroke [39].
- **Diabetic Retinopathy:** A leading cause of blindness due to damage to the retinal blood vessels [40].
- **Diabetic Neuropathy:** Nerve damage resulting in pain, numbness, and loss of function, particularly in the extremities [41].
- **Diabetic Nephropathy:** Progressive kidney disease leading to renal failure if untreated [42].
- **Diabetic Foot Ulcers:** Poor circulation and nerve damage increase the risk of infections and amputations [43].

1.3.3 Management and Treatment

Diabetes management focuses on controlling blood glucose levels to prevent complications. Key strategies include:

1.3.3.1 Lifestyle Modifications

A healthy diet, regular exercise, and weight management are essential in diabetes prevention and control. Diets rich in fiber, whole grains, and lean proteins help regulate blood sugar levels [44].

1.3.3.2 Pharmacological Treatments

Different medications are used based on the type and severity of diabetes:

- **Insulin Therapy:** Essential for Type 1 diabetes and sometimes required for Type 2 diabetes in advanced stages.
- **Oral Hypoglycemic Agents:** Metformin, sulfonylureas, and SGLT2 inhibitors help manage Type 2 diabetes by improving insulin sensitivity or reducing glucose production [45].
- **GLP-1 Receptor Agonists:** These drugs enhance insulin secretion and promote weight loss, benefiting Type 2 diabetes patients [46].

1.4. Diabetic Retinopathy

DR is a serious microvascular complication of diabetes that affects the retina, potentially leading to vision impairment and blindness if left untreated. It is caused by prolonged high blood glucose levels, which damage the small blood vessels in the retina, leading to leakage, swelling, and abnormal vessel growth. DR is one of the leading causes of preventable blindness worldwide, particularly among working-age adults [**?**]. Early detection and timely intervention are crucial for effective management and prevention of severe visual loss.

1.4.1 Normal Eye vs. Diabetic Eye

Before we delve into the pathological changes associated with diabetic eye disease, it is important for us to understand the structure and function of a normal, healthy eye. This foundational knowledge will allow us to better recognize the anatomical and physiological alterations that occur in individuals affected by diabetes. In the following section, we will outline the key components and roles of the normal eye, which will serve as a reference point when comparing the effects of diabetes on ocular health.

1.4.1.1 Structure and Function of a Normal Eye

The human eye is a **highly specialized sensory organ** that enables vision by detecting and processing light. It consists of multiple interconnected structures that work in harmony to form clear images. The **retina**, a thin layer of tissue at the back of the

eye, is responsible for converting light into electrical signals, which are then transmitted to the brain via the **optic nerve**.

A healthy eye maintains **stable and well-structured retinal blood vessels**, ensuring a continuous supply of oxygen and nutrients while efficiently removing waste products. This vascular integrity is crucial for preserving **sharp and accurate vision**.

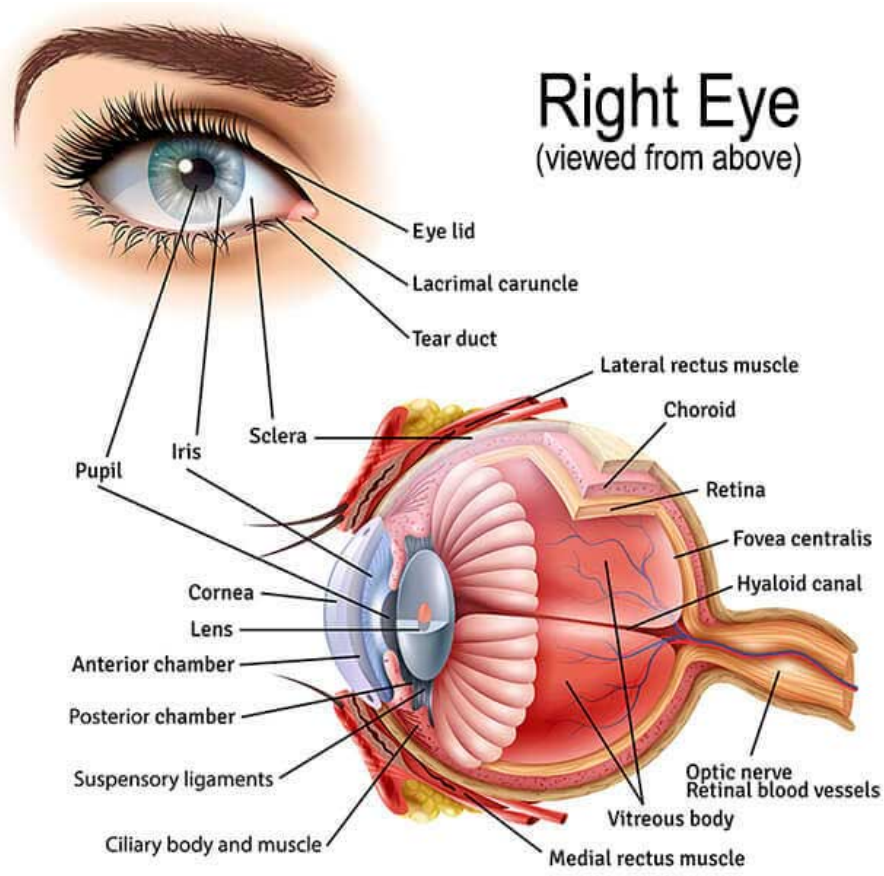


Figure 1.1: High-resolution fundus image of a normal eye [1].

1.4.1.2 Effects of Diabetes on the Eye

In individuals with diabetes, prolonged high blood sugar levels can cause progressive damage to the retinal blood vessels, leading to DR. This damage manifests in multiple ways:

- **Microaneurysms:** Tiny bulges in retinal blood vessels that can rupture and leak blood.
- **Hemorrhages:** Small spots of bleeding in the retina due to weakened vessel walls.
- **Exudates:** Fatty deposits formed from leaking blood vessels, leading to vision impairment.
- **Neovascularization:** Abnormal growth of fragile blood vessels, which may rupture and cause severe complications, such as vitreous hemorrhage and retinal detachment.

1.4.1.3 Key Differences Between a Normal and Diabetic Eye

Feature	Normal Eye	Diabetic Eye
Retinal Blood Vessels	Well-structured, ensuring proper oxygen supply.	Damaged, with leakage and potential blockage.
Oxygen Supply	Adequate, supporting normal function.	Reduced, leading to ischemia and abnormal vessels.
Vision	Clear and sharp, no visual obstructions.	Blurred or distorted due to hemorrhages.
Macular Condition	Macula remains unaffected, ensuring sharp vision.	Swelling (Macular Edema), causing central vision loss.

Table 1.1. Comparative Analysis of a Healthy Eye vs. a Diabetic Eye.

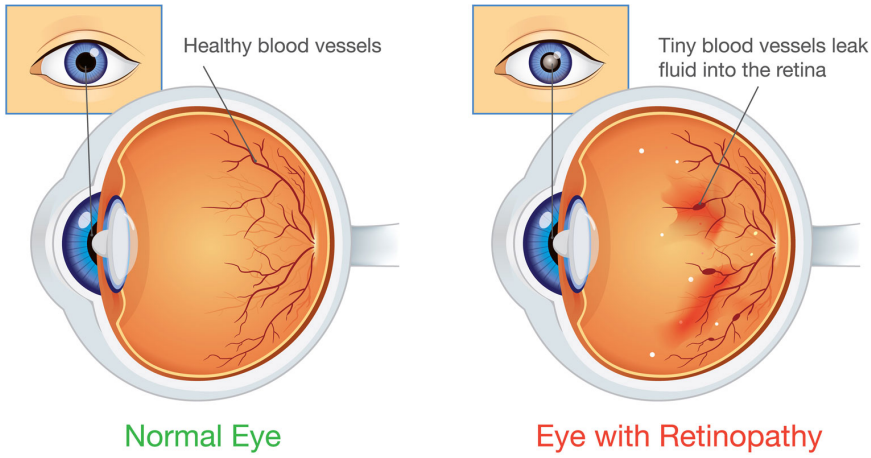


Figure 1.2: Comparison between a normal eye and an eye with diabetic retinopathy [2].

1.4.2 Stages of Diabetic Retinopathy

DR progresses through different stages, categorized based on the severity of retinal damage. DR develops through the following stages:

- **Normal Retina:** A healthy retina with well-structured blood vessels that ensure an adequate supply of oxygen and nutrients. There are no signs of microaneurysms, hemorrhages, or abnormal vessel growth.
- **Mild Non-Proliferative DR (NPDR):** The earliest stage, characterized by the presence of microaneurysms—tiny bulges in the retinal blood vessels. Small amounts of fluid leakage and minor hemorrhages may occur but do not significantly impair vision.
- **Moderate NPDR:** At this stage, the number of microaneurysms and hemorrhages increases. Blockages in some retinal blood vessels begin to restrict oxygen supply, leading to localized damage within the retina.
- **Severe NPDR:** A significant portion of retinal blood vessels becomes blocked, causing a substantial reduction in oxygen delivery (ischemia). In response, the retina signals for the growth of new blood vessels, increasing the risk of progression to the proliferative stage.

- **Proliferative DR (PDR):** The most advanced stage, characterized by the growth of new, fragile blood vessels (neovascularization). These abnormal vessels are prone to rupture, leading to complications such as vitreous hemorrhages, retinal detachment, and severe vision impairment.

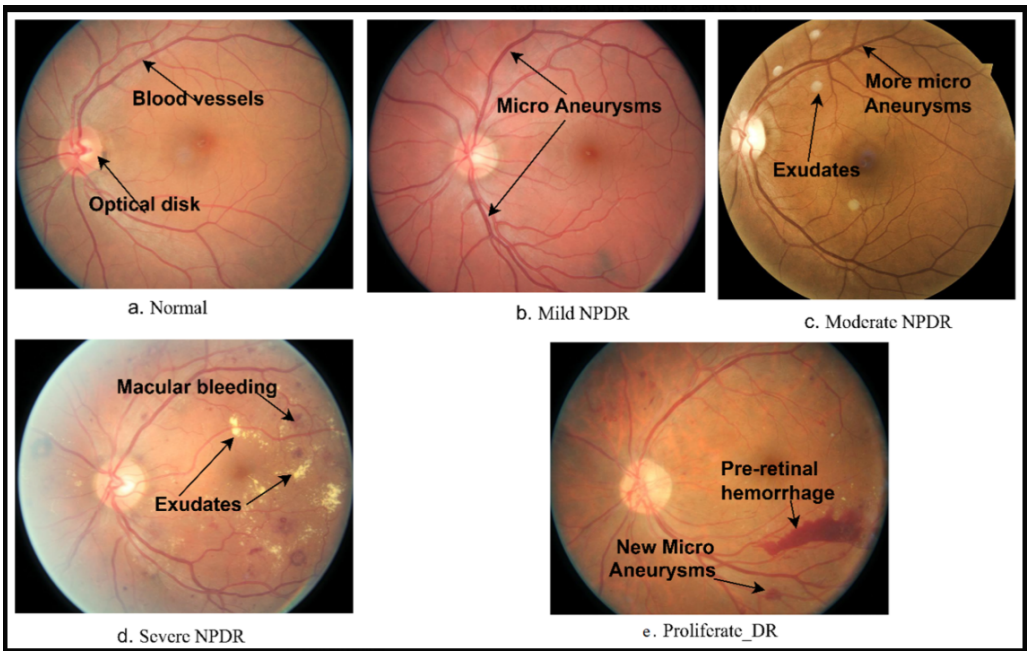


Figure 1.3: Stages of Diabetic Retinopathy [3].

1.5. DR Imaging Techniques

The early detection of DR largely depends on advanced retinal imaging techniques that enable ophthalmologists to identify and assess retinal abnormalities. Several key imaging modalities are commonly used in clinical practice, including fundus photography , Optical Coherence Tomography (OCT) , Fluorescein Angiography (FA) , and Ultra-Widefield Imaging . Each of these techniques offers unique advantages in evaluating the severity of DR and monitoring its progression over time [47].

1.5.1 Fundus Photography:

Fundus photography captures high-resolution images of the retinal structures, allowing clinicians to detect early signs of DR such as microaneurysms, hemorrhages, and exudates , which are considered key clinical indicators of the disease [48]. This technique is non-invasive, cost-effective, and widely implemented in large-scale screening programs due to its ease of use and reproducibility.

Figure 1.4: Diagram of the Transfer Learning process. Source: ResearchGate [4]

1.5.2 Optical Coherence Tomography (OCT):

OCT is a non-invasive imaging technique that provides high-resolution, cross-sectional images of the retina, enabling detailed visualization of its layered structure. This modality is particularly valuable in the detection and monitoring of diabetic macular edema (DME) , a sight-threatening complication of DR [49].

By offering precise measurements of retinal thickness and identifying fluid accumulation in the macula, OCT plays a critical role in guiding treatment decisions and assessing therapeutic response in patients with DR.

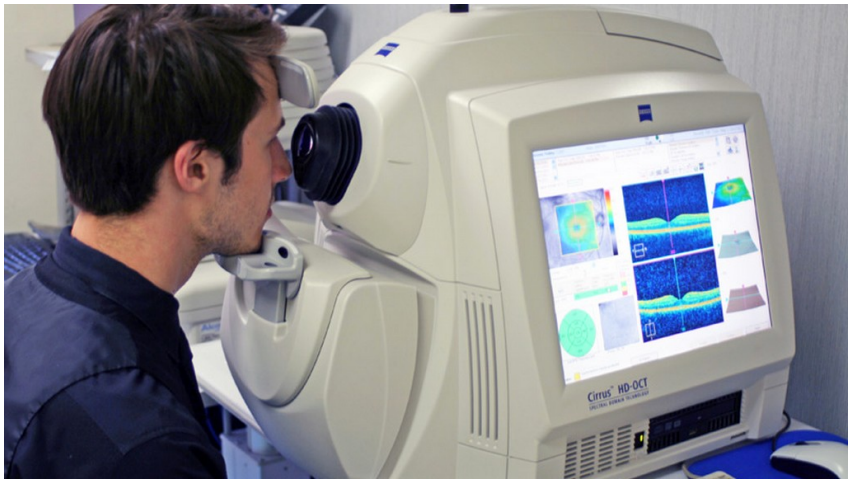


Figure 1.5: Optical Coherence Tomography (OCT) scan showing retinal thickening [5].

1.5.3 Fluorescein Angiography (FA):

FA is a diagnostic imaging technique that involves the intravenous injection of fluorescein dye, which circulates through the retinal vasculature and becomes visible under blue light illumination [50].

This method enables the visualization of abnormal blood vessels, vascular leakage, and ischemic areas in the retina , making it an essential tool in the diagnosis and management of proliferative diabetic retinopathy (PDR) and other vision-threatening complications of diabetes.

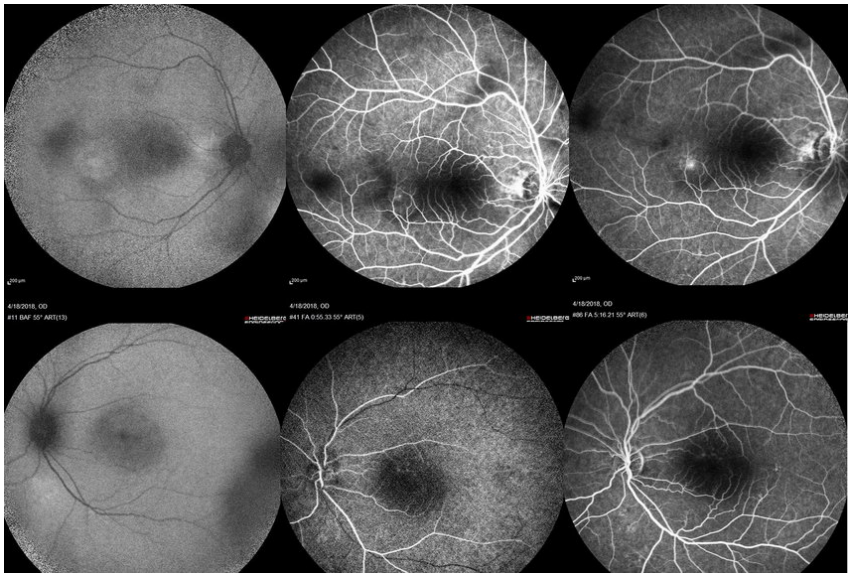


Figure 1.6: Fluorescein angiography image showing retinal leakage [6].

1.5.4 Ultra-Wide Field Imaging (UWFI)

UWFI is an advanced retinal imaging technique that provides a panoramic view of the retina , capturing up to 200 degrees of the retinal surface in a single capture [51]. This expanded field of view enables the detection of peripheral retinal lesions , such as neovascularization and ischemic areas, which are often overlooked using conventional fundus photography. Due to its enhanced sensitivity in identifying early pathological changes, UWFI has become an increasingly valuable tool in the screening and management of DR.

hyperref

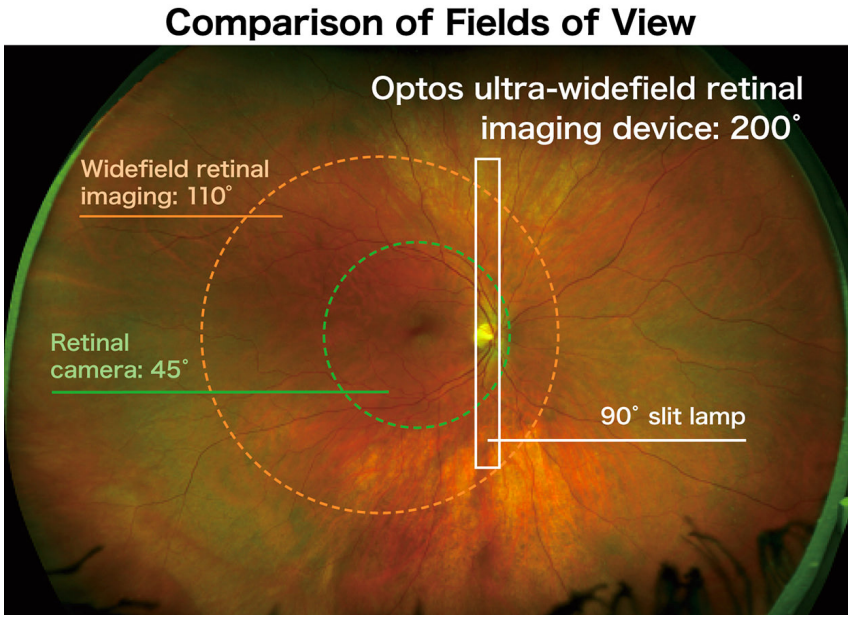


Figure 1.7: Ultra-wide field imaging capturing peripheral retinal lesions [7].

1.6. Machine Learning in Healthcare

1.6.1 Overview of Machine Learning Paradigms

ML has transformed healthcare by enabling faster diagnosis, predicting patient outcomes, and personalizing treatment plans. ML techniques are broadly classified into four categories: supervised, unsupervised, semi-supervised, and reinforcement learning.

1.6.1.1 Supervised Learning

Supervised learning is a type of ML where a model learns from labeled training data. It includes algorithms like Linear Regression, Logistic Regression, Decision Trees, and Support Vector Machines (SVM) [52].

hyperref

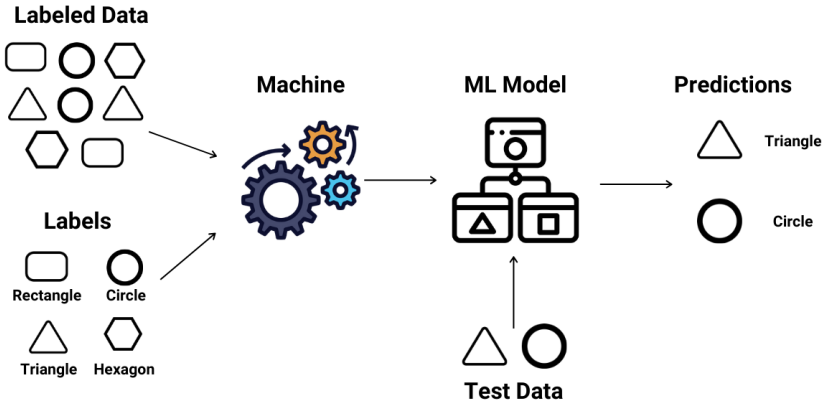


Figure 1.8: Illustration of supervised ML workflow [8].

1.6.1.2 Unsupervised Learning

Unsupervised learning deals with unlabeled data, identifying patterns without explicit supervision. Common techniques include clustering (K-Means, Hierarchical Clustering) and dimensionality reduction (PCA, Autoencoders) [53].

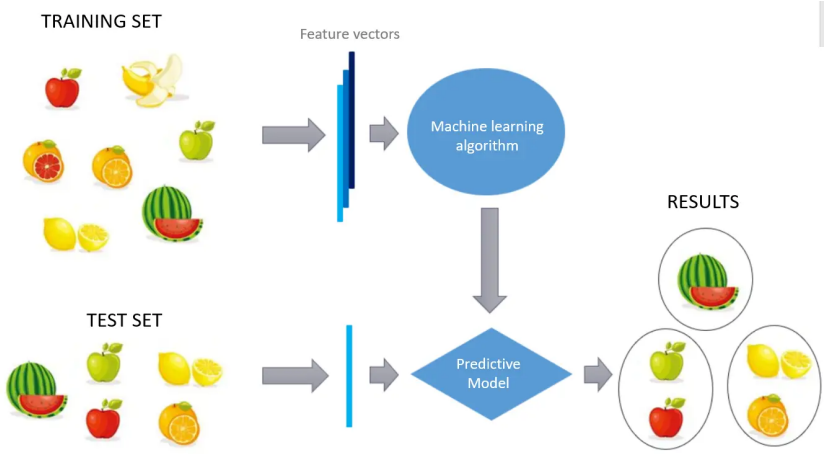


Figure 1.9: Illustration related to unsupervised learning techniques [9].

1.6.1.3 Semi-Supervised Learning

Semi-supervised learning combines both labeled and unlabeled data to improve model performance, commonly used in medical imaging applications [54].

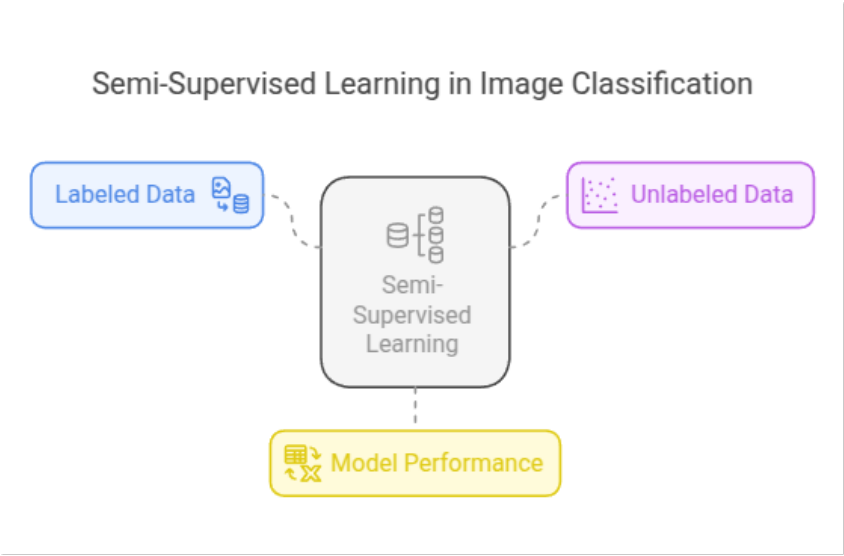


Figure 1.10: Conceptual diagram of Semi-Supervised Learning.

1.6.1.4 Reinforcement Learning (RL)

RL is an ML paradigm where an agent learns by interacting with the environment and receiving rewards. It is widely used in robotic-assisted surgery and adaptive treatment planning [55].

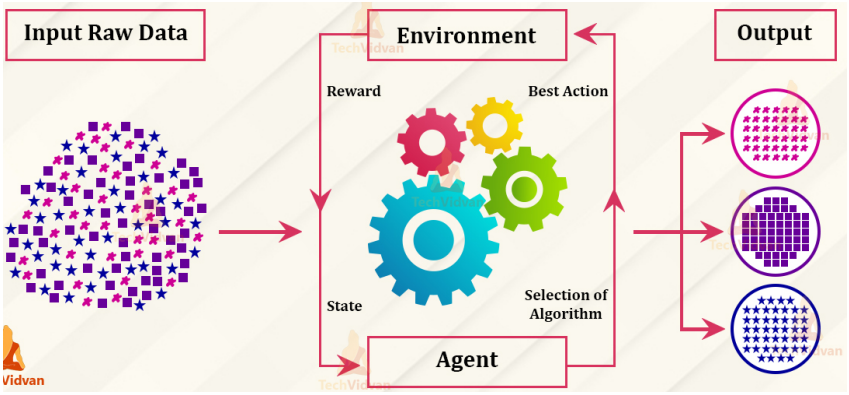


Figure 1.11: Illustration related to reinforcement learning [10].

1.6.2 Application of Machine Learning in DR Diagnosis

DR is a severe eye condition resulting from complications of diabetes, which can lead to irreversible vision loss if not detected early. Due to the increasing global prevalence of diabetes, early and accurate detection of DR is a critical public health priority. ML has emerged as a transformative technology for automating the diagnosis of DR, enabling timely and cost-effective screening.

1.6.2.1 Role of ML in DR Screening

Traditional DR screening relies on manual inspection of fundus images by ophthalmologists. This process is time-consuming and subject to inter-observer variability. ML-based models, especially those trained on large labeled datasets, offer a scalable and standardized solution by:

- Automating Image Analysis
- Improving Diagnostic Consistency
- Reducing Screening Costs
- Facilitating Early Detection

1.6.2.2 Common ML Techniques Used in DR Diagnosis

Supervised Learning: Supervised ML models are trained on labeled fundus images with known DR severity grades. These models learn to classify new images into categories such as No DR, Mild, Moderate, Severe, and Proliferative DR.

Algorithms used:

- Support Vector Machines (SVM)
- Random Forests
- K-Nearest Neighbors (KNN)
- Logistic Regression

1.7. Deep Learning

1.7.1 History of Deep Learning

DL has undergone significant advancements over the decades, driven by improvements in computational power, algorithmic innovations, and the availability of large-scale datasets. The evolution of DL can be divided into three major phases: **early foundations (1940s–1960s), resurgence with multi-layer networks (1980s–2000s), and the modern DL revolution (2006–present).**

1.7.1.1 Early Foundations (1940s - 1960s)

The concept of artificial neural networks began with theoretical models inspired by biological neurons.

- **1943:** Warren McCulloch and Walter Pitts proposed the first artificial neuron, known as the *McCulloch-Pitts model*, which demonstrated how simple logical functions could be computed using neural networks [56].
- **1958:** Frank Rosenblatt developed the *Perceptron*, the first trainable neural network, which could classify simple patterns by adjusting its weights based on input data [57].
- **1969:** Marvin Minsky and Seymour Papert published *Perceptrons*, a book highlighting the limitations of single-layer networks, particularly their inability to solve non-linearly separable problems like XOR. This led to a decline in neural network research for over a decade [58].

1.7.1.2 Resurgence with Multi-Layer Networks (1980s - 2000s)

In the 1980s, research in neural networks regained momentum with the introduction of multi-layer perceptrons and advanced training techniques.

- **1986:** Geoffrey Hinton, David Rumelhart, and Ronald Williams introduced the *Backpropagation algorithm*, which enabled multi-layer networks to adjust their weights efficiently, significantly improving neural network performance [59].
- **1990s:** Despite the advances, traditional ML models such as *Support Vector Machines (SVMs)* and *Decision Trees* outperformed neural networks in practical applications, leading to another AI winter.
- **2000s:** Computational advancements and larger datasets reignited interest in DL, paving the way for more complex models.

1.7.1.3 Modern Deep Learning Revolution (2006 - Present)

The breakthrough in DL came with advancements in hardware (GPUs) and novel architectures.

- **2006:** Geoffrey Hinton introduced *Deep Belief Networks (DBNs)*, demonstrating that deep architectures could be efficiently trained in a layer-wise manner, reviving DL research [60].
- **2012:** Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton developed *AlexNet*, a deep CNN that achieved groundbreaking results in the ImageNet competition, proving the potential of CNNs for image classification [61].
- **2014:** Ian Goodfellow introduced *GAN*, a new approach to generating synthetic data, with applications in medical imaging and data augmentation [62].
- **2017:** Google researchers introduced *Transformers*, an attention-based DL model that revolutionized natural language processing (NLP) and led to models like BERT and GPT [63].
- **2020 - Present:** Large-scale AI models, including *GPT-4* and *Vision Transformers (ViTs)*, have expanded DL applications across multiple domains, including healthcare, robotics, and autonomous systems.

1.7.1.4 Summary of Key Milestones

Year	Milestone	Key Contribution
1943	McCulloch-Pitts Model	First computational neuron model [56].
1958	Perceptron	First trainable neural network [57].
1969	Perceptron Limitations	Highlighted fundamental flaws in single-layer networks [58].
1986	Backpropagation	Enabled deep networks with efficient training [59].
2006	Deep Belief Networks	Revived interest in DL [60].
2012	AlexNet	CNN breakthrough in image recognition [61].
2014	GANs	Introduced adversarial learning for synthetic data [62].
2017	Transformers	Revolutionized NLP with attention mechanisms [63].
2020-Present	Large AI Models	Scaling DL across multiple fields.

Table 1.2: Key milestones in deep learning history.

DL continues to evolve, with ongoing research focusing on improving model efficiency, interpretability, and ethical AI development.

1.7.2 Introduction to DL and Neural Networks

DL is a branch of ML that involves training artificial neural networks to perform tasks such as classification, prediction, and pattern recognition. These networks consist of layers of neurons that mimic the structure of the human brain. In the context of (DR) detection, DL models are used to analyze and classify retinal images based on the presence of diabetic complications.

1.7.2.1 Main Layers of Deep Learning

DL models consist of three fundamental layers, as illustrated in Figure 1.12:

- **Input Layer:** Receives the raw data (e.g., images, text, numerical values) and serves as the entry point to the network. Each input feature is represented as a neuron.
- **Hidden Layers:** These layers perform **feature extraction and transformation** by applying weights, biases, and activation functions [53]. Each neuron in a hidden layer is connected to neurons in the next layer, allowing complex representations to be learned.
- **Output Layer:** Produces the final result based on the processed features. This can be a **classification label** (e.g., cat vs. dog), a regression value, or any other type of prediction.

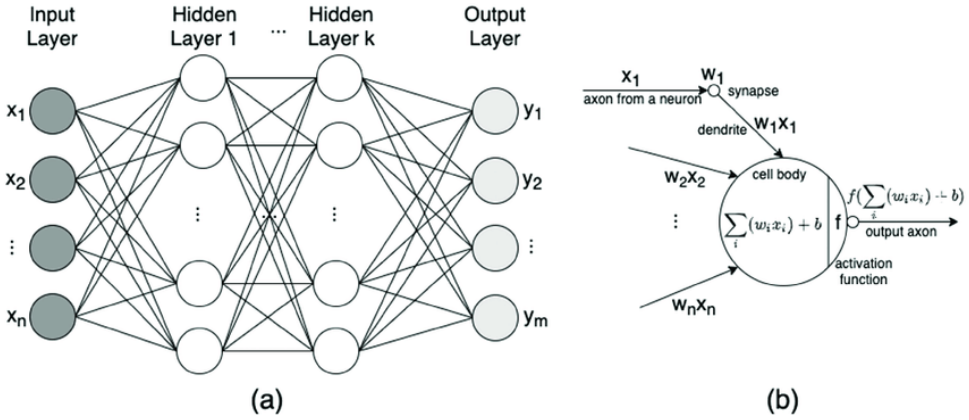


Figure 1.12: Visualization of DL Architecture: Layers and Neuron Interactions [11].

1.7.2.2 CNN Architecture

CNNs are a class of DL models designed to process grid-structured data, such as images and videos. They efficiently extract spatial hierarchies of features, making them particularly effective in image classification, object detection, and medical imaging [64].

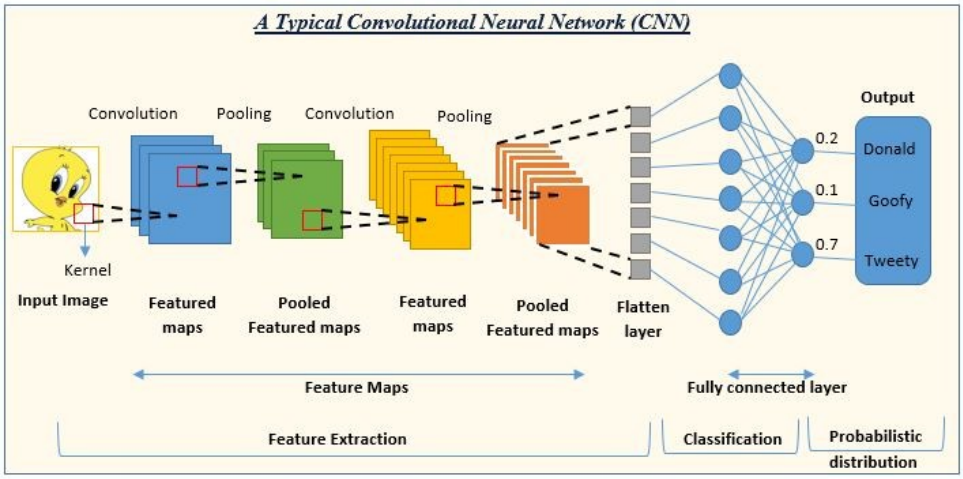


Figure 1.13: CNN Architecture Overview [12].

Convolutional Layer: is the core building block of **CNN**. It is responsible for automatically learning spatial hierarchies of features by applying a set of **learnable filters (kernels)** to the input data.

Key Components

- **Filters (Kernels):** Small matrices that slide over the input to extract patterns such as edges, textures, and shapes.
- **Stride:** Defines the step size of the filter movement across the input. A higher stride reduces spatial dimensions.
- **Padding:** Adds extra pixels around the input to preserve spatial dimensions after convolution.
- **Activation Function:** Typically, the **ReLU (Rectified Linear Unit)** function is applied to introduce non-linearity and improve feature extraction.

Mathematical Representation The convolution operation can be mathematically expressed as:

$$Y(i,j) = \sum_{m} \sum_{n} X(i+m, j+n) W(m,n) \tag{1.1}$$

where:

- $Y(i,j)$ represents the output feature map at position (i,j).

- X is the input matrix.
- W is the kernel (filter).

This process enables convolutional layers to extract meaningful representations from input data, making CNNs highly effective for **image classification, object detection, and segmentation** [64].

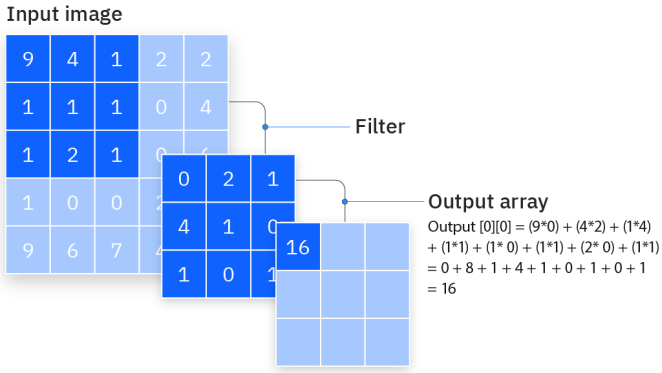


Figure 1.14: Convolutional layer curve [13].

ReLU Activation Function: DL models rely on activation functions to introduce non-linearity, enabling them to learn complex patterns. One of the most widely used activation functions is the **Rectified Linear Unit** .

The ReLU is defined as:

$$f(x) = \max(0, x) \tag{1.2}$$

It returns x if $x > 0$ and 0 otherwise. This function is computationally efficient and helps mitigate the vanishing gradient problem, making it ideal for deep networks [65].

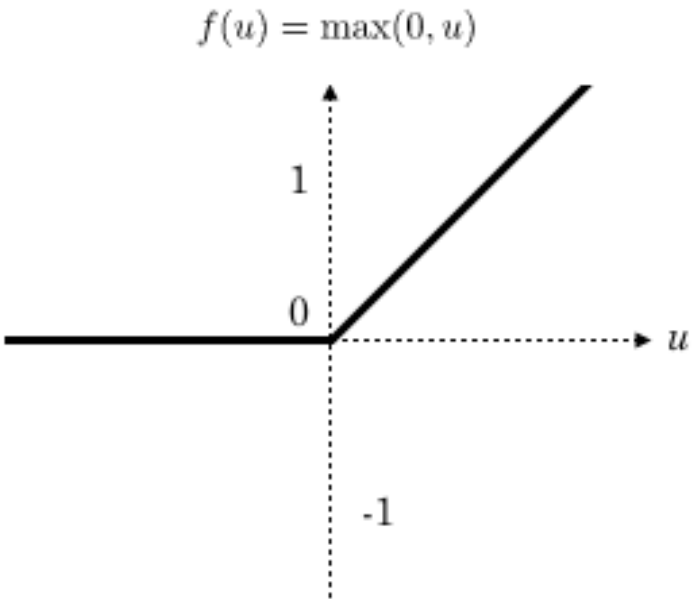


Figure 1.15: Visualization of the ReLU Activation Function [14].

Batch Normalization: Batch Normalization stabilizes training by normalizing activations across mini-batches:

$$\hat{x} = \frac{x - \mu}{\sigma} \tag{1.3}$$

where μ is the batch mean, and σ is the standard deviation. The final transformation is:

$$y = \gamma\hat{x} + \beta \tag{1.4}$$

which helps in reducing internal covariate shift and improving model regularization [66].

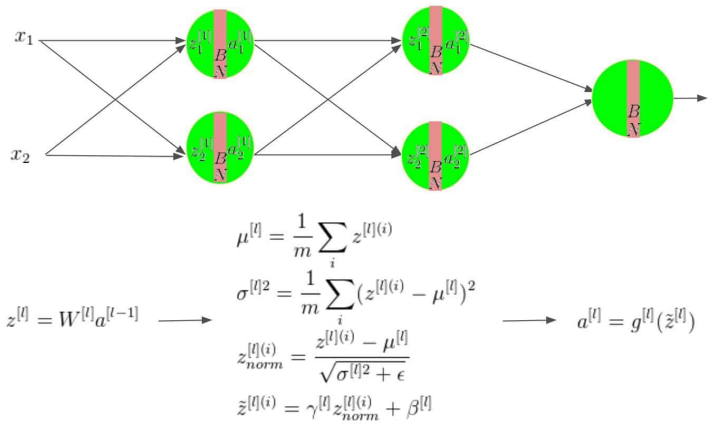


Figure 1.16: Batch Normalization in Deep Networks [15].

Pooling Layer : Pooling layers are essential in CNNs for reducing the spatial dimensions of feature maps while retaining key information. This process lowers computational cost, minimizes overfitting, and improves generalization [53].

There are several pooling methods. **Max Pooling** selects the highest value in a window, highlighting dominant features [64]. **Average Pooling** computes the mean, providing a smoother feature representation [67]. **Global Pooling** (Max or Average) condenses each feature map into a single value, often used before classification [68].

hyperref

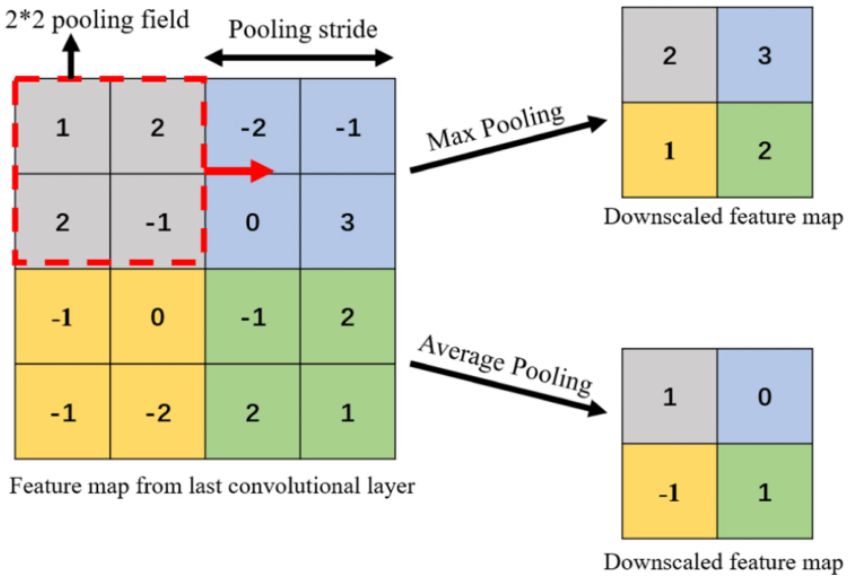


Figure 1.17: Max and Average Pooling Layer Examples [16].

Flattening Layer: Is a crucial component in DL architectures, particularly in CNN. It is responsible for converting multi-dimensional outputs from convolutional and pooling layers into a one-dimensional vector that can be processed by fully connected layers.

After convolutional and pooling layers extract spatial hierarchies and important features from the input data, the Flattening Layer transforms the high-dimensional feature maps into a one-dimensional array. This transformation enables the network to utilize fully connected layers for classification or regression tasks [64].

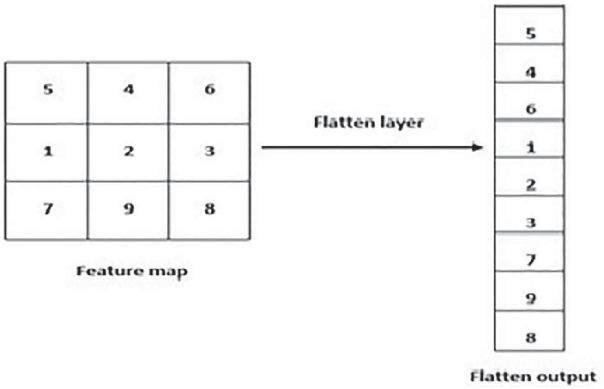


Figure 1.18: Illustration of the Flattening Layer in a CNN architecture [17].

Fully Connected Layer: The **Fully Connected Layer** is a crucial component in deep neural networks, where each neuron is connected to every neuron in the previous layer. This dense connectivity allows the model to learn complex representations by computing a weighted sum of input features, adding a bias term, and applying an activation function such as ReLU or Softmax. According to [53], fully connected layers play a vital role in final decision-making processes within DL architectures.

However, the FC layer introduces a large number of parameters, increasing the risk of overfitting. To address this, [69] proposed the **Dropout** technique, which randomly deactivates neurons during training to enhance generalization. Additionally, [66] introduced **Batch Normalization**, a method that stabilizes activations, accelerates convergence, and improves model robustness.

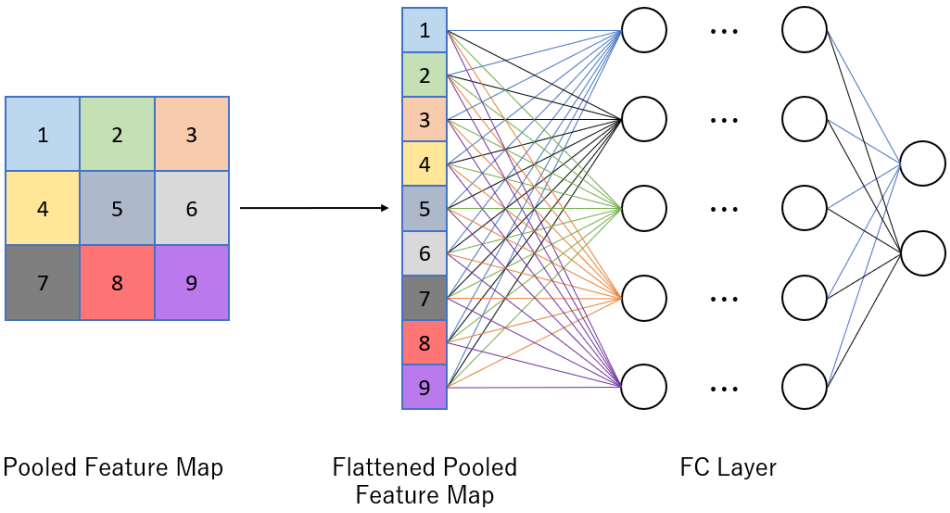


Figure 1.19: Fully Connected Layer in a CNN [18].

1.7.2.3 Significance of CNNs in Medical Image Analysis

CNNs have revolutionized medical image analysis by automating feature extraction and achieving high accuracy in identifying conditions such as DR. Unlike traditional methods that rely on manual feature engineering, CNNs learn directly from the images, making them powerful tools for analyzing complex retinal images. This makes them ideal for large-scale screenings and providing real-time decision support for physicians.

1.8. Transfer Learning

TL is a strong way to use in deep learning. It works by taking a model made for one job and using it as the start for another job that is kind of the same. It's very good when there's not much labeled data for the new job. This lets the model use what it has learned before [70].

In CNN, TL often means taking a model trained on a big dataset like ImageNet. The first parts, which pick up basic stuff like lines, textures, and shapes, are kept. The parts that come later are fine-tuned or changed to fit the new job and data [71].

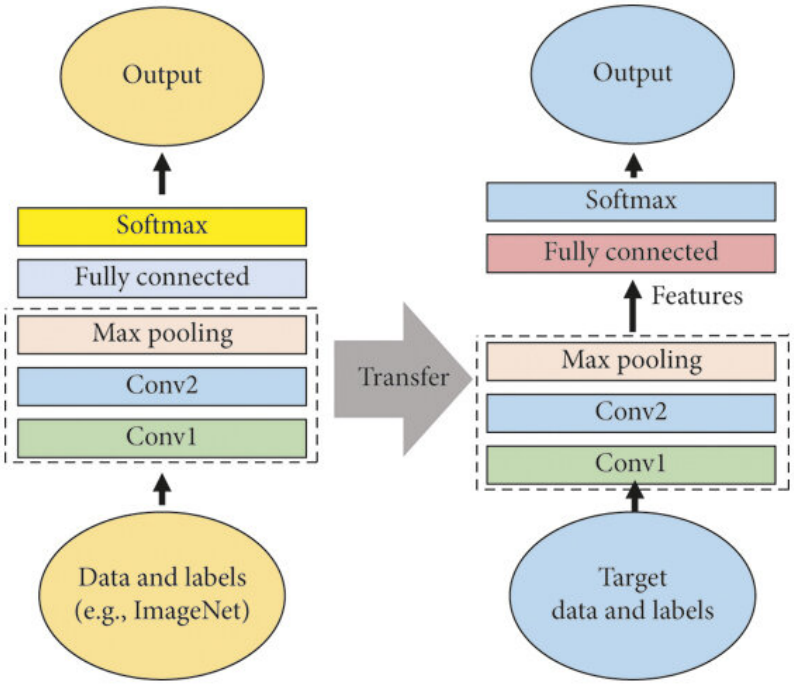


Figure 1.20: Diagram of the Transfer Learning process. Source: ResearchGate [4]

1.9. Pretrained Model

PM are DL models that have been trained on large datasets, often to perform general tasks like image classification, object detection, or language modeling. These models, such as ResNet or VGGNet, can be adapted for specific tasks like DR detection through a technique known as TL.

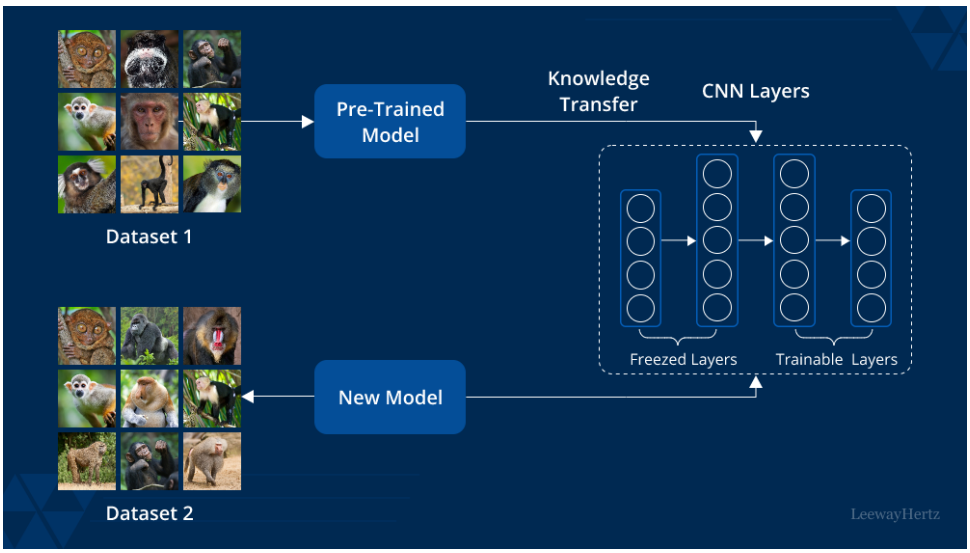


Figure 1.21: pre-trained model for TL [19].

1.9.1 The Concept of Pretraining on Large Datasets

Pretraining involves training a model on a large and diverse dataset, enabling it to learn fundamental features like edges and textures that are useful across various tasks. Once trained, the model can be fine-tuned or "adapted" on a smaller, domain-specific dataset for DR detection, saving time and computational resources.

1.9.2 Benefits of TL for Medical Imaging Tasks

Transfer learning offers several benefits, particularly in medical imaging:

- **Reduced Training Time:** Pretrained models reduce the time required to train a model from scratch.
- **Improved Performance:** Models fine-tuned with specific medical images typically outperform models trained from scratch.
- **Data Efficiency:** TL makes it possible to leverage smaller medical datasets effectively.

1.10. Building and Using Pretrained Models for Diabetic Retinopathy Detection

DL models have played a transformative role in the automated detection of DR. Pretrained architectures—originally developed for general image recognition tasks—have been widely adapted for analyzing retinal fundus images with high diagnostic accuracy. These models serve as powerful tools for early detection, severity grading, and clinical decision support in DR management.

1.10.1 Core Architectures for Vision Tasks in DR

Several DL frameworks have become foundational in medical image analysis, particularly in the context of retinal disease detection. The most notable include **CNN** such as ResNet, VGGNet, and InceptionNet, as well as emerging **Vision Transformer (ViT)-based models**, which are increasingly being explored for their effectiveness in medical imaging tasks.

1.10.1.1 Deep Dive into CNN Models for Medical Imaging

Convolutional Neural Networks (CNNs) form the backbone of most medical image classification systems due to their ability to automatically extract spatial features from visual data. Architectures like **ResNet**, **VGGNet**, and **InceptionNet**, initially trained on large-scale natural image datasets such as ImageNet, have been fine-tuned on retinal image datasets for DR detection. These networks can identify key pathological indicators—such as microaneurysms, hemorrhages, and exudates—that correlate with DR severity and progression.

1.10.1.2 Vision Transformers (ViTs): Adapting Transformers for Image Recognition

Originally designed for natural language processing, **Transformer-based models** have been successfully adapted for computer vision through the introduction of **Vision Transformers (ViTs)** [72]. ViTs divide input images into fixed-size patches, linearly embed them, and use self-attention mechanisms to model relationships across the entire image.

This allows ViTs to capture long-range dependencies and global contextual patterns, making them especially effective when trained on large and diverse medical image datasets. Recent studies indicate that ViTs can achieve performance comparable to or even surpassing CNNs in certain DR classification tasks [73].

1.10.1.3 Comparison Between CNNs and ViTs in DR Classification Tasks

While CNNs remain dominant in many medical imaging applications due to their strong feature extraction capabilities and interpretability, ViTs offer promising advantages, particularly in modeling global contextual information. CNNs excel at capturing local spatial hierarchies and are often more suitable for smaller datasets and structured feature learning. In contrast, ViTs generally require larger training sets to perform optimally but may outperform CNNs when sufficient data is available—especially in detecting subtle and complex patterns associated with advanced stages of DR [74].

The choice between CNNs and ViTs depends on factors such as dataset size, computational resources, and task complexity. Hybrid approaches combining both architectures

are also being explored to leverage the strengths of each.

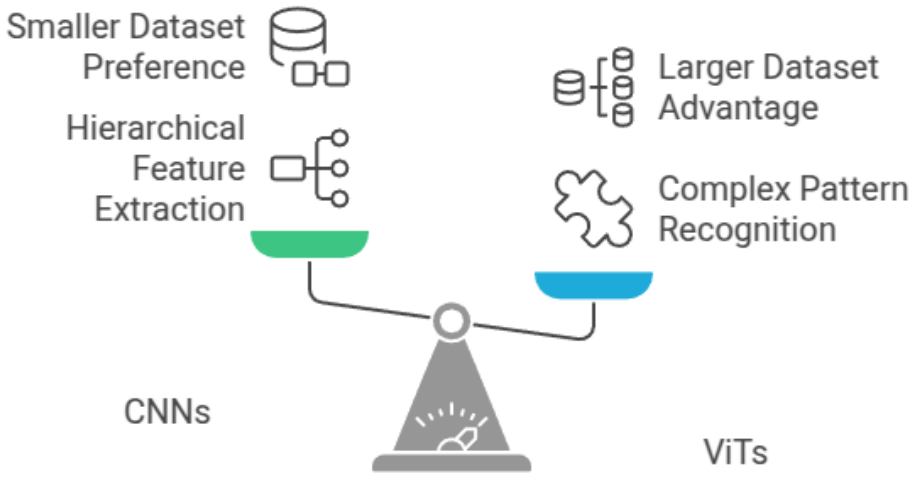


Figure 1.22: Comparing CNN ViTs in DR Classification

1.10.2 Training Techniques and Methods

Training DL models for DR detection involves several key considerations:

Supervised Learning Objectives for Classification/Segmentation

Supervised learning is the most widely used approach in DR classification tasks. Models are trained on labeled images, where each input image is paired with its corresponding DR severity label. Common objectives include:

- **Cross-Entropy Loss:** The most widely used loss function for classification tasks, measuring the difference between predicted and actual classifications.
- **Dice Coefficient:** A loss function commonly used in segmentation tasks, such as segmenting the retina region affected by DR.

The Importance of Dataset Size and Diversity

To ensure deep models generalize well, it is essential to have a large and diverse dataset. A dataset with varying image conditions (e.g., lighting, image quality, demographics) enables the model to learn robust features applicable to real-world scenarios.

Introduction to Scaling Laws

Scaling laws in DL refer to the relationship between dataset size, model size, and performance. As data and model complexity increase, performance typically improves—though with diminishing returns. These concepts are especially relevant when adapting models for DR detection, as large datasets can improve accuracy but also demand substantial computational resources.

1.10.3 Fine-Tuning Pretrained Models: Optimization Strategies

Fine-tuning is the process of adapting a PM for a specific task, such as DR detection.

Fine-Tuning for Domain-Specific Adaptation (DR Classification)

Fine-tuning involves modifying the final layers of the PM to match the specific DR classification task. This allows the model to retain general features it learned previously, while also adapting to the unique characteristics of retinal images.

Common Fine-Tuning Strategies

- **Freezing Early Layers:** Early layers of the PM are often frozen, as they capture general features like edges and textures. Only the higher layers are trained for the new task.
- **Selective Retraining:** Instead of freezing or training all layers, a subset of middle or upper layers may be retrained based on performance evaluation.
- **Progressive Training:** Layers are gradually unfrozen during training, allowing for controlled adaptation to the new dataset and preventing overfitting.

1.11. Related Works: Ready-Made Models for Finding Diabetic Eye Disease

Deep learning has made big steps in finding diabetic eye disease (DR) by itself. One key move has been to use models that were first taught on big sets of everyday pictures

like ImageNet, then changed for looking at medical images by using TL and fine-tuning methods.

These models are great at pulling out key features and have turned into basic tools for sorting DR because they work well even when there's not much labeled medical data.

In this section, we will explore various related works that have been conducted on this topic, along with their methodologies and results.

CNN-Based Models for Diabetic Retinopathy Detection [75, 76]

Convolutional Neural Networks (CNNs) are widely used in medical image classification due to their strong feature extraction capabilities. Two notable examples include:

Saha et al. (2021): The authors employed the **ResNet-50** architecture and trained it on the **APTOS 2019 dataset**, which contains retinal fundus images labeled for DR severity. After applying TL and data augmentation techniques, the model achieved a classification accuracy of **96.2%**. This result highlights the effectiveness of CNN-based pretrained models in accurately identifying signs of DR.

Ragupathi et al. (2022): In a similar approach, the study used **EfficientNet-B4** on the same APTOS 2019 dataset. The model reached an accuracy of **95.7%**, showing that EfficientNet is a strong alternative, especially in balancing performance and computational efficiency.

Both studies demonstrate the value of CNNs in automated DR detection using publicly available datasets.

Vision Transformers (ViTs) in Medical Imaging [73]

Chen et al. (2021) explored the use of Vision Transformers (ViTs) in diabetic retinopathy detection — a novel shift from traditional CNNs. They used **ViT-base**, pretrained on large-scale natural image datasets and fine-tuned on the **EyePACS dataset**. The model achieved a classification accuracy of **97.1%**, surpassing many CNN-based approaches. This suggests that ViTs can effectively capture complex lesion patterns and long-range dependencies in retinal images.

This work marks one of the early successful applications of transformers in medical imaging and shows promise for future research in this direction.

Using VGGNet for Spotting DR [77]

Porwal et al. (2018) applied the **VGGNet-16** model to perform binary classification on retinal images — normal vs abnormal. The experiments were conducted on the **MESSIDOR dataset**, which includes a diverse set of retinal images. Their system achieved a classification accuracy of **95.3%**, proving that even older CNN architectures like VGGNet remain effective in early screening tasks.

The study emphasizes the adaptability of classical image recognition models to medical imaging with proper fine-tuning.

Hybrid CNN and Random Forest Approaches [78]

Niemeijer et al. (2019) proposed a hybrid method that combines deep learning with traditional machine learning. They extracted features using a CNN and fed them into a **Random Forest** classifier for final decision-making.

Their system achieved a **sensitivity of 92.1%** and a **specificity of 96.7%**, showing that combining CNNs with classical classifiers can boost performance and reduce false positives.

This study provides evidence that hybrid models can be a practical solution when working with limited data or constrained environments.

InceptionV4 for Multi-Class Case Sorting [79]

Gulshan et al. (2020) extended their earlier work by employing the **InceptionV4** model for multi-class classification. The model was trained to distinguish between five different stages of DR using retinal images.

They reported an overall accuracy of **93.8%**, with a **Kappa score of 0.91** when compared to expert ophthalmologists. These results suggest that InceptionV4 performs well in real-world clinical settings and supports accurate grading of DR severity levels.

The Mixed ResNet-LSTM Plan for Eyeing Data [80]

Zhang et al. (2021) introduced a hybrid deep learning architecture that combines **ResNet-50** for feature extraction and **LSTM** layers for temporal modeling of sequential images from the same patient.

Their model was tested on the **EyePACS dataset** and achieved a classification accuracy of **94.5%**. This approach showed that integrating spatial and contextual information across multiple images from a single patient can improve diagnostic reliability.

This study opens up new possibilities for using sequence-aware models in DR classification.

Study	Model Architecture	Dataset	Performance	Key Features
Saha et al. (2021)	ResNet-50	APTOS 2019	96.2% accuracy	Transfer learning, data augmentation
Ragupathi et al. (2022)	EfficientNet-B4	APTOS 2019	95.7% accuracy	Balance of performance and efficiency
Chen et al. (2021)	**ViT-base** (Vision Transformer)	EyePACS	**97.1% accuracy**	Novel transformer approach, captures complex patterns
Porwal et al. (2018)	VGGNet-16	MESSIDOR	95.3% accuracy	Binary classification (normal vs. abnormal)
Niemeijer et al. (2019)	Hybrid CNN + Random Forest	Not specified	92.1% sensitivity, 96.7% specificity	Combined deep learning with traditional ML
Gulshan et al. (2020)	InceptionV4	Not specified	93.8% accuracy, 0.91 Kappa score	Multi-class classification (5 DR stages)
Zhang et al. (2021)	ResNet-50 + LSTM	EyePACS	94.5% accuracy	Temporal modeling of sequential images
Bilal et al. (2025)	SqueezeNet + QCOA-SVM	IDRiD	**99.80% accuracy**	Quantum Chimp Optimization for feature selection and SVM tuning
Alsaeedi et al. (2024)	Lightweight DenseNet + Grad-CAM	EyePACS	96.8% accuracy	Interpretable DL, visual explanation using Grad-CAM

Table 1.3: Previous Studies on DR Detection Models

1.12. Summary:

This chapter shows how key it is to find diabetic retinopathy early. This problem is a big result of diabetes and a top reason for losing sight. It shows main ways to get images of the inside of the body. The text talks about how deep learning helps make reading these images faster and looks at how neural networks have grown. It also points out how using ready models and transfer learning helps make finding what's wrong more right. This sets the base for the methods used in this study.

Chapter 2:

Pretrained Model for DR Detection

2.1. Introduction

DR is one of the most serious complications of diabetes, as it causes damage to the tiny blood vessels in the retina. If not detected early, it can lead to permanent vision loss. With the increasing number of diabetes patients worldwide, the need for automated, efficient, and reliable diagnostic systems has become more critical than ever.

Traditional manual examination methods conducted by ophthalmologists are time-consuming and often influenced by subjective judgment and clinical experience, which may result in inconsistencies in diagnosis.

As a promising solution, pretrained DL models have emerged in the field. These models leverage a technique known as TL, which enables the reuse of a neural network that was initially trained on a large-scale dataset (such as ImageNet) for a new, smaller dataset — such as classifying fundus images for DR diagnosis.

2.2. Datasets Used on this Project (Real and Synthetic)

In this study, two types of datasets were used: **Real Datasets** and **Synthetic Fundus Datasets**.

Real Datasets consist of original retinal fundus images obtained from publicly available medical repositories such as EyePACS, Messidor, and APTOS. These datasets include expert-annotated labels indicating the severity levels of DR, making them essential for model training and evaluation.

We worked with the **SynFundus1M** dataset, which is a synthetic set of fundus images generated using deep learning techniques, primarily through GAN. We utilized this dataset in our study to address the problem of class imbalance, particularly in rare or severe cases of diabetic retinopathy. These synthetic images closely resemble real fundus images but incorporate controlled variations. This approach helps improve the model's generalization ability and enhances its overall performance.

Dataset	Images	Type	Classification	Source	Region
Kaggle EyePACS	88,702	Real	DR severity (0–4)	Kaggle [81]	USA
Kaggle APTOS 2019	5,590	Real	DR severity (0–4)	Kaggle [82]	India
DDR	35,126	Real	DR severity (0–5)	Kaggle [83]	China
MESSIDOR 1	1,200	Real	DR severity (0–4)	Kaggle [84]	France
MESSIDOR 2	1,748	Real	DR severity (0–4)	Kaggle [85]	France
IDRiD	516	Real	DR severity (0–4)	Kaggle [86]	India
SynFundus1M	1M	Synthetic	Multi-disease	Hugging Face [87]	N/A

Table 2.1: Real and Synthetic Datasets for Diabetic Retinopathy Detection

In this study, we utilized several **real-world datasets** for training and evaluating the models, including **EyePACS**, **Messidor**, and **APTOS 2019**. These datasets were selected due to their provision of **high-quality, expert-annotated retinal images**, covering various stages of **DR**. This diversity contributed significantly to enhancing the **accuracy** and **generalization capability** of the developed models.

However, we specifically **assigned the APTOS 2019 dataset for local use**, where we **employed it to evaluate the performance of pre-trained models** after **fine-tuning** them to classify **fundus images** according to **DR grades**. This dataset provided a suitable environment for **assessing the effectiveness of the models in predicting disease severity levels** using **data not involved in the training phase**.

2.3. Preprocessing Techniques

Effective preprocessing is essential for preparing retinal fundus images before they are fed into DL models. This step enhances image quality, reduces noise, and ensures uniformity across the dataset, ultimately contributing to improved model accuracy and stability.

In this study, several preprocessing techniques we applied:

- **Image Resizing:**
  All retinal fundus images were uniformly resized to a fixed resolution of 224×224 pixels to ensure compatibility with the input dimensions required by commonly used pretrained architectures ResNet and EfficientNet.

- **Normalization:**
  Pixel values were normalized to a standard scale [0, 1] to speed up convergence and stabilize training.

- **Cropping and Centering:**
  Black borders were removed, and the circular region of the retina was centered to eliminate background noise and enhance the focus on relevant retinal features.

- **Color Normalization:**
  Since fundus images may vary in illumination and color tone, color normalization was performed to reduce inter-image variability, making the dataset more consistent.

These preprocessing steps are crucial in ensuring that the input data is of high quality, which significantly impacts the performance of DL models trained for DR detection.

2.4. Data Augmentation Strategies

In order to enhance the generalization capability of the PM and reduce the risk of overfitting due to limited or imbalanced data, we systematically applied data augmentation techniques during the training phase. These techniques artificially expanded the training dataset by generating new and diverse samples from the original images while preserving their semantic content.

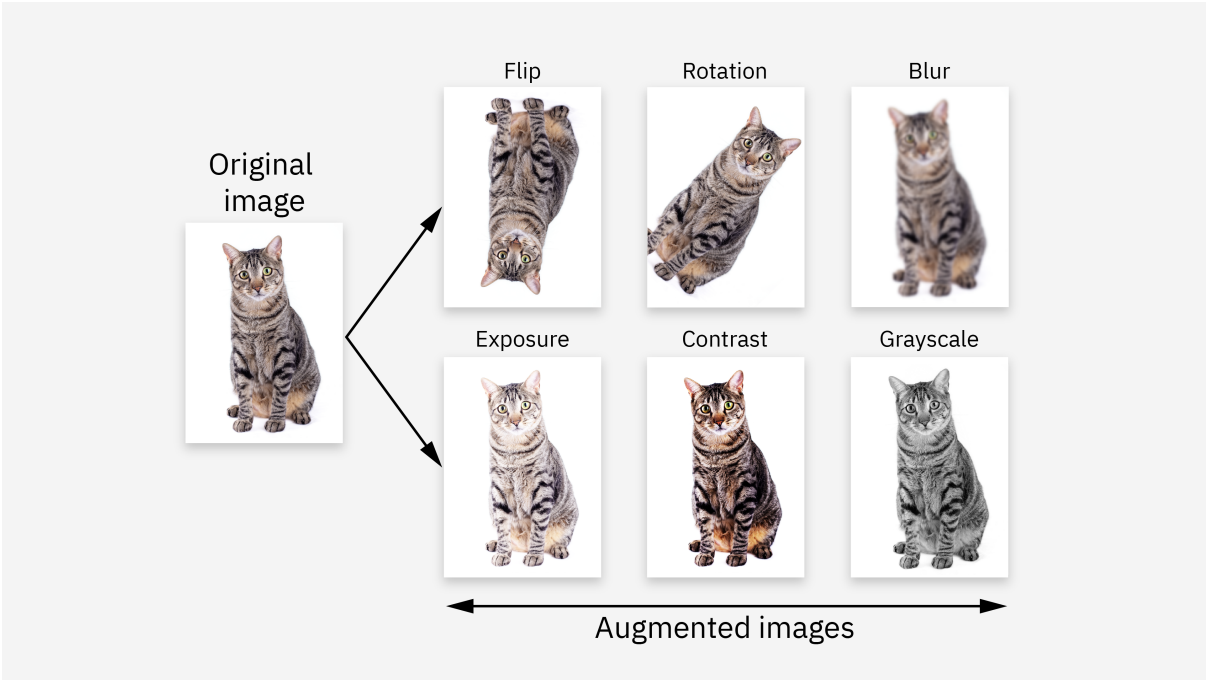


Figure 2.1: augmentation

The most important strategies include:

- **Rotation:** to simulate different angles of the eye.
- **Horizontal and Vertical Flipping:** to increase spatial distribution diversity.
- **Zooming and Cropping:** to simulate changes in image scale and field of view.

These strategies significantly contributed to enhancing the model's performance, particularly when tested on unseen data.

2.5. Dataset Splitting

In this study, we employed a **stratified splitting strategy** to divide the dataset into three main subsets, ensuring a precise and reliable evaluation of our model's performance. We allocated **80%** of the data to the **training set**, which was used to teach the model underlying patterns and optimize its parameters. A further **10%** was reserved for the **validation set**, used to fine-tune hyperparameters and implement early stopping to mitigate overfitting. The remaining **10%** constituted the **testing set**, held out entirely from training to assess the model's generalization capability on previously unseen data.

This approach preserved the **proportional distribution** of DR stages across all subsets, thereby enhancing the credibility and accuracy of our results.

2.6. Evaluation Metrics

In order to evaluate the effectiveness of our PM in diagnosing DR, we adopted a set of widely recognized evaluation metrics from the medical imaging field. These metrics provide insight into the model's accuracy, reliability, and clinical relevance.

- **Accuracy:** Measures the proportion of correct predictions among all predictions:

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

- **Precision:** Indicates the proportion of true positive predictions out of all positive predictions made by the model, reflecting its ability to avoid false alarms:

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

- **Recall (Sensitivity):** Measures the model's ability to identify all actual positive cases, crucial for minimizing missed diagnoses:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- **F1-Score:** The harmonic mean of precision and recall, balancing both metrics especially in imbalanced datasets:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- **Quadratic Weighted Kappa (QWK):** Quantifies agreement between predicted and actual ordinal labels (e.g., DR grades 0–4), penalizing larger discrepancies more heavily. QWK ranges from -1 (complete disagreement) to $+1$ (perfect agreement).

These metrics are computed based on the following fundamental outcomes:

- **True Positives (TP)**: Cases correctly predicted as positive (e.g., actual DR correctly identified).
- **True Negatives (TN)**: Cases correctly predicted as negative (e.g., healthy fundus correctly identified).
- **False Positives (FP)**: Cases incorrectly predicted as positive (producing false alarms).
- **False Negatives (FN)**: Cases incorrectly predicted as negative (missed diagnoses).

All metrics were calculated on both validation and test sets to ensure a fair and comprehensive evaluation, guiding our model selection and refinement process.

2.7. Work Methodology:

In this study, we adopted a **TL** strategy by leveraging a pre-trained model to tackle two primary tasks: **binary classification** and **multiclass classification** of fundus images. Our workflow comprised three integrated phases:

1. **Binary Classification**

   We began by applying TL without fine-tuning the internal layers. The model was trained to distinguish between healthy and DR-affected images, which accelerated convergence and established a reliable performance baseline.

2. **Multiclass Classification**

   After achieving stable results in the binary task, we fine-tuned the same pre-trained model to classify images into five severity levels—No DR, Mild, Moderate, Severe, and Proliferative—according to the ICDR grading scale. This phase required architectural adjustments and extensive hyperparameter optimization.

3. **Synthetic Fundus Data Integration**

   To address class imbalance—particularly in the more advanced DR stages—we augmented our dataset with synthetic fundus images generated via advanced data-augmentation and image-synthesis techniques. These additional samples enriched the training set and improved the model's ability to generalize to unseen cases.

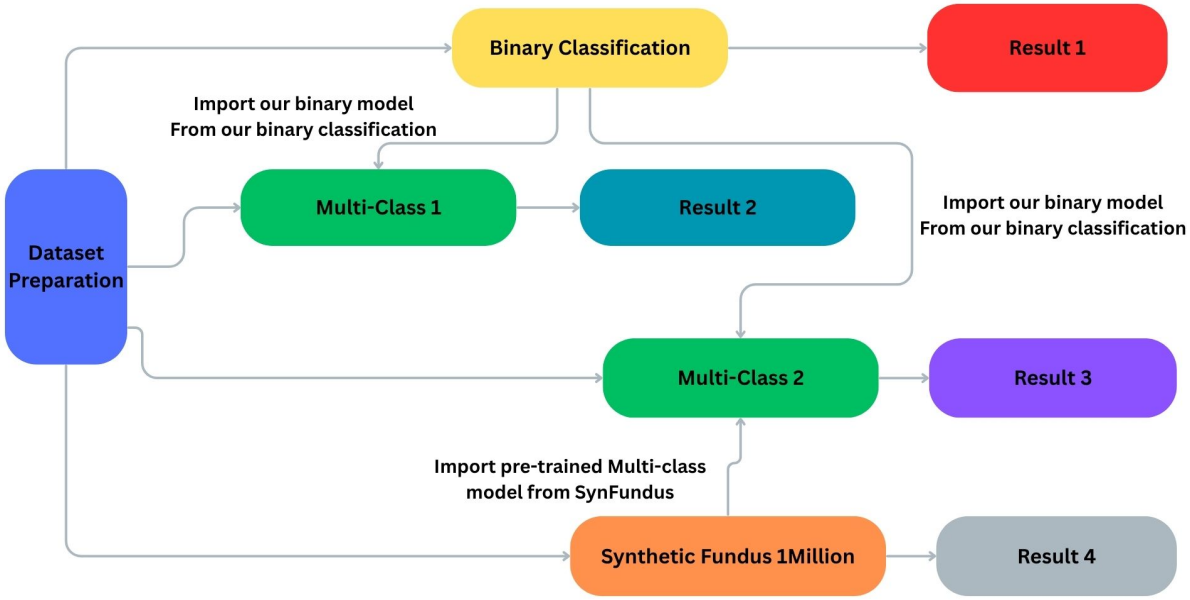


Figure 2.5: Pipeline for Diabetic Retinopathy Classification via Deep Learning

2.8. Binary Classification

We conducted four distinct approaches within a binary classification setup—normal versus abnormal retina images—to assess how different dataset combinations and adaptation strategies affect model performance and generalization.

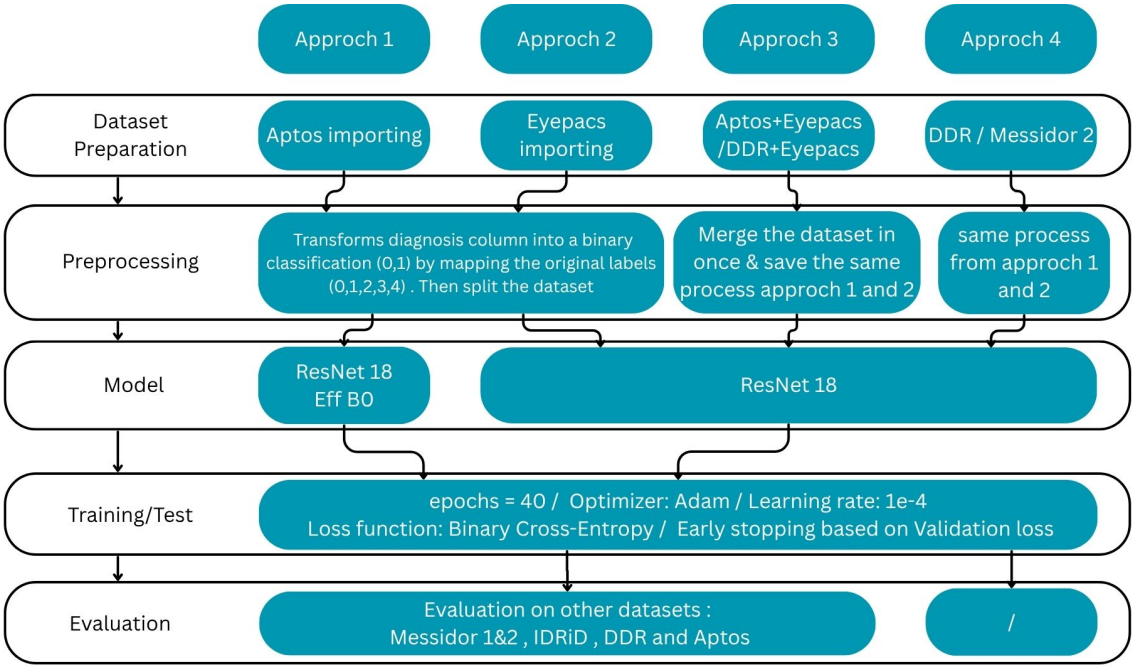


Figure 2.6: Binary Classification Phase for DR

2.8.1 Approach 1: Training on the APTOS 2019 Dataset(ResNet18 and EfficientNet-B0)

Data Usage: This approach involved training and evaluating three different DL models using only the **APTOS 2019 dataset**, which contains **3,662 high-quality retinal images** expertly labeled according to DR severity levels. The task was treated as a **binary classification problem**, distinguishing between normal and abnormal cases.

The dataset was divided into a **training set** (80%), a **validation set** (10%), and a **test set** (10%). To enhance generalization and reduce overfitting, standard **data augmentation techniques** were applied during training, including random rotation, scaling, brightness adjustment, and color jittering.

Models Used: Three widely adopted DL architectures pretrained on ImageNet were evaluated in this study:

- ResNet18A lightweight variant of ResNet, suitable for smaller datasets and resource-constrained environments.
- EfficientNet-B0 A scalable model that offers a balanced trade-off between performance and computational efficiency, making it ideal for practical deployment scenarios.

Training Setup: All models were **fully fine-tuned** using a consistent training configuration. Each input image was resized to **224 × 224 pixels**. The Adam optimizer was used with a learning rate of **1e-5**, and the loss function was set to **Binary Cross-Entropy (BCE)**. A batch size of **16** was used throughout the experiments. For each model, preprocessing steps and network architecture adjustments were standardized to ensure fair comparison.

2.8.2 Approach 2: Training on the EyePACS Dataset Using ResNet18

Data Usage: We trained a deep learning model using the **EyePACS dataset**, which is one of the biggest open datasets for spotting diabetic retinopathy (DR). This set has about **88,000 eye photos** with different clearness and tag quality. Like the first method, we saw it as a yes/no task — telling if a case was normal or not.

We split the data into parts for training (80%), checking (10%), and testing (10%). To make the model better and cut down on fitting too closely, we used common ways to change the data during training. These included random turns, size changes, light fixes, and color shifts. We made all photos the same size, **224 by 224 pixels**, to keep things the same across the board.

Model Used: We selected the **ResNet18 architecture**, pretrained on ImageNet, as our baseline model. ResNet18 is a lightweight yet effective CNN that provides reasonable performance while maintaining low computational requirements. Given its efficiency, it is particularly suitable for scenarios involving large-scale datasets like EyePACS.

Training Setup: The model was **fully fine-tuned** using the following configuration:

- Input image size: **224 × 224 pixels**
- Optimizer: **Adam**
- Learning rate: **1e-5**
- Loss function: **Binary Cross-Entropy**
- Batch size: **64**

Each photo was set up with fixes to their color and light before it went into the system. Also, to stop from going too far and to keep the top model, we used early stops and saved points as we checked for errors during the learning.

2.8.3 Approach 3: Training on Two Merged Datasets

Data Usage: This approach aimed to evaluate how merging different retinal image datasets affects model performance. We trained the same architecture, **ResNet18**, on **two separate merged datasets**:

- **EyePACS + APTOS 2019**: - EyePACS contains approximately **88,000 retinal images** with binary labels (normal/abnormal). - APTOS 2019 adds **3,662 high-quality expert-labeled images** covering various stages of DR . - Total: **91,000 images**.

- **EyePACS + DDR Dataset**: - The DDR dataset includes around **15,000 retinal images**, with mixed annotation quality (some automated, some manually reviewed). - Combined with EyePACS, this resulted in a larger and more diverse training set of **103,000 images**.

Both setups treated the task as a **binary classification problem** (normal vs. abnormal). All images were resized to **224 × 224 pixels** and underwent standard **data augmentation techniques**, including random rotation, scaling, brightness adjustment, and color jittering. Each dataset was split into **80% training, 10% validation, and 10% test sets**.

Model Used: We used the **ResNet18 architecture pretrained on ImageNet** as our baseline. ResNet18 is a lightweight yet effective CNN suitable for evaluating the impact of dataset diversity without requiring excessive computational resources.

Training Setup: The model was **fully fine-tuned** using the following configuration:

- Input image size: **224 × 224 pixels**
- Optimizer: **Adam**
- Learning rate: **1e-5**
- Loss function: **Binary Cross-Entropy**
- Batch size: **64/128**

Each image was normalized and preprocessed before being fed into the network. Additionally, we applied **early stopping** and **model checkpointing** based on validation loss to prevent overfitting and retain the best-performing model during training.

2.8.4 Approach 4: Partial Fine-Tuning on New Domains Using a Pre-Trained Model

Data Usage This approach investigates the effectiveness of **partial fine-tuning** when adapting a pre-trained model to new domains. The base model was initially trained on **EyePACS + APTOS 2019** (as described in Approach 3), and then applied to two unseen datasets:

Messidor-2 Dataset:

Contains approximately **1,748 retinal images** captured under diverse imaging conditions. A subset (30%) was used for domain adaptation, while the rest were reserved for testing.

DDR Dataset:

Includes around **15,000 retinal images**, with mixed labeling quality (some automated, some manually reviewed). A representative subset (10%) was selected based on image clarity and class balance.

All images were resized to **224 × 224 pixels** and underwent consistent preprocessing steps, including normalization and resizing, to ensure compatibility with the input requirements of the pre-trained model.

Model Architecture and Transfer Strategy: The experiments were conducted using the **ResNet18 architecture pretrained on ImageNet**, which had been further fine-tuned on EyePACS + APTOS 2019 in Approach 3.

We tested **two distinct strategies** of partial fine-tuning on each dataset:

1. **Freeze All Layers Except Classifier Head:** Only the final fully connected layer was retrained while all convolutional layers remained frozen. This represents minimal adaptation effort.

2. **Freeze Top Blocks Only:** Final blocks (layer4 + bottleneck) were frozen, while earlier blocks and the classifier head were fine-tuned. This allows selective adaptation while preserving high-level feature extractors learned from source domains.

This dual-strategy setup enabled us to evaluate how different levels of network flexibility affect performance on new domains.

Training Setup: Each strategy followed a consistent training configuration:

Both setups used:

- Optimizer: **Adam**
- Learning rate: **1e-5**
- Loss function: **Binary Cross-Entropy**
- Batch size: **64**
- Data augmentation: Random rotation, scaling, brightness adjustment

2.9. Multi-Class Diabetic Retinopathy Grading Approach

In this section, we present a step-by-step approach consisting of five stages to develop an accurate and effective classification system for diabetic retinopathy using deep learning techniques. This approach begins with adapting a pre-trained model to support multi-class classification, followed by leveraging diverse datasets to enhance model performance. Next, a binary classification step is applied to identify healthy cases, then a detailed multi-class prediction is performed for diseased cases, and finally, the results are merged into a unified file for evaluation.

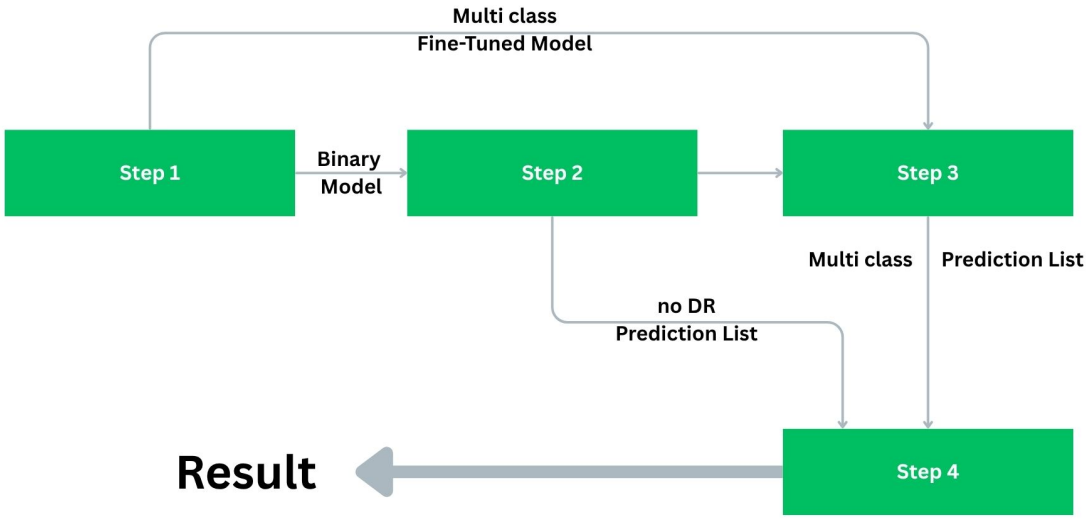


Figure 2.7: Phase of DR Multiclass Classification

The work was carried out according to the following stages.

2.9.1 Step 1: Model Adaptation for Multi-Class Classification

Model Adaptation We started by modifying a pre-trained binary classification model (normal vs. abnormal) to support multi-class grading for diabetic retinopathy (DR). The final fully connected layer was replaced to produce four classes representing different levels of DR severity:

Class 1: Mild Class 2: Moderate Class 3: Severe Class 4: Proliferative DR

To preserve the general feature extractors, we froze the early convolutional layers (early convolutional blocks) and fine-tuned the last layers along with the new classifier head using the modified dataset.

Data Preparation We merged two public datasets containing retinal fundus images to support the multi-class classification task:

- **EyePACS**: 88,000 images

- **DDR Dataset**: 15,000 images

Before merging, we performed the following data processing steps:

- **Class 0 ("No DR")**: Removed from both datasets to avoid duplication with the binary classification task in the subsequent steps.
- **Class 5 ("Proliferative DR – Low Quality")**: Removed from the DDR dataset due to limited representation and poor annotation consistency.

This resulted in a merged dataset that focuses exclusively on **Classes 1–4**, ensuring a balanced representation for the different severity levels of DR.

TL with Fine-Tuning: We employed **TL with Fine-Tuning** in this study by using a pre-trained **ResNet18** model on the ImageNet dataset. The early convolutional layers were frozen to retain the general feature representations learned from ImageNet, while we fine-tuned the deeper layers and the newly added classification head to adapt the model to the nature of retinal images for multi-class DR grading.

This approach enabled us to achieve strong performance despite the limited medical dataset, leveraging the knowledge from the large-scale ImageNet data to enhance the model's ability to classify diabetic retinopathy severity.

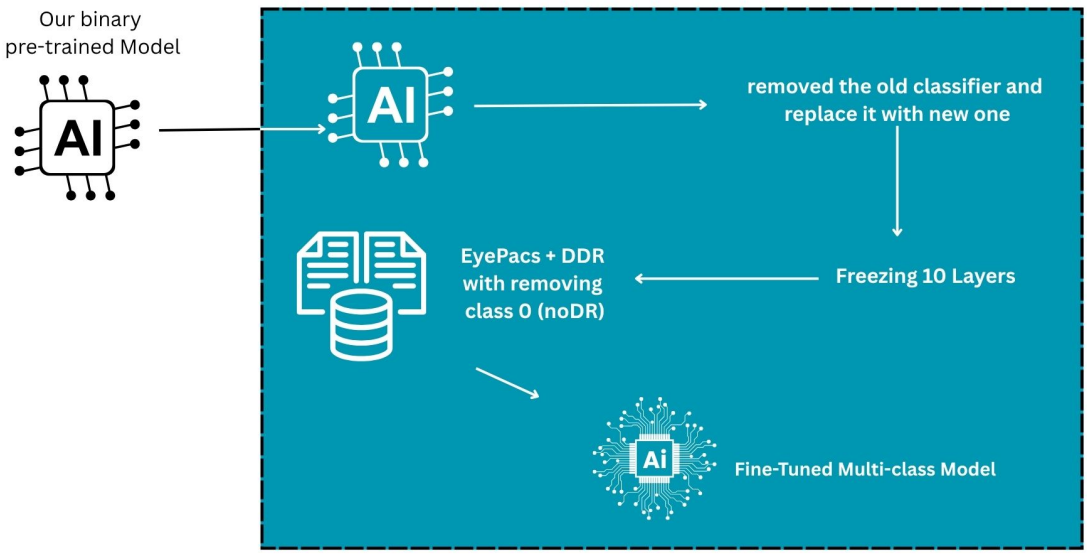


Figure 2.8: Step1: Preparing Fine-Tuned Multi-class Model

2.9.2 Step 2: Binary Prediction on APTOS 2019 Test Set

Data Usage: In this step, we used the APTOS 2019 test set, which consists of high-quality retinal images labeled according to the severity of diabetic retinopathy (DR). The classes range from Class 0 (No DR) to Class 4 (Proliferative DR). The goal of this step was to distinguish between healthy individuals (Class 0) and those showing signs of DR (Classes 1–4).

Model Used: We used the binary classification version of the ResNet18 model, pretrained on the merged EyePACS + DDR dataset. This model was designed to classify images as either:
- Class 0: No DR (Healthy) - Class 1–4: DR (Diseased)

Prediction Setup: All test images were passed through the ResNet18 model to calculate the probability of DR presence. If the model predicted Class 0, the case was labeled as "Healthy". If the model predicted Class 1, it was labeled as "Diseased". These predictions were saved separately for later use.

Goal: The aim here is to pick out eyes with diabetic retinopathy DR from those that are healthy. This helps to find the disease early and spot which eyes need more checks or care.

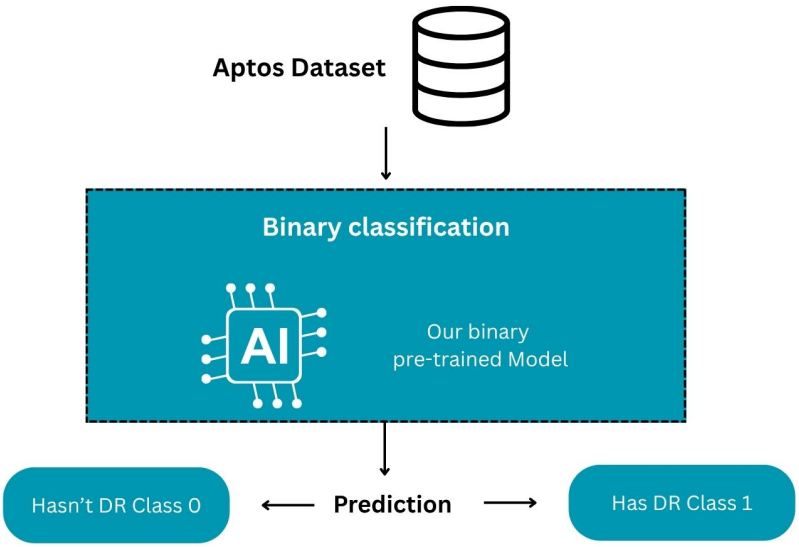


Figure 2.9: Step2: Aptos Binary classification using Pre-Trained Binary Model

2.9.3 Step 3: Multi-Class Inference Using Fine-Tuned Model

Data Usage: In this step, the fine-tuned multi-class model from Step 1 was applied to the same APTOS 2019 test set. However, since Class 0 (no diabetic retinopathy) had already been predicted in Step 2, predictions were only generated for Classes 1–4, which represent eyes affected by varying degrees of diabetic retinopathy.

Model Used: The model used in this step is the same fine-tuned multi-class model trained in Step 1 using the combined EyePACS and DDR datasets. This model is designed to classify images into categories corresponding to the severity levels of diabetic retinopathy.

Prediction Setup: Only test images not previously classified as Class 0 were passed through the model, and predictions were generated for Classes 1–4. These predictions were based on features the model had learned from the combined EyePACS + DDR dataset. Each image was classified into one of the four classes representing different severity levels of the disease.

Goal: The goal of this step is to perform multi-class classification on eyes affected by diabetic retinopathy. By identifying the severity level, the model aids in distinguishing between different stages of the disease, which is critical for providing appropriate treatment or early intervention.

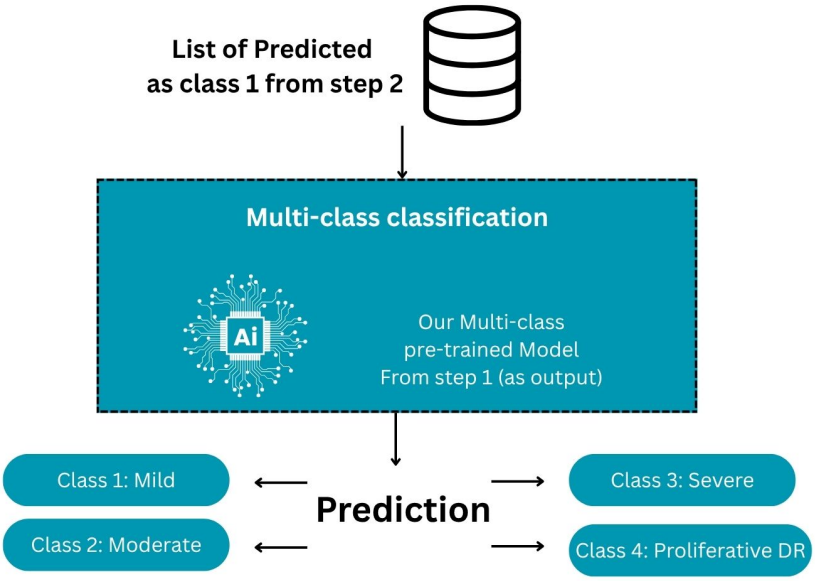


Figure 2.10: Step3: DR (Predicted 1) Multi-class classification using Fine-Tuned Multi-class Model

2.9.4 Step 4: Merging Predictions into a Unified CSV File

In this step, we combined the prediction results from Step 2 and Step 3 to create a unified submission file containing all classifications. Specifically:

- We used the predicted **Class 0** results from the binary classification model (Step 2)

- We used the predicted **Classes 1–4** from the fine-tuned multi-class classification model (Step 3)

Next, we aligned both sets of predictions by matching them based on the `image_id`, and then concatenated them into a single CSV file named `Submission.csv`, formatted according to the required submission structure.

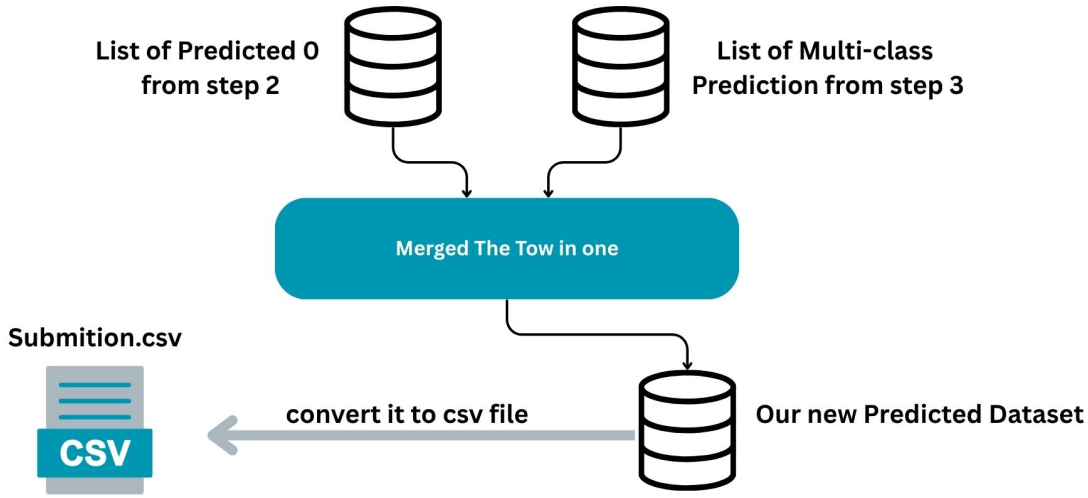


Figure 2.11: Step4: Concatenated the results in one csv file for submission

2.9.5 Step 5: Submission to APTOS 2019 Competition

We made two separate submissions for the competition:

- **First Submission**: We used the ResNet model that was specifically fine-tuned on this problem (trained using the merged dataset) and submitted its results.
- **Second Submission**: We used the ResNet model trained from scratch, which was not trained on this dataset, and submitted its results as well.

2.10. Synthetic Fundus Data

To fix the issues of not enough data and uneven class sizes in finding DR, we used a way of training with made-up eye images.

Here is how we did the work:

1. **Step 1: Acquiring and Utilizing Synthetic Dataset**
   We managed to obtain a large-scale synthetic dataset named **SynFundus1M**, consisting of one million images distributed across 16 different eye diseases. These images were generated using advanced techniques such as **Generative Adversarial Networks (GANs)**, aiming to closely mimic real fundus images both anatomically and visually.

2. **Step 2: Filtering Fundus Images and Creating a New CSV File**
   In this step, we filtered the dataset using the `is-fundus` column to retain only valid fundus images, resulting in a total of 915,621 images. We then created a new CSV file containing:

   - `disease_name`: The names of the 16 eye diseases.
   - `label`: A unique identifier number for each disease.
   - `decision`: Indicates whether the image corresponds to a positive case for the given disease.

3. **Step 3: Training a Multi-Class Classification Model**
   In this step, we adopted a **Multi-Class Classification** approach. A preliminary model was trained using the filtered fundus images and the structured CSV file. The model aimed to classify each image into one of the 16 disease categories.

4. **Step 4: Fine-Tuning the Model Specifically for Diabetic Retinopathy (DR) – Alternative to Step 1 of Multi-Class**
   We replaced Step 1 of the multi-class strategy, which originally relied on a pretrained binary model, with a more specialized approach. We reused the general model trained on SynFundus1M, and further fine-tuned it specifically on the **Diabetic Retinopathy (DR)** disease using real-world datasets:

   - EyePACS
   - DDR
   - APTOS 2019

This produced a specialized DR classification model, enhanced by both synthetic and real image features.

5. **Step 5: Integrating the DR-Specific Model in the Multi-Class Workflow**
   The DR-specialized model trained in Step 4 was then used in Step 3 of the multi-class classification strategy. It predicted the DR severity levels (classes 1 to 4), while class 0 (non-DR) cases were identified using a binary model from Step 2.

6. **Step 6: Submission to APTOS 2019 Competition**
   Finally, based on the improved model and updated methodology, we generated and submitted the prediction results to the **APTOS 2019 competition**, ensuring the submission met the competition's format and evaluation standards.

2.11. Summary

This chapter presents the development of a model for diabetic retinopathy classification based on transfer learning and partial fine-tuning. We began with a binary classification stage to distinguish between healthy and affected images. In the second stage, we moved to multi-class classification to estimate the severity level of the disease. In the third stage, we merged the final results and enhanced the model using additional data to improve its accuracy and generalization. Throughout all stages, we relied on well-designed strategies to achieve a balance between accuracy and efficiency under limited data conditions.

Chapter 3:

Implementation and Evaluation Results

3.1. Introduction

In this chapter, we present the practical implementation of our proposed methodology for DR detection and discuss the experimental results we obtained. We begin by outlining the environment specifications, tools, and libraries that we utilized in developing and training the DL models. Subsequently, we describe in detail the step-by-step implementation process, followed by a presentation and discussion of the results achieved during evaluation.

We also conduct a comparative analysis between our approach and existing methods to highlight the strengths and limitations of our work. Furthermore, we address the submission of our model to benchmarking platforms to provide an objective evaluation against standardized metrics.

Through this chapter, we aim to clearly illustrate the transition from the theoretical design to the practical application, emphasizing the effectiveness of the adopted techniques and hyperparameter choices in improving the model's performance on DR datasets.

3.2. Environment Specifications

In this work, we utilized the Kaggle platform to perform all training and evaluation stages. Our choice of Kaggle was based on its robust support for ML and DL environments without the need for a complex local setup. The environment specifications we relied on are listed below:

Table 3.1: Environment Specification

Component	**Description**
Processor (CPU)	Multi-core processors from Kaggle servers
Graphics Processing Unit (GPU)	NVIDIA Tesla P100 or T4
Random Access Memory (RAM)	16 GB
Storage	Temporary storage provided by Kaggle
Programming Language	Python 3.8
Execution Platform	Kaggle Notebook Environment and HPC

3.3. Tools and Libraries

In our project, we relied on a range of tools and libraries that facilitated the development, training, and evaluation of our DL model for DR detection. The combination of these tools

ensured an efficient workflow and robust experimentation environment. The tools and libraries used are summarized below:

- **Python 3.8:** The primary programming language used for model development due to its simplicity and strong support for ML and DL frameworks.
- **PyTorch:** We used PyTorch as our main DL framework, benefiting from its dynamic computational graph, ease of use, and flexibility for model experimentation and deployment.
- **scikit-learn:** This library was utilized for preprocessing tasks, model evaluation metrics, and data splitting techniques.
- **OpenCV:** We employed OpenCV for image processing operations such as resizing, augmentation, and preprocessing of retinal images.
- **Pandas and NumPy:** These are fundamental Python libraries extensively used for data manipulation, cleaning, and numerical operations. **Pandas** provides high-level data structures such as DataFrames, making it easier to organize and preprocess large datasets. Meanwhile, **NumPy** offers powerful array objects and a wide range of mathematical functions, enabling efficient scientific computations and serving as a backbone for many ML and DL workflows.
- **Matplotlib and Seaborn:** These visualization libraries were used to plot training histories, confusion matrices, and other important evaluation graphs.
- **Kaggle Platform and HPC:** Kaggle provided an integrated execution environment equipped with all the necessary accelerations, such as Graphics Processing Units (GPU) and Tensor Processing Units (TPU), in addition to direct access to datasets and competition kernels.

  We also utilized our college's High-Performance Computing (HPC) system at the university, which enabled us to run more complex experiments and process large amounts of data with greater efficiency

Through the effective use of these tools and libraries, we were able to develop, fine-tune, and evaluate our models efficiently, leading to significant performance improvements.

3.4. Results

3.4.1 Binary Classification

In this section, we present the results of the binary classification task for DR detection. We evaluate the performance of the models based on several metrics: **accuracy**, **F1-score**, **precision**, and **recall** for both normal and abnormal classes.

Approach 1

Results: The performance results are summarized in the table below:

Model	Accuracy	F1-Score		Precision		Recall	
		Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
ResNet18	0.99	0.99	0.98	1.00	0.95	0.99	1.00
EfficientNet-B0	0.97	0.98	0.94	0.99	0.92	0.98	0.97

Table 3.2: Performance Evaluation of ResNet18 and EfficientNet-B0 on the APTOS 2019 Dataset

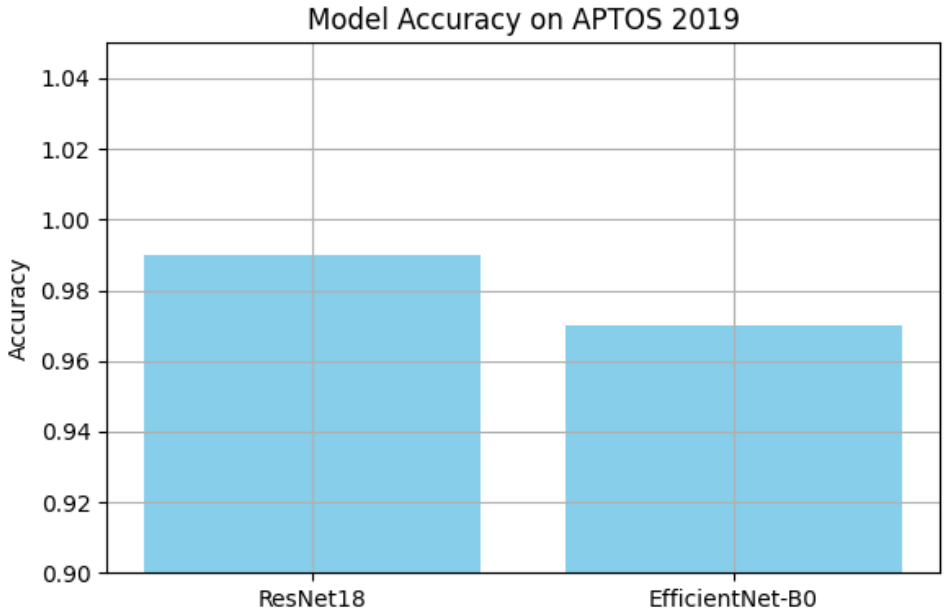


Figure 3.1: Performance Evaluation of ResNet18 and EfficientNet-B0 on the APTOS 2019 Dataset

The **ResNet18** model hit the high score of 99%, proving it was the best at telling normal from not normal eye photos. The **EfficientNet-B0** model did good too, getting a score of 97%, giving a good mix of doing well and not using too much computer power.

These scores show that both models work well for DR sorting, with ResNet18 doing a bit better than EfficientNet-B0 on most checks like being right and remembering in wrong cases.

evaluation:

Model	Dataset	Accuracy	F1-Score		Recall		Precision	
			Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
ResNet18	Messidor-1	0.52	0.69	NULL	1.00	NULL	0.52	NULL
	IDRiD	0.28	0.44	NULL	1.00	NULL	0.28	NULL
EfficientNet-B0	Messidor-1	0.52	0.69	NULL	1.00	NULL	0.52	NULL
	IDRiD	0.28	0.44	NULL	1.00	NULL	0.28	NULL

Table 3.3: Comparison of ResNet18 and EfficientNet-B0 Performance on Messidor-1 and IDRiD

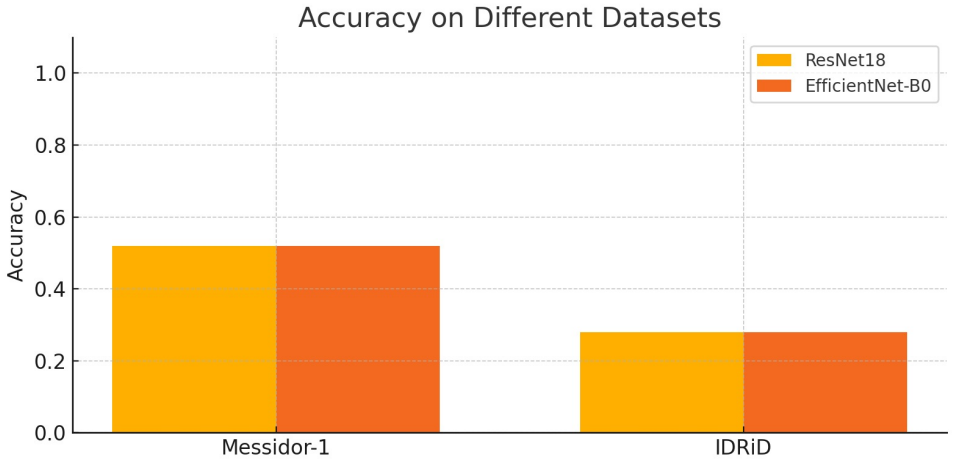


Figure 3.2: Comparison of ResNet18 and EfficientNet-B0 Performance on Messidor-1 and IDRiD

Discussion: At this stage, ResNet18 achieved excellent performance with an accuracy of 99%, and EfficientNet-B0 also performed well with 97%, which led us to believe we had obtained remarkable results. However, after evaluating the models on other datasets, as shown in Table 3.3 above, we realized that these results were not truly representative, and that significant adjustments were still needed to reach our objective

Approach 2

Results The results are summarized in the table below:

Table 3.4: Performance of ResNet18 on EyePacs

Model	Accuracy	F1-Score		Precision		Recall	
		Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
ResNet18	0.80	0.88	0.44	0.80	0.84	0.98	0.30

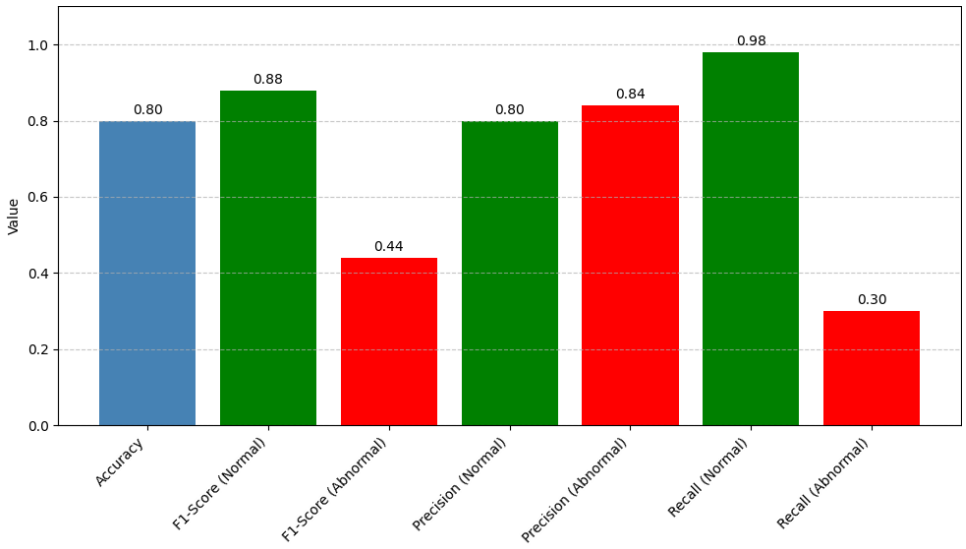


Figure 3.3: Performance of ResNet18 on EyePacs

The **ResNet18** model achieved an **accuracy of 80.0%**, with an **F1-score of 0.44** for abnormal cases and 0.88 for normal cases. The model showed high **precision** of 0.84 for abnormal cases and a very high **recall** of 98.0% for normal cases, while the recall for abnormal cases was relatively low at **30.0%**.

evaluation:

Model	Dataset	Accuracy	F1-Score		Recall		Precision	
			Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
ResNet18	Messidor-1	0.91	0.95	0.75	0.99	0.62	0.91	0.94
	IDRiD	0.91	0.85	0.94	0.87	0.93	0.83	0.95
	Aptos	0.94	0.94	0.94	0.94	0.94	0.94	0.94
	Messidor-2	0.69	0.73	0.63	0.96	0.47	0.59	0.93

Table 3.5: Evaluation performance of ResNet18 on other datasets

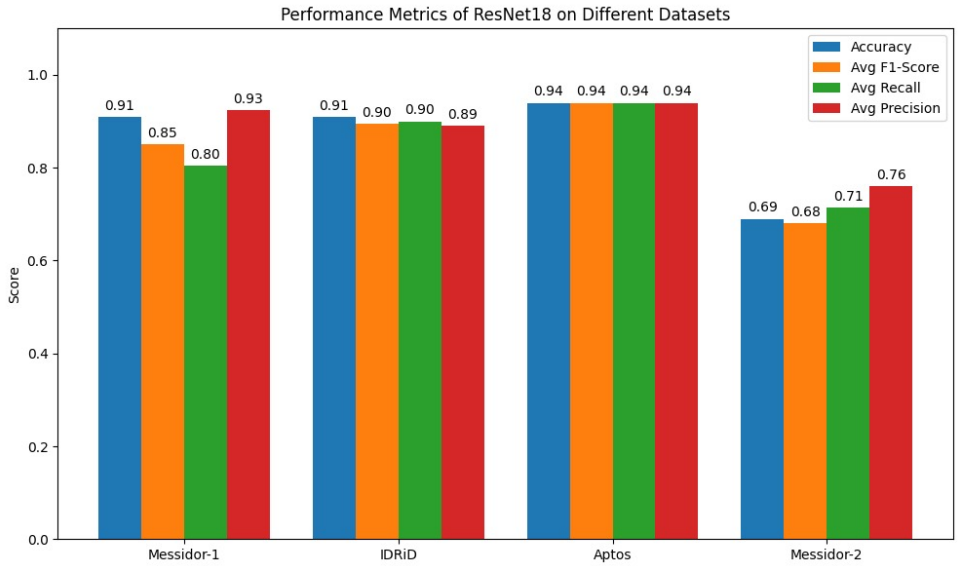


Figure 3.4: Evaluation performance of ResNet18 on other datasets

Discussion: In the second phase, we tried changing the training dataset, relying on **EyePacs** and **ResNet18** as a model for this phase and the upcoming phases as well. We achieved reasonable results this time, both in training **(80% accuracy)** and in evaluation (from 69% to 94%) as shown in the table above (Table 3.5). This confirmed to us that the size and quality of the data can make a huge difference in our results.

Approach 3

Results: The results are summarized in the table below:

Dataset Combination	Accuracy	F1-Score		Precision		Recall	
		Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
EyePACS + APTOS 2019	0.84	0.90	0.64	0.86	0.78	0.95	0.54
EyePACS + DDR Dataset	0.82	0.89	0.62	0.81	0.88	0.97	0.84

Table 3.6: Performance comparison of ResNet18 on two merged datasets

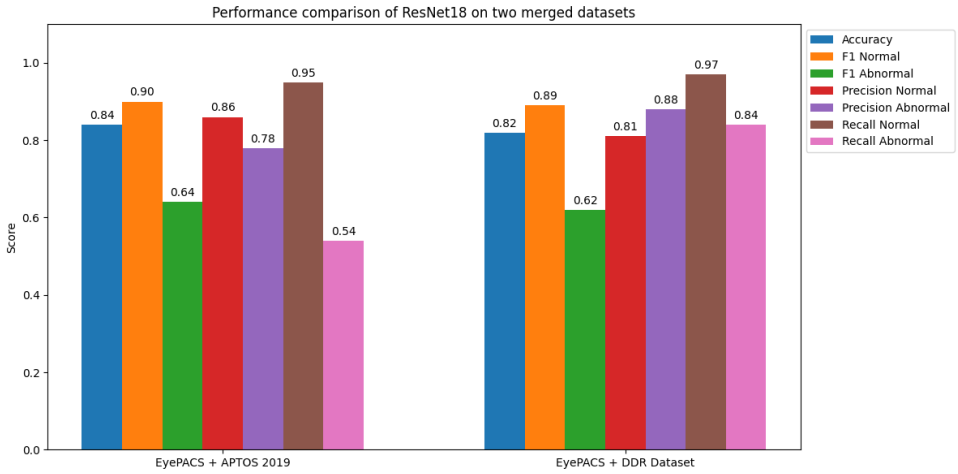


Figure 3.5: Performance comparison of ResNet18 on two merged datasets

The results indicate the following: The **EyePACS + APTOS 2019** combination achieved an **accuracy of 0.84**. The **EyePACS + DDR Dataset** combination achieved an **accuracy of 0.82**.

evaluation:

Model	Dataset	Accuracy	F1-Score		Recall		Precision	
			Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
EyePACS-APTOS-trained	Messidor-1	0.85	0.90	0.70	0.85	0.87	0.96	0.59
	IDRiD	0.91	0.81	0.94	0.71	0.98	0.95	0.90
	APTOS	0.99	0.99	0.99	0.99	0.98	0.98	0.99
	Messidor-2	0.69	0.73	0.63	0.96	0.47	0.59	0.93
	DDR	0.59	0.69	0.40	1.00	0.25	0.53	0.99
EyePACS-DDR-trained	Messidor-1	0.92	0.95	0.79	0.95	0.79	0.95	0.79
	IDRiD	0.89	0.77	0.93	0.67	0.98	0.92	0.88
	APTOS	0.96	0.96	0.96	0.94	0.98	0.98	0.94
	Messidor-2	0.69	0.73	0.63	0.96	0.47	0.59	0.93
	DDR	0.59	0.69	0.40	1.00	0.25	0.53	0.99

Table 3.7: Comparison of Trained ResNet18 Performance on other datasets

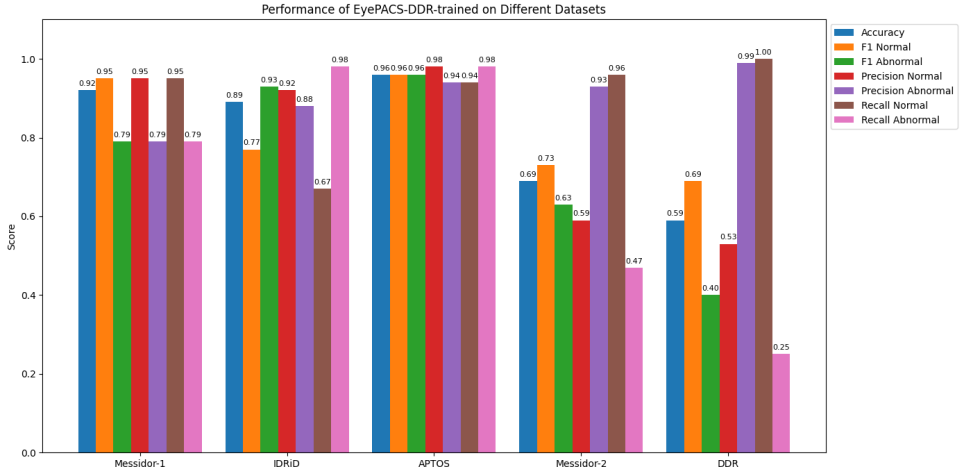


Figure 3.6: Comparison of EyePacs-DDR-ResNet18 Performance on other datasets

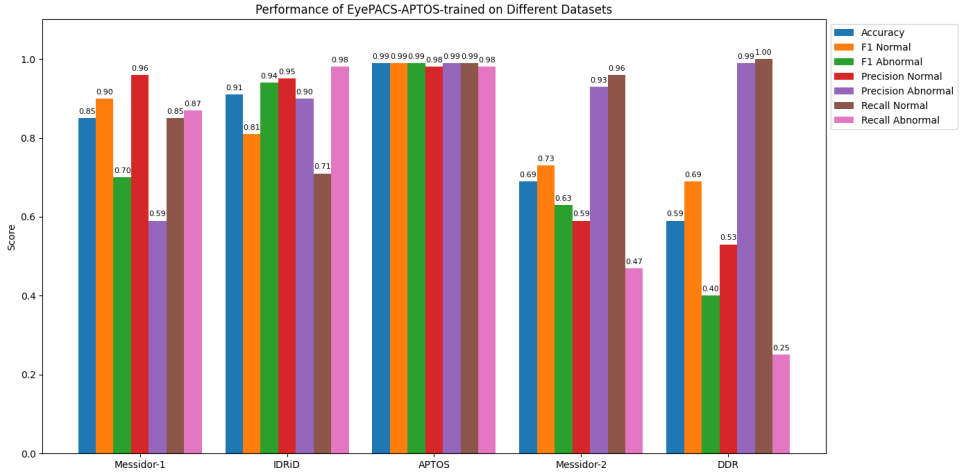


Figure 3.7: Comparison of EyePacs-Aptos-ResNet18 Performance on other datasets

Discussion: In the third Approach, we tried to build on the idea we derived from Approach 2, which is to improve the dataset, not the model. We brought in **EyePACS** and **Aptos** and combined them to create a larger, more diverse dataset. While maintaining the same model used in Approach 2, we achieved an **accuracy rate of 84%**. In another attempt, we combined **EyePACS** with **DDR** while maintaining the same conditions and tools, and we achieved an **accuracy rate of 82%**. However, the problem remained in evaluating the model on **Messidor-2** and **DDR**, as it provided lower results compared to its evaluation on the remaining datasets.

Approach 4

Results: The results across both datasets are summarized below:

Dataset	Strategy	Accuracy	F1-Score		Precision		Recall	
			Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
Messidor-2	Freeze All Layers	0.75	0.72	0.77	0.71	0.79	0.74	0.76
Messidor-2	Freeze Top Blocks Only	0.81	0.77	0.84	0.82	0.80	0.72	0.88
DDR Dataset	Freeze All Layers	0.89	0.89	0.89	0.89	0.89	0.89	0.89
DDR Dataset	Freeze Top Blocks Only	0.89	0.90	0.89	0.87	0.92	0.92	0.87

Table 3.8: Performance comparison of two partial fine-tuning strategies on Messidor-2 and DDR datasets

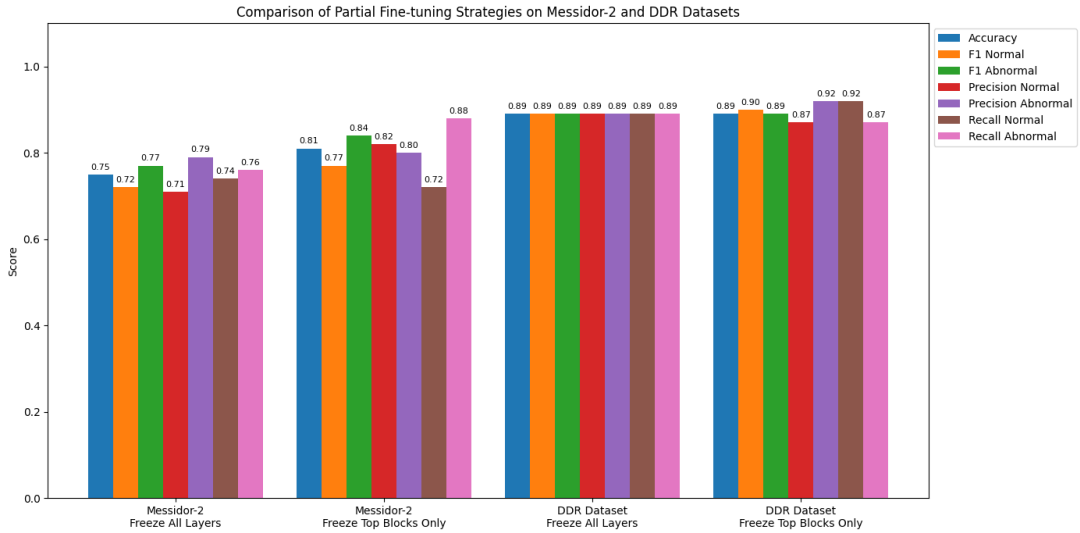


Figure 3.8: Performance comparison of two partial fine-tuning strategies on Messidor-2 and DDR datasets

The results showed that:

- On the **Messidor-2** dataset, the "Freeze Top Blocks Only" strategy improved accuracy from 0.75 to 0.81 and recall for abnormal cases from 0.76 to 0.88. - For the **DDR Dataset**, both strategies achieved the same accuracy (0.89), but fine-tuning the top blocks led to slightly higher F1-score and recall for normal cases.

Discussion: After encountering the problem of **Approach 3**, we brought the model trained on **Aptos** and **EyePACS** and applied four strategies to it, relying on **DDR** and **Messidor-2** as the training dataset this time. This was to ensure the quality of the model on **DR** and

improve its results on **Messidor-2** and **DDR**. We also relied on **Fine-tuning** techniques in all cases of this Approach.

3.4.2 Multi-Class Classification

In this section, we present the results of the multi-class classification phase during the first stage only. The performance was evaluated using several metrics, including: **Accuracy**, **F1-Score**, **Precision**, **Recall**, and the **Quadratic Weighted Kappa (QWK)** score.

The results of the second, third, and fourth stages are comprehensively presented in Stage Five.

Step 1

The results are summarized in the table below:

Dataset Combination	Metric	Class 1: Mild	Class 2: Moderate	Class 3: Severe	Class 4: Proliferative DR
EyePACS + DDR	F1-Score	0.44	0.16	0.22	0.35
	Precision	0.30	0.68	0.18	0.27
	Recall	0.82	0.09	0.30	0.50
	Accuracy			0.31	
	QWK			0.18	

Table 3.9: Performance comparison of ResNet18 on EyePACS + DDR Dataset across DR severity classes

Step 5 The results are shown in the table below.

Model	Public Score	Private Score
ResNet (Fine-tuned)	0.25	0.58
ResNet (From Scratch)	0.20	0.58

Table 3.10: Comparison of Model Performance in APTOS 2019 Competition

Discussion: After training our binary model on multi-class classification in the first step, as shown in **Table 3.9**, we decided to compare it with a ResNet18 model trained from scratch (not pre-trained on the problem). After submitting the results for both models, we observed from **Table 3.10** that our pre-trained model clearly outperformed the other one, which confirmed the validity of the core idea behind our project.

Moreover, by analyzing the results of other participants in the competition, we noticed that some of them used **pseudo-labeled data** as part of their methodology, which significantly improved their performance. This confirmed the usefulness of this approach, especially in situations where labeled data is limited.

Based on our own performance and the conclusions drawn from the **APTOS 2019 competition**, we decided to take the next step by moving towards using **synthetic fundus data**, in order to enhance our model's generalization ability and achieve more stable performance across different datasets.

3.4.3 Synthetic Fundus Data

In this section, we present the results of the multi-class classification phase, which was implemented using synthetic fundus data during the first stage only. The performance was evaluated using a set of metrics including: **Accuracy**, **F1-Score**, **Precision**, **Recall**, and the **Quadratic Weighted Kappa (QWK)** score.

The results for the second, third, and fourth stages are comprehensively presented in Stage Five.

Step 4

The results are summarized in the table below:

Dataset Combination	Metric	Class 1: Mild	Class 2: Moderate	Class 3: Severe	Class 4: Proliferative DR
EyePACS + DDR + Aptos	F1-Score	0.47	0.52	0.17	0.40
	Precision	0.35	0.71	0.18	0.35
	Recall	0.71	0.42	0.16	0.46
	Accuracy			0.46	
	QWK			0.38	

Table 3.11: Performance comparison of (Pre-Trained on SynFundus1M) ResNet18 on EyePACS + DDR + Aptos Dataset across DR severity classes

Step 6

Model	Public Score	Private Score
ResNet (Fine-tuned)	0.49	0.76

Table 3.12: Performance of fine-tuned ResNet18 model on the APTOS 2019 competition leaderboard.

3.5. Summary

This chapter summarizes the results of the **approach used in our research** for detecting DR using pretrained deep learning models. Initially, we relied on real-world datasets such as APTOS, EyePACS, and DDR, which were obtained from the Kaggle platform. However, during training, we encountered a major challenge related to class imbalance, especially in the advanced stages of the disease. This led to the need for additional data, which we addressed by incorporating synthetic fundus images from the SynFundus1M dataset, generated using GAN.

The training and experimentation processes were conducted using both Kaggle's cloud-based environment and the HPC infrastructure available at our university. This dual setup enabled efficient handling of large-scale data and accelerated the training process. The results showed a clear improvement in classification performance when combining real and synthetic data. The final model was submitted to the APTOS 2019 competition, where it outperformed models trained from scratch.

General Conclusion

General Conclusion

In this research, We worked on pretrained model developement for the diagnosis of diabetic retinopathy using both real and synthetic data. Initially, we relied on real datasets such as APTOS, EyePACS, and DDR, achieving high binary classification accuracy ranging between 97% and 99%. However, after evaluating our models on other datasets from different sources, we found that some results did not reflect the true model performance in broader and more diverse environments, indicating challenges in model generalization and data scarcity.

To overcome the issue of insufficient data, especially in advanced cases, we were able to provide synthetic data through the SynFundus1M dataset generated using GAN. This synthetic data significantly improved the models' performance in multi-class classification. Before using SynFundus1M, the fine-tuned ResNet18 model achieved scores of 0.25 (public) and 0.58 (private) in the APTOS 2019 competition leaderboard, while the ResNet18 model trained from scratch scored 0.20 and 0.58 respectively. After integrating the synthetic data, the performance greatly improved to 0.49 (public) and 0.76 (private), confirming the vital role of synthetic data in enhancing the model's generalization capability and accuracy across diabetic retinopathy severity levels. However, time constraints prevented us from conducting more extensive experiments with the synthetic data despite the promising results.

We conducted the training processes using the Kaggle cloud platform along with the HPC infrastructure available at our university, which enabled more efficient and faster model training.

Pretrained models played a crucial role in accelerating training and enhancing performance, allowing us to leverage prior learning on large and diverse datasets. This enabled the extraction of robust features from fundus images despite the scarcity of local data. Although we did not achieve perfect results, using these models clearly improved classification accuracy compared to models trained from scratch.

While the classification performance across all stages of our work did not reach very high levels, the results were realistic and encouraging, supporting our research hypotheses and confirming the effectiveness of the adopted methodology. We also believe that this

approach is generalizable and can be applied to diagnose other diseases and across various medical fields, opening broad horizons for the use of artificial intelligence in healthcare improvement.

For future work, we see the necessity of creating a comprehensive local dataset that reflects the unique characteristics of patients in our healthcare environment, as well as exploring advanced learning techniques such as Few-Shot Learning and employing modern models like Vision Transformers. Additionally, improving model interpretability and reliability should be a focus to facilitate integration into clinical practice. This study represents an important initial step toward building intelligent and reliable systems to support early and accurate diagnosis of DR .

Bibliography

[1] Central Florida Retina, "Eye anatomy," https://www.centralfloridaretina.com/patient-resources/education/eye-anatomy/, 2025, accessed: 2025-05-14.

[2] Lighthouse Blindness Charitable Trust, "What is diabetic retinopathy?" https://lhblind.org/what-is-diabetic-retinopathy/, 2025, accessed: 2025-05-13.

[3] MDPI, "Stages of diabetic retinopathy," *MDPI*, 2025, accessed: 2025-05-14. [Online]. Available: https://www.mdpi.com/2075-4418/13/17/2783

[4] A. Name(s), "Diagram of the transfer learning process," https://www.researchgate.net/figure/Diagram-of-the-Transfer-Learning-Process-Source-ResearchGate_fig1_349908655, accessed: [Date of access].

[5] Medikal.net, "Oct optical coherence tomography device," accessed: 2025-05-13. [Online]. Available: https://www.medikal.net/en/eye-health-equipment/optical-devices/oct-optical-coherence-tomography-device/

[6] R. F. Spaide, "Fluorescein angiography showing pinpoint leakage in od," 2020, accessed: 2025-05-13. [Online]. Available: https://www.researchgate.net/figure/Fluorescein-angiography-FA-shows-pinpoint-leakage-in-OD-which-increases-in-size-and_fig3_341468098

[7] P. S. Silva and L. P. Aiello, "Ultra-wide field imaging for diabetic retinopathy screening," *Current Diabetes Reports*, vol. 20, no. 6, pp. 18–29, 2020.

[8] D. Solanki, "Supervised machine learning: A beginner's guide," 2024,

accessed: 2025-05-13. [Online]. Available: https://medium.com/@dhara732002/supervised-machine-learning-a-beginners-guide-9ac0b07eccbb

[9] D. Calvo, "Learning non-supervised," 2024, accessed: 2025-05-13. [Online]. Available: https://www.diegocalvo.es/en/learning-non-supervised/

[10] TechVidvan, "Reinforcement learning tutorial," 2024, accessed: 2025-05-13. [Online]. Available: https://techvidvan.com/tutorials/reinforcement-learning/

[11] M. Publishers, "Breast cancer early detection comparison with deep learning and machine learning models: A case of study," 2021, accessed: 2025-05-13. [Online]. Available: https://medwinpublishers.com/JQHE/breast-cancer-early-detection-comparison-with-deep-learning-and-machine-learning-models-a-case pdf

[12] Lukmaan IAS, "Convolutional neural networks (cnn): An in-depth exploration," 2024, accessed: 2025-05-13. [Online]. Available: https://blog.lukmaanias.com/2024/12/18/convolutional-neural-networks-cnn-an-in-depth-exploration/

[13] IBM, "Convolutional neural networks," 2025, accessed: 2025-05-13. [Online]. Available: https://www.ibm.com/think/topics/convolutional-neural-networks

[14] ResearchGate, "Relu activation function," 2017, accessed: 2025-05-13. [Online]. Available: https://www.researchgate.net/figure/ReLU-activation-function_fig3_319235847

[15] LearnOpenCV, "Batch normalization in deep networks," https://learnopencv.com/batch-normalization-in-deep-networks/, 2025, accessed: 2025-05-13.

[16] ResearchGate, "Examples for max and average pooling layer," https://www.researchgate.net/figure/Examples-for-max-and-average-pooling-layer_fig5_356816561, 2025, accessed: 2025-05-13.

[17] ——, "Conversion of convoluted layer to flatten layer," https://www.researchgate.net/figure/Conversion-of-convoluted-layer-to-flatten-layer_fig5_336320439, 2019, accessed: 2025-05-13.

[18] A. I. Aramendía, "Convolutional neural networks (cnns): A complete guide," https://medium.com/@alejandro.itoaramendia/convolutional-neural-networks-cnns-a-complete-guide-a803534a1930, 2023, accessed: 2025-05-13.

[19] LeewayHertz, "Transfer learning: Applications and benefits," https://www.leewayhertz.com/transfer-learning/, 2024, accessed: 2025-05-13.

[20] M. D. Abràmoff, Y. Lou, A. Erginay *et al.*, "Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning," *Investigative Ophthalmology Visual Science*, vol. 59, no. 10, pp. 4304–4312, 2018.

[21] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[22] V. Gulshan, L. Peng, M. Coram *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.

[23] D. S. W. Ting, C. Y. Cheung, G. Lim *et al.*, "Development and validation of a deep learning system for diabetic retinopathy assessment in primary care settings," *JAMA*, vol. 318, no. 22, pp. 2211–2223, 2017.

[24] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, 2017.

[25] W. H. Organization, *World Health Statistics 2020: Monitoring Health for the SDGs*. World Health Organization, 2020.

[26] B. Starfield, "Primary care: Balancing health needs, services, and technology," *Oxford University Press*, 1998.

[27] D. S. Kringos *et al.*, "The strength of primary care in europe: an international comparative study," *British Journal of General Practice*, vol. 63, no. 616, pp. e742–e750, 2013.

[28] J. Frenk, "The global health system: Strengthening national health systems as the next step for global progress," *PLoS Medicine*, vol. 7, no. 1, p. e1000089, 2010.

[29] E. J. Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*, 1st ed. New York: Basic Books, 2019.

[30] ——, "High-performance medicine: The convergence of human and artificial intelligence," *Nature Medicine*, vol. 25, no. 1, pp. 44–56, 2019.

[31] D. S. Ting *et al.*, "Deep learning in ophthalmology: The technical and clinical considerations," *Progress in Retinal and Eye Research*, vol. 72, p. 100759, 2019.

[32] J. E. Hollander and B. G. Carr, "The virtual visit: Use of telehealth during the covid-19 pandemic," *JAMA*, vol. 323, no. 18, pp. 1785–1786, 2020.

[33] J. Kvedar *et al.*, "Digital medicine: a primer on telehealth and telemedicine," *npj Digital Medicine*, vol. 1, pp. 1–10, 2014.

[34] A. K. Jha *et al.*, "Use of electronic health records in u.s. hospitals," *New England Journal of Medicine*, vol. 360, pp. 1628–1638, 2009.

[35] W. H. Organization, *Global Report on Diabetes*. World Health Organization, 2021.

[36] M. Knip *et al.*, "Pathogenesis of type 1 diabetes—an update," *Diabetes Research and Clinical Practice*, vol. 165, p. 108249, 2020.

[37] E. Adeghate *et al.*, "Type 2 diabetes mellitus: Etiology, pathophysiology, and complications," *Pathophysiology*, vol. 13, pp. 119–129, 2006.

[38] T. A. Buchanan *et al.*, "Gestational diabetes mellitus: Risks and management," *The Journal of Clinical Endocrinology & Metabolism*, vol. 97, pp. 41–47, 2012.

[39] C. A. Emdin *et al.*, "Diabetes and cardiovascular disease: Pathophysiology and implications for treatment," *Journal of the American College of Cardiology*, vol. 66, pp. 1723–1730, 2015.

[40] D. S. Ting *et al.*, "Diabetic retinopathy: Global prevalence, major risk factors, screening, and treatment," *The Lancet Global Health*, vol. 5, pp. e888–e897, 2017.

[41] A. I. Vinik *et al.*, "Diabetic neuropathy: Pathophysiology and therapy," *Diabetes Care*, vol. 36, pp. 245–252, 2013.

[42] P. Rossing *et al.*, "Diabetic nephropathy: Pathophysiology and treatment," *Kidney International*, vol. 62, pp. 2041–2053, 2002.

[43] W. J. Jeffcoate *et al.*, "Diabetic foot ulcers: Pathogenesis and treatment," *The Lancet Diabetes & Endocrinology*, vol. 6, pp. 159–170, 2018.

[44] J. Tuomilehto *et al.*, "Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance," *New England Journal of Medicine*, vol. 344, pp. 1343–1350, 2001.

[45] D. M. Nathan *et al.*, "Medical management of hyperglycemia in type 2 diabetes: A consensus algorithm," *Diabetes Care*, vol. 32, pp. 193–203, 2009.

[46] D. J. Drucker, "The glp-1 receptor agonists: Mechanism of action," *Endocrinology and Metabolism Clinics*, vol. 35, no. 4, pp. 747–768, 2006.

[47] M. D. Abramoff, M. S. Malik, and M. F. Fleming, "Automated analysis of retinal images for detection of diabetic retinopathy," *JAMA Ophthalmology*, vol. 136, no. 10, pp. 1153–1160, 2018.

[48] M. D. Abramoff and J. Doe, "Deep learning in diabetic retinopathy diagnosis," *Ophthalmology AI*, vol. 2, no. 3, pp. 50–62, 2018.

[49] E. Sohn, N. M. Bressler, and S. B. Bressler, "Optical coherence tomography in diabetic retinopathy: A review of current applications," *Middle East African Journal of Ophthalmology*, vol. 23, no. 4, pp. 335–341, 2016.

[50] J. D. M. Gass, *Stereo Atlas of Macular Diseases: Diagnosis and Management*, 4th ed. St. Louis: Mosby-Year Book, 1997.

[51] M. M. Wessel, G. D. Aaker, G. Parlitsis, N. Patel, D. J. D'Amico, and S. Kiss, "Ultrawide-field angiography improves clinical decision-making in patients with diabetic retinopathy," *Retina*, vol. 34, no. 5, pp. 867–875, 2014.

[52] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[53] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[54] J. E. Van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Machine Learning*, vol. 109, pp. 373–440, 2020.

[55] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT press, 2018.

[56] W. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[57] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychological Review*, vol. 65, no. 6, pp. 386–408, 1958.

[58] M. Minsky and S. Papert, *Perceptrons: An Introduction to Computational Geometry*. MIT Press, 1969.

[59] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986.

[60] G. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[61] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2012.

[62] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.

[63] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[64] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[65] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," *ICML*, 2010.

[66] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *ICML*, 2015.

[67] R. Zhang, P. Isola, and A. A. Efros, "Making convolutional networks shift-invariant again," *International Conference on Machine Learning (ICML)*, 2019.

[68] M. Lin, Q. Chen, and S. Yan, "Network in network," in *International Conference on Learning Representations (ICLR)*, 2013.

[69] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[70] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[71] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Advances in neural information processing systems*, 2014, pp. 3320–3328.

[72] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations*, 2020.

[73] Y. Chen, Y. Zhang, M. Li, J. Wang, and X. Liu, "Vision transformer-based automatic detection of diabetic retinopathy," *IEEE Access*, vol. 9, pp. 123 456–123 464, 2021.

[74] M. O. Irfanoglu, G. Karakaya, and Z. Aydın, "A comparative study of cnn and vision transformer-based models for diabetic retinopathy detection," *Computer Methods and Programs in Biomedicine*, vol. 213, p. 106503, 2022.

[75] M. Saha, D. Mitra, C. Chakraborty *et al.*, "Automated diabetic retinopathy detection using deep learning models: A comparative study," *IEEE Access*, vol. 9, pp. 45 678–45 690, 2021.

[76] K. Ragupathi, S. Prabha, and V. Balaji, "Efficientnet-based model for automated detection of diabetic retinopathy," *Journal of Medical Systems*, vol. 46, no. 2, pp. 1–11, 2022.

[77] A. Porwal, R. B. Pachori, and K. Raja, "A survey on computer aided detection of diabetic retinopathy," *Biomedical Signal Processing and Control*, vol. 45, pp. 1–15, 2018.

[78] M. Niemeijer, B. van Ginneken, S. R. Russell *et al.*, "Automated detection of diabetic retinopathy using random forest and cnn features," *IEEE Transactions on Medical Imaging*, vol. 38, no. 1, pp. 121–131, 2019.

[79] V. Gulshan, L. Peng, T. Corrado *et al.*, "Deep learning for diabetic retinopathy detection and analysis," *Ophthalmology*, vol. 127, no. 8, pp. 1082–1089, 2020.

[80] W. Zhang, X. Li, Y. Wang *et al.*, "Hybrid deep learning model for multi-class classification of diabetic retinopathy," *Computer Methods and Programs in Biomedicine*, vol. 202, p. 105978, 2021.

[81] Kaggle, "Diabetic retinopathy detection dataset (eyepacs)," 2015, accessed: 2025-06-19. [Online]. Available: https://www.kaggle.com/competitions/diabetic-retinopathy-detection

[82] ——, "Aptos 2019 blindness detection," 2019, accessed: 2025-06-19. [Online]. Available: https://www.kaggle.com/competitions/aptos2019-blindness-detection

[83] H. Li *et al.*, "Ddr: Diabetic retinopathy image dataset," 2019, accessed: 2025-06-19. [Online]. Available: https://www.kaggle.com/datasets/yin1933/idriddiseasegrading

[84] Messidor, "Messidor 1," 2008, accessed: 2025-06-19. [Online]. Available: https://www.kaggle.com/datasets/miraj24hossain/messidor1-data

[85] ——, "Messidor 2," 2011, accessed: 2025-06-19. [Online]. Available: https://www.kaggle.com/datasets/google-brain/messidor2-dr-grades

[86] P. Porwal *et al.*, "Indian diabetic retinopathy image dataset (idrid)," 2018, accessed: 2025-06-19. [Online]. Available: https://www.kaggle.com/datasets/mariaherrerot/idrid-dataset

[87] Y. Zhao *et al.*, "Synfundus1m: Synthetic fundus image dataset," 2023, accessed: 2025-06-19. [Online]. Available: https://huggingface.co/datasets/synfundus/SynFundus1M