الجمهورية الجزائرية الديمقراطية الشعبية
People's Democratic Republic of Algeria
وزارة التعليم العالي والبحث العلمي
Ministry of Higher Education and Scientific Research
جامعة الشهيد الشيخ العربي التبسي -تبسة
Echahid Cheikh Larbi Tebessi University

Faculty of Science and Technology

Department of Electrical Engineering

Faculty of Exact Sciences and Natural and Life Sciences

Department of Computer Science

Academic Master Thesis

Realtime Retinopathy Detection via a Mobile Fundus Camera

Defended by:

RAHMOUNI Oualid
Domain: **Sciences and Technology**
Science field: **Automatic**
Option: **Automatic and Systems**

MAALEM Abdelmouaaz
Domain: **Mathematics and Computer**
Science field: **Computer Science**
Option: **Systems and Multimedia**

Board of Examiners:

Dr. DJABRI Riadh	MCB	President
Dr. LAIMECHE Lakhdar	MCA	Examiner
Dr. BOUCHEMHA Amel	MCA	Supervisor
Dr. BOUALLEG Yaakoub	MCB	Supervisor
Pr. DJEDDI Chawki	Prof	Co- Supervisor

Academic Year 2023-2024

We would like to dedicate this thesis to our families and friends.

Acknowledgements

In the name of ALLAH, the Most Gracious, the Most Merciful.

All praise and thanks are due to ALLAH, the Almighty, for His countless blessings and for granting us the strength, knowledge, and perseverance to complete this master's journey. We express profound gratitude for His unwavering direction throughout this challenging journey.

We would like to extend our heartfelt appreciation to the individuals and institutions that have supported us during our pursuit of this master's degree.

We express our profound gratitude to our supervisors for their essential mentorship, expertise, and unwavering support. Their perceptive critique and support have played a crucial role in influencing our research and our development as scholars.

We express our gratitude to our fellow master's students for their engaging discussions, cooperative attitude, and instances of mutual support, which have enhanced the pleasantness of our journey.

Lastly, we would like to extend our sincere appreciation to our families and friends for their unwavering affection, patience, and moral encouragement during this demanding process. Your support has been our source of strength.

This achievement would not have been attainable without the mercy of Allah and the support of all these people. Alhamdulillah.

Abstract

Diabetic Retinopathy (DR) is a frequent complication of diabetes mellitus that compromises retinal function in more than 50% of type 2 diabetic patients. It occurs when the retina's blood vessels deteriorate. These altered vessels can dilate, leak fluid (plasma, lipids, and/or blood), and even clog, leaving part of the retina without blood flow. All these phenomena that occur as a result of diabetes can cause progressive damage to the structures of the eyeball, leading to a severe reduction in vision and even, without appropriate treatment, to blindness in the working age. In our work, we propose a framework for diabetic retinopathy detection based on rerinal lesions using advanced deep learning. We use the U-Mamba architecture for retinal and blood vessel segmentation, achieving high F1 scores for various lesion types. Our Swin Transformer-based classification model, incorporating lesion segmentation masks, demonstrated exceptional performance across multiple datasets, with up to 97.75% accuracy on the EyePAC dataset. This approach outperformed existing models across various datasets, showing promise for clinical DR diagnosis.

Keywords: Diabetic Retinopathy Detection, Retinal Lesions Segmentation, Blood vessel segmentation, Deep Learning, Retina Image Analysis.

Résumé

La rétinopathie diabétique, une complication courante du diabète, affecte la fonction rétinienne chez plus de la moitié des patients atteints de diabète de type 2. Cette pathologie résulte de la dégradation des vaisseaux sanguins rétiniens, entraînant dilatation, fuites de fluides et occlusions vasculaires. Sans traitement adéquat, ces altérations peuvent conduire à une perte de vision sévère, voire à la cécité chez les adultes en âge de travailler. Notre étude propose une approche novatrice pour la détection de la rétinopathie diabétique, basée sur l'analyse des lésions rétiniennes par apprentissage profond avancé. Nous avons employé l'architecture U-Mamba pour la segmentation de la rétine et des vaisseaux sanguins, obtenant des scores F1 élevés pour différents types de lésions. Notre modèle de classification, fondé sur le Swin Transformer et intégrant des masques de segmentation, a démontré des performances remarquables sur plusieurs jeux de données, atteignant une précision allant jusqu'à 97,75%.
Mots Clée : Détection de la rétinopathie diabétique, Segmentation des lésions rétiniennes, Segmentation vasculaire rétinienne, Apprentissage profond, Analyse d'images rétiniennes.

ملخص

يُعد اعتلال الشبكية السكري من أكثر مضاعفات داء السكري شيوعا، حيث يؤثر على ما يزيد عن ٥٠٪ من مرضى السكري من النوع الثاني. تتجلى هذه الحالة في تلف الأوعية الدموية الدقيقة في شبكية العين، مما قد يؤدي إلى توسعها أو تسريبها للسوائل أو انسدادها. نتيجة لذلك، قد تتعرض أجزاء من الشبكية لنقص في التروية الدموية، مما يهدد سلامة العين ووظائفها البصرية. وفي غياب التشخيص والعلاج المبكرين، قد يتطور الأمر إلى فقدان حاد للبصر أو حتى العمى، مما يؤثر بشكل كبير على جودة حياة المرضى.

يقدم بحثنا طريقة مبتكرة للكشف عن اعتلال الشبكية السكري باستخدام تقنيات التعلم العميق المتقدمة. اعتمدنا على بنية U-Mamba لتقسيم صور الشبكية والأوعية الدموية، محققين نتائج عالية الدقة في تحديد مختلف أنواع الآفات الشبكية. كما طورنا نموذجا للتصنيف يعتمد على تقنية Swin Transformer، مع دمج معلومات تقسيم الآفات، مما أدى إلى تحقيق أداء متميز عبر مجموعات بيانات متنوعة. وصلت دقة النموذج إلى ٥٧.٧٩٪ عند اختباره على مجموعة بيانات EyePACS، متفوقا بذلك على النماذج الحالية في العديد من مجموعات البيانات الأخرى. تُظهر هذه النتائج إمكانات واعدة لتطبيق نهجنا في التشخيص السريري لاعتلال الشبكية السكري، مما قد يسهم في تحسين رعاية المرضى والكشف المبكر عن المرض، وبالتالي الحد من مخاطر فقدان البصر.

الكلمات المفتاحية: اعتلال الشبكية السكري، تحليل صور الشبكية، التعلم العميق، تقسيم الآفات الشبكية، تشخيص أمراض العين.

Contents

List of Figures i

List of Tables iii

General Introduction 1

1 Basic Concepts 4

1.1 Medical Background 4

1.1.1 Diabetes Disease 4

1.1.2 The Human Eye 7

1.1.3 Diabetic Retinopathy 9

1.2 Deep Learning in Medical Image Analysis 15

1.2.1 Introduction to Deep Learning 15

1.2.2 Deep Learning for Image Classification 22

1.2.3 Deep Learning for Image Segmentation 28

2 State of the Art 33

2.1 Retinal Image Segmentation 34

2.1.1 Retinal Blood Vessel Segmentation 34

2.1.2 Retinal Lesions Segmentation 42

2.2 Diabetic Retinopathy Classification 46

2.2.1 End-to-End Learning 46

2.2.2 Retinal Lesions based Methods 50

3 Contributions and Results 55

3.1 Proposed Framework 55

3.1.1 Retinal Image Preprocessing 56

3.1.2 Retina Image Segmentation Phase 58

3.1.3 Diabetic Retinopathy Classification Phase 63

3.2 Experimental Setup 65

3.2.1 System Configuration 65

3.2.2 Datasets 65

3.2.3 Evaluation Metrics 66
3.3 Experimental Results 67
3.3.1 Image Segmentation 67
3.3.2 Diabetic Retinopathy Detection 76
3.4 Discussion 83

General Conclusion **85**

Bibliography **86**

List of Figures

1.1 Illustration of the human eye anatomy. 7
1.2 A retinal fundus image and its main anatomical features. 9
1.3 Example of retinal in each diabetic retinopathy stage. 10
1.4 Different retinal lesions according to each diabetic retinopathy stage [1]. . 13
1.5 Architecture of the Artificial Neural Network [2]. 17
1.6 Illustration of a Multi-layer Perceptrons Architecture [3]. 19
1.7 Illustration of CNN architecture: example of VGG16 model. 20
1.8 Principle of filter sliding in the convolution layer over an image. 21
1.9 Pooling layer principle: example of performing Max pooling function. . . 21
1.10 Illustration of Fully-Connected Layers structure [4]. 22
1.11 History of The CNN architectures evolution khan2020survey. 24
1.12 Architecture of a fully convolutional network [5]. 29
1.13 Architecture of a Mask R-CNN model [6]. 30
1.14 Standard U-Net architecture [7]. 30

2.1 Overview of a method for Retinal Blood Vessel Segmentation: (a) Block diagram; (b) FCN architecture [8]. 36
2.2 History of the automatic segmentation of retinal vessels [9]. 41

3.1 Proposed Framework for Retinal Lesions-based Diabetic Retinopathy Detection. 56
3.2 Illustration of contrast enhancement using Contrast Limited Adaptive Histogram Equalization technique left: image before, right: image after. . . . 56
3.3 U-Mamba architecture. 59
3.4 Detailed architecture of U-Mamba block. 60
3.5 Mamba Block. 61
3.6 Structured state-space models. 62
3.7 An illustration of the shifted window. 63
3.8 Swin-Transformer block layer. 64
3.9 Swin-Transformer architecture. 65
3.10 application of Croping and Flipping 66

3.11 Behaviour of DSC and loss with respect to epochs in training the modified U-mamba model for Lesion segmentation of microaneurysm images. . . . 70
3.12 Behaviour of DSC and loss with respect to epochs in training the modified U-mamba model for Lesion segmentation of Hard-Exudates images. . . . 71
3.13 Behaviour of DSC and loss with respect to epochs in training the modified U-mamba model for Lesion segmentation of Soft-Exudates images. 72
3.14 Behaviour of DSC and loss with respect to epochs in training the modified U-mamba model for Lesion segmentation of Hemorrhage images. 73
3.15 Behaviour of DSC and loss with respect to epochs in training the modified U-mamba model for Lesion segmentation of blood vessels images. 75
3.16 Confusion matrix for 5 class using Swin Transforme for EyePac Dataset. 79
3.17 Confusion matrix for 5 class using Swin-Transformer for Idrid Dataset. . 79
3.18 Confusion matrix for 5 class using Swin Transforme for Aptos2019 Dataset. 79
3.19 Accuracy and loss curves in train and validation on (A=EyePac, B=Idrid, C=Aptos2019) Datasets. 80
3.20 Bar shart of Sensitivity and Specificity for each class on (A=EyePac, B=Idrid, C=Aptos2019) Datasets. 80

List of Tables

1.1 Non-Proliferative Diabetic Retinopathy Grades and Clinical Features. . . 11
1.2 Comparison of Common Activation Functions 18
1.3 Types of Attention Mechanisms for Image Classification 25

2.1 Synthesis of Retinal Blood Vessel Segmentation Methods 38
2.2 Synthesis of Retinal Lesion Segmentation Methods 44
2.3 Summary of End-to-End Learning Techniques for DR Classification . . . 47
2.4 Summary of Retinal Lesion-Based Methods for DR Classification 52

3.1 Segmentation Dataset details. 66
3.2 Classification Dataset details (TS=Total Size) (TrS=Train Size)(TsS=Test Size)(IS=Image Size)(IT=Image Type) 66
3.3 Hyperparameters of U-mamba model for segmentation on IDRID and DRIVE datasets. 69
3.4 Results of Microaneurysm Segmentation. 69
3.5 Results of Hard-Exudates Segmentation. 70
3.6 Results of Soft-Exudates Segmentation. 71
3.7 Results of Hemorrhage Segmentation. 72
3.8 Comparison of our proposition with the state-of-the-art methods on the Idrid dataset. 74
3.9 Results of Blood Vessel Segmentation. 75
3.10 Comparison of U-mamba model with the state-of-the-art methods on the DRIVE dataset, the best result is in bold. 76
3.11 Metrics values for different classification datasets. 78
3.12 Comparison of Swin-Tansformer model for classification with the state-of-the-art methods on the Idrid dataset. 81
3.13 Comparison of Swin-Tansformer model for classification with the state-of-the art methods on the EyePAc dataset, the best result is in bold. 82
3.14 Comparison of the Swin-Tansformer model for classification with the state-of-the-art methods on the Aptos2019 dataset, the best result is in bold. . 82

General Introduction

Introduction

The main cause of vision impairment and blindness in diabetics is diabetic retinopathy (DR). The blood vessels in the retina may be harmed by uncontrolled diabetes. Prompt medical care and early identification are essential for preventing further visual impairment. Diabetic retinopathy (DR) is a pathology that affects patients with type 1 and type 2 diabetes, as a result of the most common complication of microangiopathy. It occurs when the retina's small blood vessels are damaged. These altered vessels can dilate, leak fluid (plasma, lipids, and/or blood), and even clog, leaving part of the retina without blood flow [10].

Diabetic retinopathy ranks as the second leading cause of blindness after age-related macular degeneration (before the age of 55). No proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR) are the two types of diabetic retinopathy that affect the retina (PDR). About 35.4% of diabetes patients worldwide have DR, with a third having vision-threatening DR and 7.6% having retinal edema [11]. The world prevalence of diabetes among adults (aged 20-79 years) will be 6.4%, affecting 285 million adults, in 2010, and will increase to 7.7%, and 439 million adults by 2030 [12].

The Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR) is a longitudinal epidemiologic study of the progression of diabetic retinopathy. The WESDR demonstrated a direct relationship between the prevalence of diabetic retinopathy and the duration of diabetes, in both types of diabetes (types 1 and 2). After 20 years of diabetes, almost 99% of type 1 diabetics and 60% of type 2 diabetics have diabetic retinopathy. Moreover, proliferative diabetic retinopathy is present in 50% of type 1 diabetics after 20 years of diabetes, and in 25% of type 2 diabetics after 25 years of diabetes. What's more, 3.6% of young patients (under 30 years of age at diagnosis) and 1.6% of patients over 30 have visual acuity of 1/10 or less. This loss of vision is attributable to diabetes in 86% of young patients (under 30) and in 33% of older patients [13].

Fundus fluorescein angiography is currently the gold-standard investigative method for determining the state of the vessels in the fundus and retina (neovascularization). Other techniques could potentially offer a safer, faster, and equally effective alternative method for diagnosing and monitoring diabetic retinopathy [14]. Optical coherence tomography

(OCT) is a non-invasive, non-contact imaging technique that produces micrometer-level resolution of ocular tissue sections. OCT is based on the reconstruction of reflected light. Similar to an ultrasound scan, the technique creates a two-dimensional image of the light that the retina's various layers have backscattered [15].

The diagnosis of this retinal pathology requires a combination of careful clinical examination and specialized imaging techniques. Early detection of diabetic retinopathy is essential to restore vision and provide timely treatment. The high number of diabetics worldwide indicates that DR remains a major factor in partial or total vision loss. Consequently, early detection followed by rapid treatment procedures for people with diabetes-related diseases is vital.

This project aims to propose an automated real-time diagnostic model based on robust and efficient artificial intelligence algorithms, which would automatically detect this complication and speed up the work of doctors. This model will then be validated by clinical evaluation by doctors and by performance criteria associated with the automatic detection model, to provide early detection and predict the risk or level of progression of the disease.

Deep Learning (DL) is a class of artificial intelligence (AI) methods inspired by the structure of the human brain and based on artificial neural networks [16]. Essentially, DL refers to methods that automatically learn the mathematical representation of intrinsic data relationships. Unlike traditional machine learning methods, deep learning methods require far less human guidance, as they are not based on the generation of hand-crafted features, a task that can be very laborious. Instead, they learn appropriate features directly from the data. What's more, DL methods scale much better than traditional machine learning methods as the amount of data increases. As part of this project, we will be developing a real automatic system for the diagnosis and detection of retinal pathologies based on the Deep Learning algorithm.

Project Contributions

Our project's main goal is to address the pressing need for accessible DR detection tools designed exclusively for healthcare providers in response to these major care-related barriers through the development of a mobile app driven by advanced AI technology. Our contributions are as follows:

- Provide OPHTALMO SCAN, a mobile application that uses a straightforward, reasonably priced smartphone camera device to enable real-time DR detection by utilizing artificial intelligence (AI). This intuitive software does more than just locate possible DR. It leads users through the process of taking retinal images and guarantees that the images are of suitable quality for precise analysis. After examining the collected image, the AI algorithm gives a prompt assessment of the possibility of DR.

- The design and development of an intelligent system to analyze retinal images for the early detection of diabetic retinopathy. This model will have the ability to detect discriminating features that enable DR to be identified using retinal images with good performance in terms of accuracy.
- Develop a set of robust methods to automate as far as possible the process of detecting and analyzing retinal diseases on retinal images.
- Assessment of the quality of the proposed approaches, by comparing them with work reported in the state of the art.

Manuscript structure

After an introduction, the manuscript will be organized as follows:

- **Chapter 1:** In this chapter, we introduce the basic concepts that provide the fundamental knowledge needed to understand this project. This chapter is divided into two parts. In the first part, we delve into the medical context by providing a detailed exploration of the human eye, focusing on its anatomy and functions to explain how diabetes can lead to retinal damage. Next, we detail the risk factors, causes, and stages of diabetic retinopathy. The second part of the chapter will be devoted to the technological context, introducing the essential concepts of Deep Learning (DL), with a focus on medical image classification and segmentation
- **Chapter 2:** This Chapter presents a comprehensive review of the most relevant research in retinal image segmentation and DR classification, focusing on two pivotal aspects: the delineation of retinal structures and the identification of pathological changes indicative of DR progression.
- **Chapter 3:** In this chapter, we introduce our proposed framework for DR detection, overcoming the limits of related works, and then we delve into the experimental study, detailing the used datasets, the experimental setups, and the exhausted experiments conducted to evaluate the performance of our contributions.

Finally, we will end this master thesis with a general conclusion encompassing the work and research prospects for future work.

Chapter 1

Basic Concepts

Introduction

In this chapter, we introduce some basic concepts of the research area, giving essential background knowledge to understand the remaining parts of this thesis. This chapter consists of two parts. In the first part, we delve into the medical background, starting with an introduction to diabetes, a chronic condition that affects millions worldwide. This part also provides a detailed exploration of the human eye, focusing on its anatomy and functions to explain how diabetes can lead to retinal damage. Subsequently, we focus on Diabetic Retinopathy (DR), discussing its causes, symptoms, stages, symptoms, and current treatment options.

The second part of the chapter transitions to the technological background, introducing essential concepts Deep Learning (DL), which form the backbone of modern medical image analysis. Then, we highlight image classification and segmentation, two critical tasks in computer vision that have profound applications in detecting DR. By bridging the medical and technological domains; this chapter sets the stage for the subsequent exploration of DR detection through advanced image processing techniques.

1.1 Medical Background

1.1.1 Diabetes Disease

Diabetes mellitus presents as a chronic metabolic disorder with high blood sugar levels (hyperglycemia) resulting from inadequate insulin production, insufficient insulin action, or a combination of both [17]. The pancreas produces insulin, which enables cells to absorb glucose for energy needs [17]. In cases of type 1 diabetes, insufficient or no insulin is produced due to the autoimmune destruction of beta cells responsible for producing insulin [18]. Conversely, type 2 diabetes is characterized by the body's cells developing resistance to the effects of insulin, often coupled with a gradual decline in insulin production over time [19]. This imbalance of insulin and glucose can lead to chronic damage, dysfunction, and eventual failure of various organs. Critical organs affected include the eyes, kidneys, nerves, heart, and blood vessels. Uncontrolled diabetes significantly increases the risk of severe complications, such as cardiovascular diseases, kidney failure, neuropathy, retinopathy, and lower-limb amputations, thereby impacting multiple organ systems [20, 21].

1.1.1.1 Diabetes Types

Having discussed the general nature of diabetes, it is important to distinguish between the major types of this metabolic disorder:

- Type 1 Diabetes: An autoimmune condition entails the immune system's attack on and destruction of the insulin-producing beta cells in the pancreas, resulting in a complete lack of insulin. It accounts for about 5-10% of all diabetes cases and often presents in childhood or adolescence, though it can occur at any age [22].

- Type 2 Diabetes: The most prevalent type, which accounts for approximately 90-95% of all diabetes cases, is characterized by insulin resistance and a relative deficiency in insulin secretion. This type often develops gradually and is associated with older age, obesity, physical inactivity, and a family history of diabetes [22].

- Gestational Diabetes: Gestational diabetes mellitus (GDM) occurs during pregnancy and usually resolves after childbirth. It is characterized by glucose intolerance and hyperglycemia due to insulin resistance, exacerbated by the placental hormones. GDM can lead to various complications for both the mother and the fetus, including preeclampsia, macrosomia, and an increased risk of developing type 2 diabetes later in life [23, 24].

- Impaired glucose tolerance and impaired fasting glycaemia: Conditions such as impaired glucose tolerance (IGT) and impaired fasting glycaemia (IFG) signify a middle ground between normal blood sugar levels and diabetes. People with IGT or IFG have a higher likelihood of developing type 2 diabetes, but it is not a guaranteed progression [25].

- Other Types of Diabetes: Monogenic diabetes syndromes, disorders of the exocrine pancreas (for example, cystic fibrosis and pancreatitis), and drug-induced diabetes resulting from medications like glucocorticoids and specific antipsychotics are all included in this category [22].

A comprehensive understanding of the various types of diabetes and their unique characteristics is crucial for effective management and prevention of complications. Early diagnosis and appropriate treatment can significantly improve health outcomes and enhance the quality of life for individuals with diabetes.

1.1.1.2 Epidemiology

Diabetes mellitus poses a significant global health challenge, with rapidly escalating prevalence rates. As of 2021, approximately 537 million adults worldwide were living with diabetes, a figure projected to surge to 783 million by 2045 [26]. Over 75% of these adults reside in low- and middle-income countries, driven by factors such as urbanization, aging populations, decreased physical activity, and rising obesity rates. Type 1 diabetes, constituting about 5-10% of all diabetes cases, is one of the most prevalent chronic diseases in children, with the highest incidence rates observed in Scandinavia and the lowest in East Asia [27]. Type 2 diabetes accounts for 90-95% of diabetes cases, primarily affecting adults but increasingly seen in younger populations, including children and adolescents. The highest prevalence of type 2 diabetes is found in low- and middle-income countries, where urbanization and lifestyle changes exacerbate the epidemic [28].

Gestational diabetes affects approximately 14% of pregnancies globally, with higher rates in certain ethnic groups, including South Asian, Hispanic, and African-American populations. The prevalence of gestational diabetes is on the rise due to increasing maternal age and obesity rates, posing long-term risks for both the mother and offspring, including the subsequent development of type 2 diabetes [24]. In Algeria, the prevalence of diabetes among adults aged 19-69 years is estimated at 14.4%, with macrovascular complications present in 7.6% and microvascular complications in 7.4% of the diabetic population. These statistics underscore the significant burden of diabetes in the country and the need for comprehensive management strategies.

The economic burden of managing diabetes and its complications is substantial. Preventive measures, including lifestyle changes and metformin, have been shown to reduce these costs over time significantly. Furthermore, studies have revealed significant gaps in diabetes-related knowledge and practices, emphasizing the necessity for improved education and management strategies to enhance patient outcomes. Regular screening and early intervention are crucial, particularly for conditions like diabetic retinopathy, which has a high prevalence among diabetic patients and can lead to severe visual impairment if left untreated [29].

Overall, these findings underscore the urgent need for comprehensive public health strategies that focus on prevention, early diagnosis, and effective management to mitigate the impact of diabetes and its complications on individuals and healthcare systems. According to the diagnosis, during the first 20 years, approximately all individuals with diabetes of type 1, and over 60% of those with diabetes of type 2, are likely to develop diabetic retinopathy (DR). At the time of diagnosis, 21% of those with type 2 show signs of diabetic retinopathy [30].

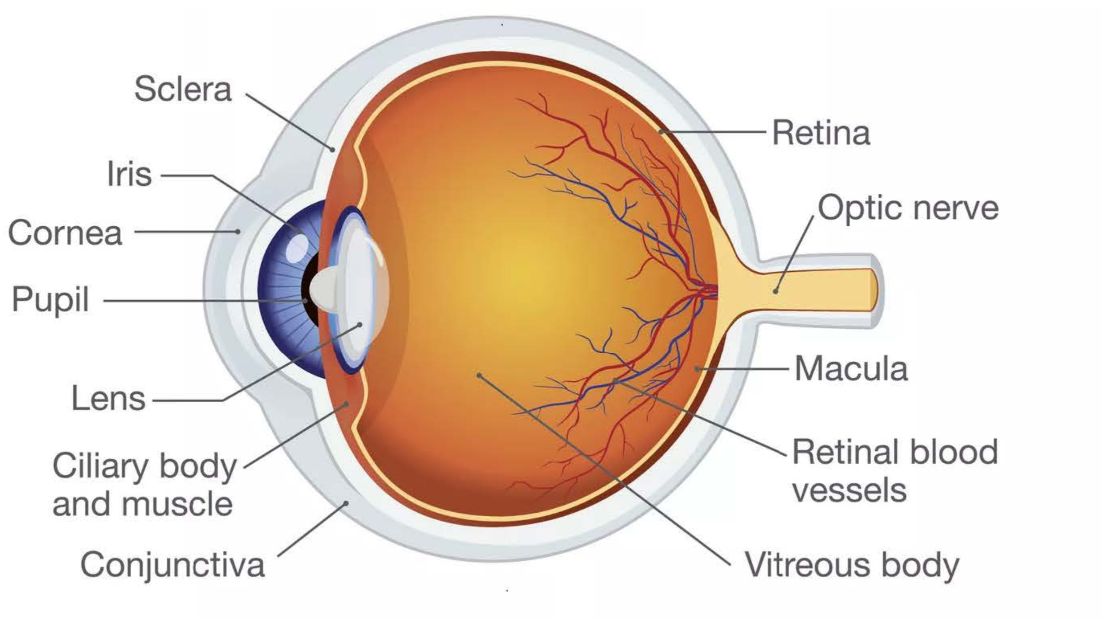

Figure 1.1: Illustration of the human eye anatomy.

1.1.2 The Human Eye

The human eye is a complex and sophisticated organ responsible for the sense of sight. It comprises various structures that work together to capture, focus, and transmit light, ultimately resulting in visual perception.

1.1.2.1 Anatomy of the Human Eye

The human eye consists of various interconnected structures that work harmoniously to facilitate vision. Figure 1.1 provides a visual representation of the anatomical components of the human eye, with a specific emphasis on its external and internal structures.

The sclera is the tough outer layer maintaining the eye's shape. The cornea is a clear, dome-shaped front part that focuses light onto the retina. The iris is the colored part of the eye surrounding the pupil, which controls the amount of light entering the eye. Behind the pupil, the lens changes shape to focus light on the retina.

The retina contains photoreceptor cells (rods and cones) that convert light into electrical signals. The macula, at the center of the retina, includes the fovea, the point of sharpest vision. The optic nerve carries visual information from the retina to the brain [31]. The vitreous humor is a gel-like substance that fills the space between the lens and the retina, helping maintain the eye's shape. The choroid supplies blood to the retina and absorbs excess light to prevent reflection within the eye [32].

Understanding eye anatomy is essential for managing eye health and treating conditions like diabetic retinopathy. High blood sugar can harm the retina's blood vessels, leading to vision loss or blindness, emphasizing the retina's crucial role in visions [33].

1.1.2.2 Retina

The retina is a complex and delicate tissue located between the vitreous body and the choroid at the posterior part of the eye. It consists of several layers of specialized cells that work together to facilitate vision. The primary function of the retina is to convert incoming light into electrical signals that the brain can interpret.

The retina contains various types of cells, including photoreceptor cells, bipolar cells, ganglion cells, and several other types of supporting cells. The two main types of photoreceptor cells are rods and cones. Rods are responsible for vision in low light conditions and are highly sensitive to light but do not discern color. Cones, on the other hand, are responsible for color vision and function best in bright light. In addition to photoreceptor cells, the retina also contains other specialized cells, such as horizontal cells and amacrine cells, which play a role in modulating and fine-tuning the visual signals as they pass through the retinal layers.

The retina is nourished by a network of blood vessels, including the central retinal artery and vein, which supply oxygen and nutrients to the retinal cells. Fundus images are used to capture detailed anatomical information about the retina. Figure 1.2 illustrates the retina's appearance in a fundus image [34]. In addition to blood vessels, the are three main anatomical parts that make up the retina:

- Macula: is a small, specialized area that is located near the center of the retina. It is responsible for central vision, which is crucial for tasks such as reading, driving, and recognizing faces. Within the macula, there is a high concentration of cone cells, that are responsible for detailed and color vision. The macula contains a small depression called the fovea.

- Optic Disk: also known as the optic nerve head, is a region on the retina where the optic nerve exits the eye. It does not contain any photoreceptor cells, making it a blind spot in our visual field. The optic disk appears as a circular or oval-shaped area and is easily identified in fundus images. It serves as a crucial landmark for assessing the health of the optic nerve.

- Fovea: is a tiny pit at the macula's center. It is the area of the retina with the highest concentration of cone cells and is responsible for our sharpest central vision. The fovea enables us to perceive fine details and provides excellent visual acuity.

1.1.2.3 Fundus Photography

Fundus photography is a technique used to capture images of the fundus, which can be used for diagnostic and research purposes. Specialized fundus cameras consisting of an intricate microscope attached to a flash-enabled camera are used in fundus photography.

The optical design of fundus cameras is based on the principle of monocular indirect ophthalmoscopy, which provides an upright, magnified view of the fundus. A typical

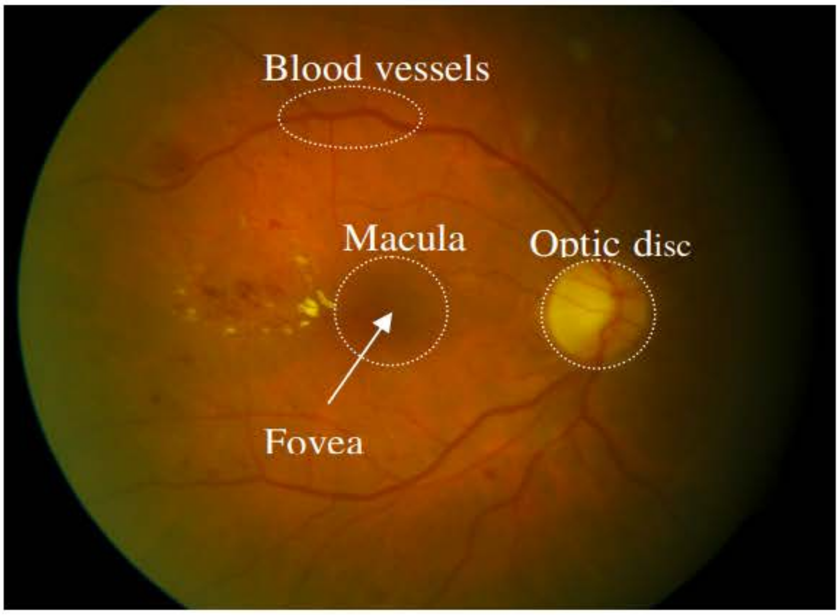

Figure 1.2: A retinal fundus image and its main anatomical features.

camera views 30 to 50 of retinal area, with a magnification of 2.5x, and allows some modification of this relationship through zoom or auxiliary lenses.

Fundus photography is a useful tool for diagnosing, educating patients, counseling, monitoring, and forecasting many ophthalmic conditions, notably diabetic retinopathy, age-related macular degeneration, retinal vascular disorders, and glaucoma.

1.1.3 Diabetic Retinopathy

Diabetic retinopathy (DR) is a microvascular complication of diabetes mellitus characterized by damage to the retinal blood vessels due to chronic hyperglycemia. In DR, prolonged elevated blood glucose levels cause pathological changes in the retinal microvasculature, including capillary basement membrane thickening, pericyte loss, and endothelial cell damage. These changes lead to increased vascular permeability, capillary occlusion, and ischemia, ultimately resulting in retinal hemorrhages, fluid leakage, and the formation of new, fragile blood vessels (neovascularization). If left untreated, DR can progress to severe visual impairment and blindness. DR is a leading cause of blindness among adults aged 20-74 years and significantly impacts the quality of life of individuals with diabetes [35].

1.1.3.1 Stages of Diabetic Retinopathy and Clinical Features

Diabetic retinopathy (DR) presents with a range of clinical manifestations that vary based on the stage and severity of the disease (see Figure 1.3). The primary stages are non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR). Additionally, diabetic macular edema (DME) is a critical complication that can occur

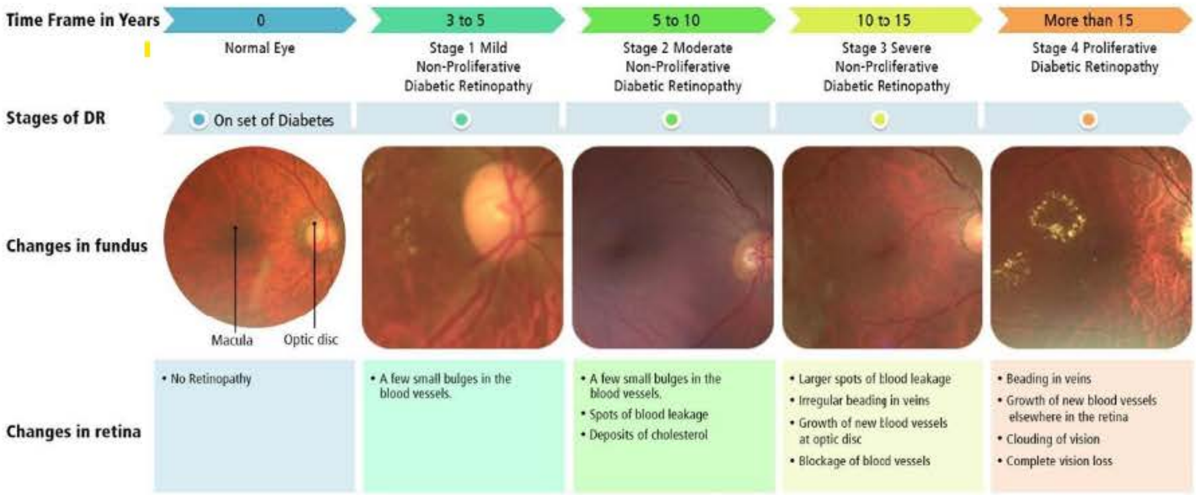

Figure 1.3: Example of retinal in each diabetic retinopathy stage.

at any stage of DR. In diabetic patients, the prevalence of different types of retinopathy increases with the age of diabetes and the patient's age. Diabetic retinopathy is rare before the age of 10. However, the risk of developing diabetic retinopathy increases after puberty.

A) Non-Proliferative Diabetic Retinopathy (NPDR):

This is the most common form of DR, also known as background retinopathy. The most common findings on fundus examination are signs of retinal ischemia (microaneurysms, hemorrhages, intraretinal microvascularization abnormalities) without the growth of new blood vessels. As illustrated in Table 1.1, the severity of NPDR is further categorized into mild, moderate, and severe stages, each with distinct clinical features that extent and severity of retinal lesions [33].

B) Proliferative Diabetic Retinopathy (PDR):

PDR is the advanced stage of DR and is mainly characterized by neovascularization, the growth of new, abnormal blood vessels on the retina's surface or optic disc. These new vessels are fragile and prone to bleeding, leading to severe visual complications that increase the risk of vision loss [37]. The PDR stage clinical features can be detailed by the following retinal lesions:

- **Neovascularization:** The primary feature of PDR is the development of new blood vessels in response to retinal ischemia. These vessels can grow on the retina, optic disc, or into the vitreous humor, the clear gel that fills the space between the lens and the retina. Neovascularization is driven by the upregulation of vascular endothelial growth factor (VEGF) and other angiogenic factors [10].

- **Vitreous Hemorrhage:** The new vessels in PDR are fragile and can rupture,

Tableau 1.1: Non-Proliferative Diabetic Retinopathy Grades and Clinical Features.

NPDR Grades	Clinical Features
Mild	• **Microaneurysms:** The earliest clinical sign, appearing as tiny red dots on the retina caused by localized capillary dilations. • **Retinal Hemorrhages:** Small, dot-blot hemorrhages occur within the deeper layers of the retina.
Moderate	• **Increased Microaneurysms and Hemorrhages:** More widespread microaneurysms and hemorrhages. • **Hard Exudates:** Lipid residues that appear as yellowish deposits resulting from leakage from damaged capillaries. • **Cotton Wool Spots (Soft Exudates):** Small, white, fluffy lesions caused by microinfarctions in the retinal nerve fibre layer due to capillary occlusion. These indicate localized ischemia and represent accumulated axoplasmic material within the nerve fibre layer [36].
Severe	• **Intraretinal Microvascular Abnormalities (IRMA):** Irregularly dilated capillaries that serve as a precursor to new vessel formation. • **Venous Beading:** Sausage-like retinal vessel dilations indicate significant retinal ischemia. • **Extensive Hemorrhages and Microaneurysms:** More numerous and widespread than moderate NPDR.

causing blood to leak into the vitreous humor. This bleeding can significantly impair vision and may require surgical intervention if not resolved spontaneously. Vitreous hemorrhage can present as sudden, painless vision loss or the appearance of dark spots or floaters in the visual field.

- **Tractional Retinal Detachment:** Scar tissue associated with neovascularization can contract and pull on the retina, causing it to detach from the underlying tissue. Tractional retinal detachment is a medical emergency and can lead to permanent vision loss if not treated promptly. Symptoms include flashes of light, a curtain-like shadow over the visual field, and a sudden increase in floaters.

- **Neovascular Glaucoma:** In some cases, neovascularization can occur in the anterior eye segment, forming new blood vessels on the iris and drainage angle. This can obstruct aqueous humor outflow, increasing intraocular pressure and neovascular glaucoma. Neovascular glaucoma is a painful and sight-threatening condition that requires prompt medical and surgical management.

1.1.3.2 Diagnosis of Diabetic Retinopathy

Diagnosing DR involves a comprehensive eye examination to assess the extent of retinal damage. Early detection through regular screenings is crucial for preventing vision loss. The primary methods for diagnosing DR include fundus examination and various imaging techniques.

A) Fundus Examination

Fundus examination is a critical component of the diagnostic process for DR. It involves visually inspecting the eye's interior surface, including the retina, optic disc, macula, and posterior pole. The examination can be conducted using several techniques:

- **Direct Ophthalmoscopy**: A handheld ophthalmoscope is used to view the retina through the pupil. This method allows for a direct view of the retinal structures but offers a limited field of view.

- **Indirect Ophthalmoscopy:** This technique uses a binocular indirect ophthalmoscope and a condensing lens, providing a wider field of view and a more detailed examination of the peripheral retina. It is beneficial for detecting proliferative DR and peripheral retinal changes.

- **Slit-Lamp Biomicroscopy with a Fundus Lens:** A slit-lamp biomicroscope combined with a specialized fundus lens allows for a detailed and magnified retina view. This method effectively identifies microaneurysms, hemorrhages, exudates, and neovascularization

During a fundus examination, ophthalmologists look for key signs of diabetic retinopathy, including microaneurysms, hemorrhages, exudates, cotton wool spots (Soft Exudates), Neovascularization and Retinal Detachment. Figure 1.4 illustrates these key signs in retinal lesions according to each DR stage.

B) Imaging Techniques

Several advanced imaging techniques complement the fundus examination to diagnose and monitor DR, providing detailed visualization of retinal structures and blood flow. Also, they are essential tools for guiding treatment decisions and assessing the response to therapy.

- **Fundus Photography:** Using a specialized fundus camera, this technique involves capturing high-resolution retina images. Fundus photography documents the presence

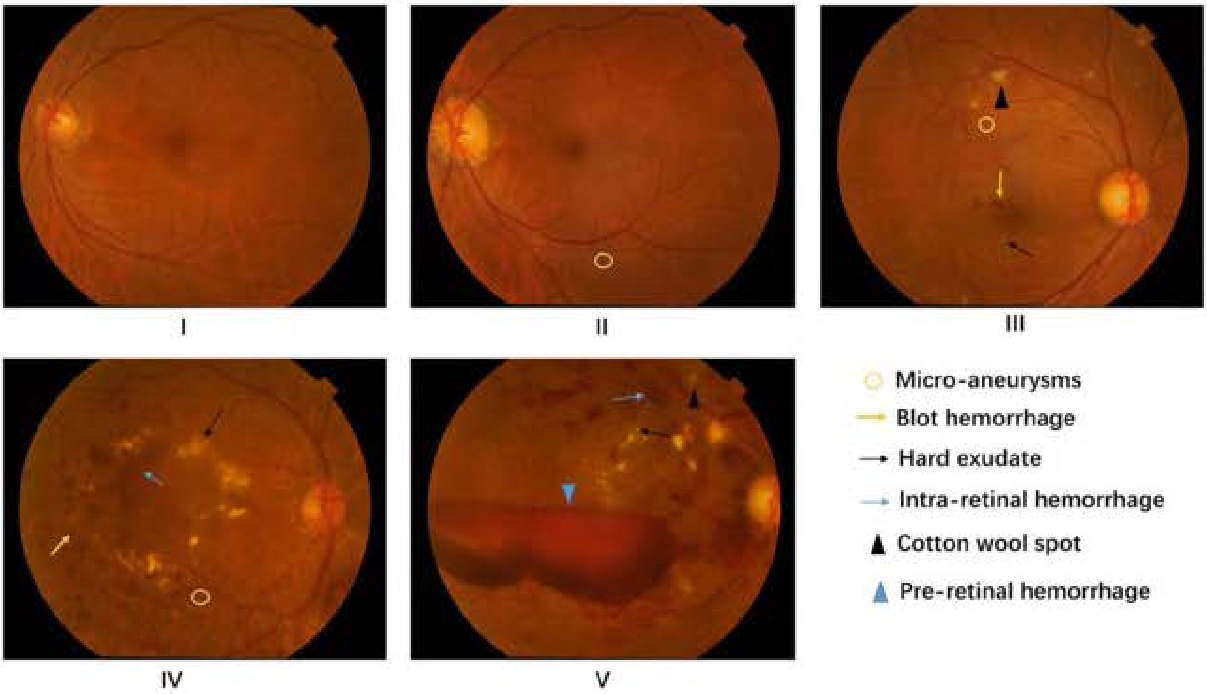

Figure 1.4: Different retinal lesions according to each diabetic retinopathy stage [1].

and progression of retinal lesions over time. It provides a permanent record that can be compared during follow-up visits [38].

- Fluorescein Angiography (FA): In this technique, a fluorescent dye (fluorescein) is injected into a vein in the arm. The dye travels to the retinal blood vessels, and photographs are taken as the dye circulates through the retinal vasculature. FA helps identify areas of leakage, capillary non-perfusion, microaneurysms, and neovascularization. It is beneficial for evaluating the extent of retinal ischemia and guiding laser treatment [39].

- Optical Coherence Tomography (OCT): OCT is a non-invasive imaging technique that uses light waves to create cross-sectional images of the retina. It provides detailed information about the retinal layers and is especially useful for detecting and monitoring diabetic macular edema (DME). OCT can reveal retinal thickening, cystoid spaces, and disruptions in the retinal architecture that are not visible on clinical examination [40].

- Optical Coherence Tomography Angiography (OCTA): OCTA is an advanced form of OCT that visualizes blood flow within the retinal and choroidal vasculature without dye injection. It provides high-resolution images of the retinal capillary networks and can detect early microvascular changes, capillary dropout, and neovascularization in DR [41].

- Ultrasound Imaging (B-Scan Ultrasonography): B-scan ultrasonography evaluates the retina in complex direct visualization, such as dense cataracts or vitreous hemorrhage. It helps detect retinal detachments, vitreous hemorrhage, and other posterior segment pathologies [42].

- Fundus Autofluorescence (FAF): FAF imaging captures the natural fluorescence emitted by certain retinal structures when exposed to specific wavelengths of light. It helps identify areas of retinal pigment epithelial (RPE) damage and atrophy, providing insights into the retina's health [43].

1.1.3.3 Treatment of Diabetic Retinopathy

The management of diabetic retinopathy (DR) involves a combination of laser treatments, pharmacologic therapies, and surgical interventions aimed at preventing further retinal damage, managing complications, and preserving vision. The choice of treatment depends on the stage of the disease and the specific retinal changes present.

A) Laser Photocoagulation

Laser photocoagulation is a well-established treatment for DR, particularly in the management of proliferative diabetic retinopathy (PDR).

- Panretinal Photocoagulation (PRP): is used primarily for treating PDR. The procedure involves applying laser burns to the peripheral retina to reduce the demand for oxygen and decrease the stimulus for neovascularization. By destroying the peripheral ischemic retina, PRP helps to prevent the growth of new, abnormal blood vessels and reduces the risk of vitreous hemorrhage and retinal detachment [44].

- Focal/Grid Laser Photocoagulation: This technique targets specific leaking microaneurysms, while grid photocoagulation is applied to areas of diffuse retinal thickening. The laser seals leaking blood vessels, reduces fluid accumulation in the macula, and stabilizes or improves vision [45].

B) Anti-Vascular Endothelial Growth Factor (VEGF) Therapy

This therapy has significantly advanced the treatment of DR, key complications of diabetes leading to vision loss. By inhibiting VEGF, these therapies reduce abnormal blood vessel growth and fluid leakage in the retina, thus improving visual acuity and reducing macular swelling. Clinical trials have demonstrated their efficacy, establishing anti-VEGF therapy as a first-line treatment for preventing DR progression [46]. Despite challenges like cost and the need for regular injections, anti-VEGF therapy remains a revolutionary approach, significantly enhancing vision preservation and quality of life for diabetic patients.

C) Surgical Interventions

Surgical interventions are reserved for advanced cases of DR that do not respond to less invasive treatments or when complications arise. Two surgical procedure can be performed.

- Vitrectomy: Vitrectomy is a surgical procedure that involves the removal of the vitreous gel from the eye. It is indicated in cases of non-clearing vitreous hemorrhage, tractional retinal detachment, and severe proliferative diabetic retinopathy with extensive fibrovascular proliferation. The procedure helps to clear hemorrhages, relieve traction on the retina, and allow for further laser treatment if necessary [47].

- Endolaser Photocoagulation: During a vitrectomy, endolaser photocoagulation can be performed to treat retinal neovascularization. This approach allows for precise application of laser therapy to areas that may not be accessible with conventional laser techniques [48].

1.2 Deep Learning in Medical Image Analysis

1.2.1 Introduction to Deep Learning

Deep Learning (DL) is a subfield of machine learning. It has emerged as a powerful tool for tackling complex tasks in various domains, including medical image analysis. DL is inspired by the structure and function of the human brain and involves the use of artificial neural networks with multiple layers to learn hierarchical representations of data. This powerful technique has demonstrated remarkable success in tasks such as image recognition, natural language processing, and speech recognition, among others.

The origins of DL can be traced back to the 1940s-1960s when artificial neural networks were first proposed as a model for biological brain activity [49]. Early ideas in DL were introduced in the 1960s [50]. However, due to the computational and conceptual limitations of that period, models that could take advantage of these ideas were not developed. The popular wave of deep learning came in the 1980s when batch processing and unsupervised learning entered the scene. In the late 90s and early 2000s when the popularity of neural networks started to decrease mainly due to the difficulties in training models and the lack of adequate data sets [51].

From 2011, DL models outpaced the performance of traditional machine learning methods in computer vision tasks and more general domains. This real revival of DL is driven by three key factors: the availability of large-scale datasets (such as ImageNet [52]), advances in computing power (especially GPUs), and improvements in network architectures (like CNNs). In which researchers demonstrated that DL networks could outperform traditional machine learning methods in tasks such as image and speech recognition [53, 54].

Today, DL is at the forefront of AI research and applications. Breakthroughs in architectures (like transformers and generative models), coupled with vast computational resources and large-scale datasets, have enabled deep learning models to achieve extraordinary performance across various domains, including healthcare, autonomous driving, and more.

1.2.1.1 Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) form the milestone of deep learning models. They are computational models inspired by the biological neural networks in the human brain. They are designed to mimic the way the brain processes information by interconnecting a large number of processing units (neurons), organized in layers that enable the network to learn from data and make predictions [2].

At the core of an ANN is the idea of learning from data by adjusting the strengths of the connections between neurons, known as weights. These weights determine the influence of one neuron's output on another neuron's input, allowing the network to capture and represent complex patterns and relationships within the data.

a) Basic Structure

The fundamental structure of an ANN is illustrated in Figure 1.5. It typically comprises three main layers: the input layer, the hidden layer(s), and the output layer.

The input layer is responsible for receiving the data or features that are to be processed by the network [55]. These input features are then passed through the hidden layers, where a series of mathematical operations are performed to transform the data, allowing the network to learn complex patterns and relationships within the data [56]. The hidden layers are composed of a variable number of neurons, which apply various activation functions to the weighted sum of their inputs, producing outputs that are then passed to the next layer till reach the output layer that provides a decision [55].

The choice of the number and size of hidden layers, as well as the specific activation functions used, are critical design decisions that can significantly impact the performance and capabilities of the ANN [57].

b) Activation Functions

Activation functions play a crucial role in introducing non-linearity into neural networks, enabling them to model and learn complex relationships within the data. Without activation functions, neural networks would essentially become linear models, severely limiting their capability to solve intricate problems [58].

The role of activation functions is to apply a non-linear transformation to the weighted sum of inputs received by a neuron. This non-linear transformation allows the neural network to capture and represent non-linear patterns present in the data. Different activation functions have distinct properties and characteristics, making them suitable for different types of problems and network architectures [59]. Table 1.2 summarizes commonly used activation functions in neural networks.

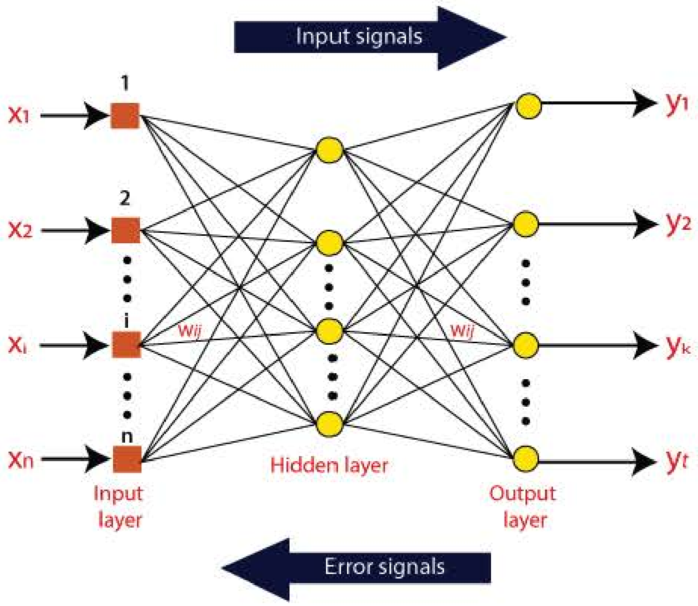

Figure 1.5: Architecture of the Artificial Neural Network [2].

c) Backpropagation Algorithm

Backpropagation is a fundamental algorithm for training ANNs. It involves a forward pass, where inputs are propagated through the network to generate predictions, and a backward pass, where the error is propagated back to update the weights [60].

In the forward pass, inputs are fed into the input layer, and the network computes outputs layer by layer until reaching the output layer. The error or loss is then calculated by comparing the output to the true labels using a loss function, such as mean squared error for regression or cross-entropy for classification.

The network's weights are adjusted then in the backward pass based on the computed error. This process uses the gradient of the loss function concerning each weight, obtained through the chain rule of calculus. This gradient is multiplied by a learning rate, a small constant that controls the step size in the weight update process. The iterative process of forward and backward passes continues until the network converges to a state where the loss is minimized. The weight update rule can be summarized by the following equation:

$$w_{new} = w_{old} - \eta \frac{\delta L}{\delta w} \tag{1.1}$$

where w represents the weight, η is the learning rate, and $\frac{\delta L}{\delta w}$ is the gradient of the loss function with respect to the weight.

Tableau 1.2: Comparison of Common Activation Functions

Activation Function	Formula	Output Range	Advantages	Desadvantages
Sigmoid	$\sigma(x) = \frac{1}{1+e^{-x}}$	(0, 1)	- Useful in binary classification tasks. - Outputs probabilities in the output layer.	- Vanishing gradient problem for large inputs. - Outputs not zero-centered, leading to slower convergence.
ReLU	$\text{ReLU}(x) = \max(0, x)$	$(0, \infty)$	- Linear for positive inputs, zero for negative inputs. - Computationally efficient.	- "Dying ReLU" problem where neurons can become inactive. - Outputs not zero-centered, causing imbalance.
Tanh	$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	(-1, 1)	- Smooth, S-shaped curve. - Maps input to range (-1, 1).	- Vanishing gradient problem for large inputs. - Slightly more complex computation than ReLU.
Softmax	$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}}$ for $i = 1, \ldots, K$	(0, 1) for each class	- Converts logits into a probability distribution over multiple classes.	- Not used in hidden layers. - Can be computationally intensive for a large number of classes.

1.2.1.2 Deep Neural Networks (DNNs)

ANNs have been successful in various applications. However, for complex problems involving large datasets and intricate patterns, traditional ANNs can be limited in their learning capacity and computational efficiency. This limitation has led to the development of deep neural networks (DNNs), which are ANNs with multiple hidden layers and specialized architectures, such as Multi-Layer Perceptrons (MLPs) and Convolutional Neural Networks (CNNs).

a) Multi-Layer Perceptrons (MLPs)

Multi-Layer Perceptrons (MLPs) are a subclass of ANNs that form the foundational architecture for many DL models. They consist of multiple layers of neurons, including input, hidden, and output layers, where each layer is fully connected to the next. MLPs are particularly significant for their ability to model complex, non-linear relationships

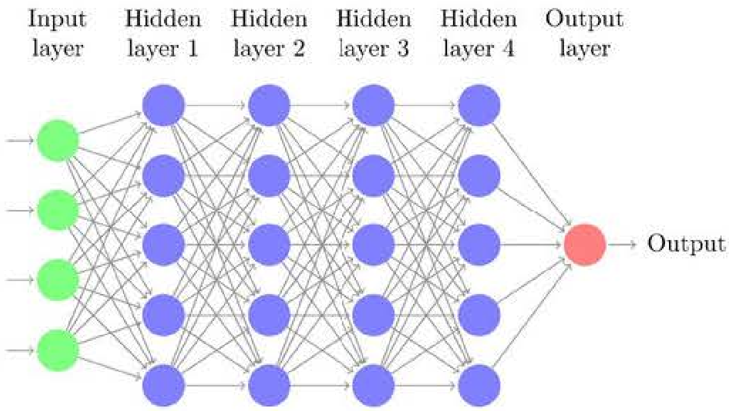

Figure 1.6: Illustration of a Multi-layer Perceptrons Architecture [3].

within data.

As shown in Figure 1.6, MLPs are composed of an input layer that receives the data, one or more hidden layers that process the data, and an output layer that produces the final result. Each layer is fully connected to the next, meaning every neuron in one layer is connected to every neuron in the subsequent layer. This dense connectivity allows MLPs to capture complex patterns in the data [61].

While MLPs are powerful for modelling a variety of tasks, they have limitations when dealing with structured data like images or sequences. For instance, they do not inherently exploit the spatial structure of images, which is where specialized architectures like Convolutional Neural Networks (CNNs) come into play. Nevertheless, MLPs are highly effective for tasks such as regression, classification, and simple pattern recognition, and they serve as the basis for understanding more complex deep learning architectures.

b) Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a specialized type of DNN designed to process grid-like data structures, particularly images. They leverage the concepts of convolution and pooling to capture spatial hierarchies and reduce the dimensionality of the data, making them more efficient and scalable for image analysis.

The CNN architecture is composed of several successive layers (Figure 1.7). There are three main types of layers that are generally observed in the CNNs architectures: Convolutional Layers (Conv), the Pooling Layer (Pool), and the Fully-Connected Layers (FCLs).

- **The Convolutional Layer:** performs a specific function of transformation on local regions in the input (receipt field) to obtain a useful representation. It functions as a feature extractor. An input image is passed through a series of sliding learnable convolution

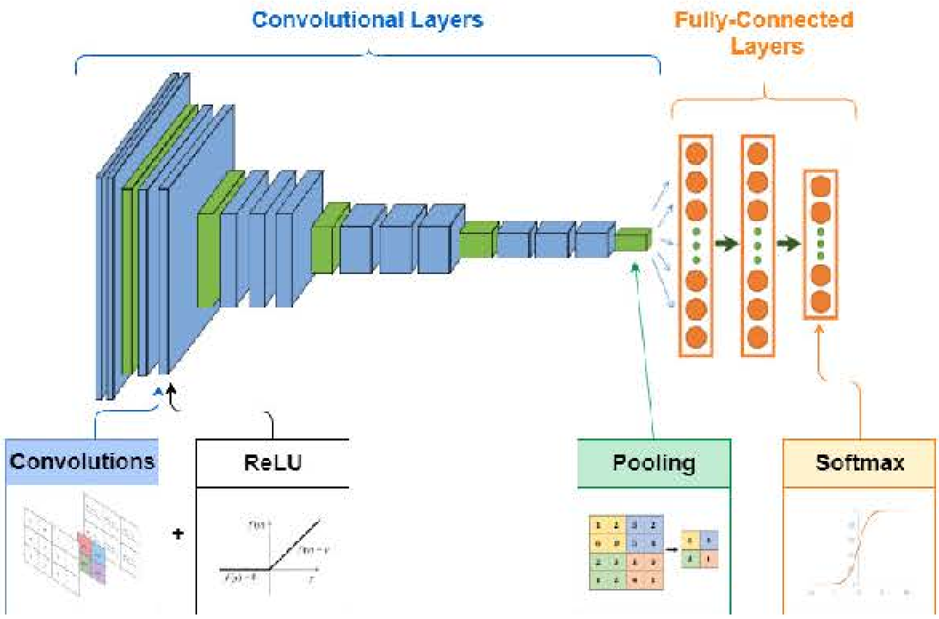

Figure 1.7: Illustration of CNN architecture: example of VGG16 model.

kernels (filters), creating as a result 3-dimensional convolution feature maps (Fmaps) (see Figure 1.8). The feature maps values are produced using the neuron activation function that can be defined by:

$$f(x) = \sum_{i=1}^{s} w_i x_i + b \tag{1.2}$$

where x_i ,w_i and b are the convolutional input values (receipt field), the weights (filter values), and the bias, respectively. s represents the filter size.

In addition, a correction operation called Rectified Linear Unit (ReLU) is also applied to the obtained feature maps. ReLU is an element-wise operation defined by (Equation 1.3). The output feature maps have non-negative values.

$$f(x) = Max(0, x) \tag{1.3}$$

- The Pooling Layer: performs a sub-sampling operation, by reducing the spatial dimensions (i.e the height and the width) of the intermediate feature maps and retaining the most important information. The pooling is an important concept for CNNs since it aims to reduce the size of the feature maps in order to minimize the number of parameters and the computation operations in the network. The pooling is generally operated as a Max, Average, or Sum function on every depth slice of the input feature maps independently. Whereas the depth dimension d is still unchanged, the height and the width dimensions of the depth slice are down-sampled using pooling filters (Figure 1.9). The

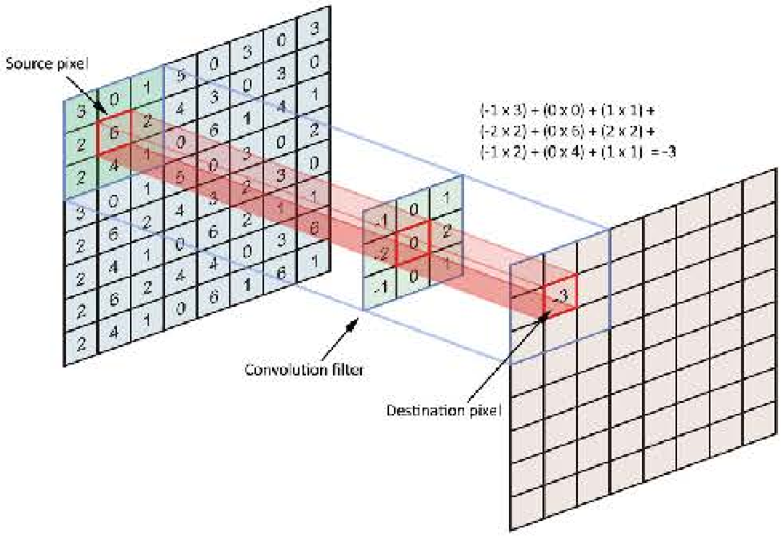

Figure 1.8: Principle of filter sliding in the convolution layer over an image.

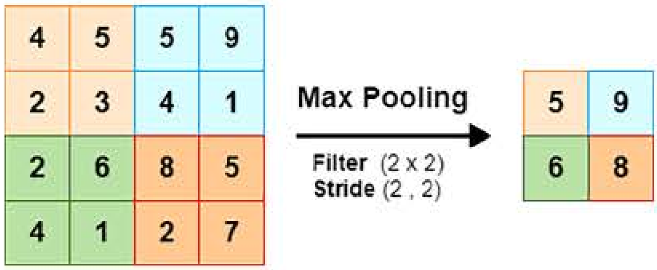

Figure 1.9: Pooling layer principle: example of performing Max pooling function.

output of this layer produces typically a 3-dimensional feature maps of the dimensions $(w \times h \times d)$ which can be defined by:

$$w = \frac{w_1 - F}{S} + 1 \tag{1.4}$$

$$h = \frac{h_1 - F}{S} + 1 \tag{1.5}$$

Where w_1, h_1, and d are the input width, height, and depth respectively, S is the stride, and F is the spatial extent. For the two pooling hyper-parameters F and S, they are commonly used in two variations: $F = 2$ and $S = 2$ as well as the overlapping pooling in which $F = 3$ and $S = 2$.

- Fully-Connected Layers (FCLs): As its name signifies, is a feed-forward neural network in which all neurons are connected to all the neurons of the next layer and have connections with all previous layer neurons (Figure 1.10. As shown in Equation 1.2, the FCLs neuron activation function can be computed using matrix multiplication added to bias offset.

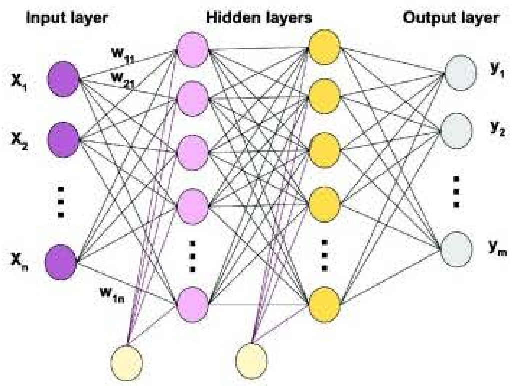

Figure 1.10: Illustration of Fully-Connected Layers structure [4].

Adding FCLs able the CNN model for end-to-end learning [62]. More precisely, after feature generation, we need for a decision. Thus, the obtained high-level features from the conv layer are fed into the FCLs structure that learns the non-linear combinations in that feature space. Over a series of back-propagation epochs, the neurons weights are determined for optimising the Softmax function. Finally, the last FCL outputs the final decision.

1.2.2 Deep Learning for Image Classification

Image classification is a core task in computer vision that involves categorizing images into predefined classes based on their visual content. This process is essential in numerous fields, including medical imaging, where classification involves assigning a label to an entire image or specific regions within it, often to identify the presence or absence of a particular disease or condition, such as identifying whether an X-ray shows signs of pneumonia or a retinal image indicates diabetic retinopathy. Thus, image classification plays a crucial role in automating diagnosis, supporting clinical decision-making, and improving patient care.

Image classification can be categorized into different types based on the number of classes and the nature of the labels:

- **Binary Classification:** The image is assigned to one of two possible categories, such as "normal" or "abnormal."
- **Multi-class Classification:** The image is assigned to one of multiple mutually exclusive categories, such as different types of tumors or different stages of a disease.
- **Multi-label Classification:** The image can be assigned to multiple labels simultaneously, indicating the presence of multiple features or conditions.

1.2.2.1 Deep Learning Techniques for Image Classification

With the advancements in deep learning and the availability of large-scale datasets, Deep learning has significantly advanced image classification tasks, achieving state-of-the-art performance across various applications. Deep neural networks have proven highly effective in learning complex features directly from raw image data. However, choosing an appropriate deep learning architecture is crucial for effective image classification. Different architectures are designed to handle the intricacies of image data and can significantly impact the performance and efficiency of the model.

A) Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are the most widely used model architectures for image classification tasks. Over time, more sophisticated architectures have been developed to improve the capabilities of CNNs. These architectures have the ability to automatically learn powerful image representations from the input images.

The successful results of CNNs on image classification task has motivated the research contributions in network architectural design. AlexNet [51] is one of the earliest successful CNNs, AlexNet achieved groundbreaking performance on the ImageNet challenge in 2012[1]. It introduced key concepts like ReLU activation functions and dropout regularization.

Since then, several CNN architectures have been proposed to achieve the stat-of-the-art performances on ImageNet dataset for different ILSVRC competition tasks, whereby each CNN architecture has tried to address the shortcomings of previous CNN architectures by adding new structural reformulations or by exploring different strategies for parameter optimization to improve the CNNs performance and reduce the computational cost. Figure 1.11 summarizes the history of the CNN architectures evolution [63].

The full training of CNN models is a high computational cost process and requires a huge amount of labelled data. Thus, several studies have examined the generalization power of the CNN architecture, demonstrating the transferability of the CNN models that are trained upon ImageNet dataset. These pre-trained CNN models are able to serve as the backbone for other recognition tasks on other datasets. In the literature, the most cited CNN networks belong to the three families: VGG, Resnet, and Inception khan2020survey.

- **VGGNet:** VGGNet [64] further improved upon AlexNet by using smaller convolutional filters and stacking them deeper, leading to better feature extraction and classification accuracy.

[1] `http://www.image-net.org/challenges/LSVRC/2012/`

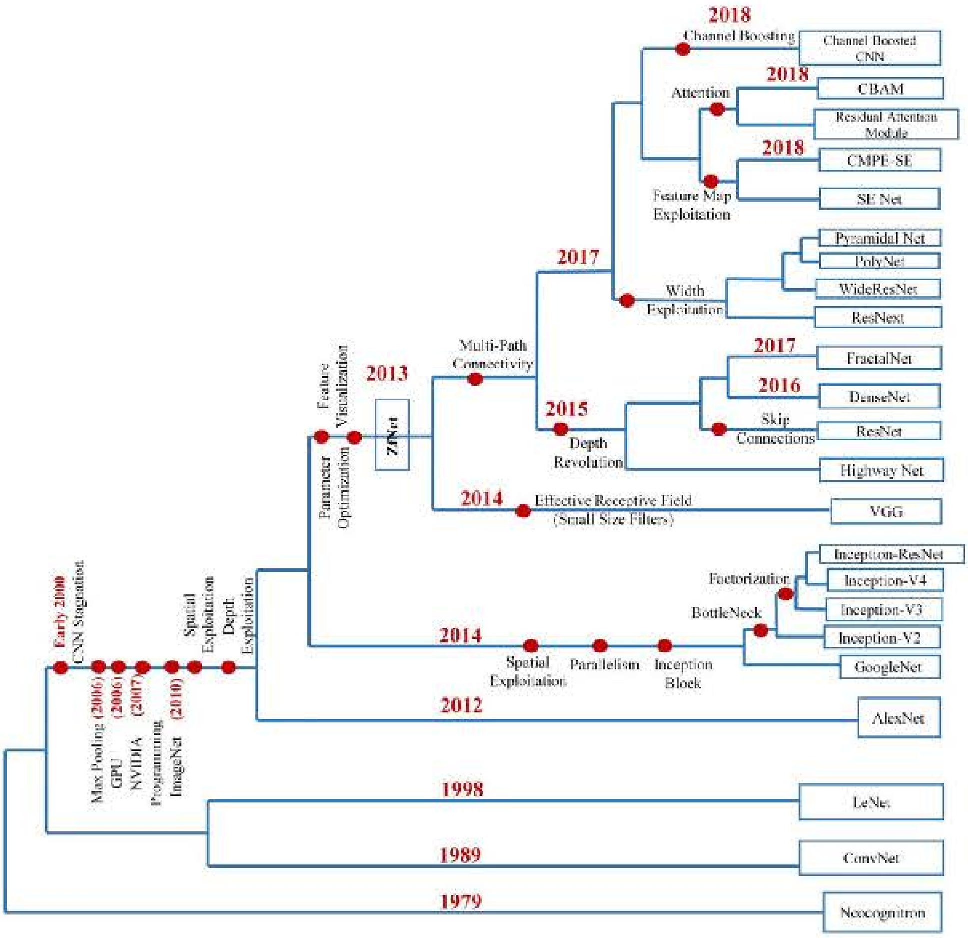

Figure 1.11: History of The CNN architectures evolution khan2020survey.

- **ResNet:** ResNet [65] addressed the vanishing gradient problem in deep networks by introducing residual connections, allowing information to flow more efficiently through the network.

- **Inception Networks:** Inception networks [66] introduced a novel architecture with multiple parallel convolutional paths, enabling the network to learn features at different scales and resolutions.

In medical image classification task, transfer learning allows leveraging the power of these pre-trained models on large datasets, such as ImageNet, to fine-tune them on smaller medical image datasets. This significantly reduces the need for large amounts of labeled medical data, making deep learning more accessible for medical applications.

B) Attention Mechanisms

While deep learning architectures like CNNs have significantly advanced image classification, they often struggle with capturing long-range dependencies and focusing on the most relevant parts of an image. Attention mechanisms, inspired by the human visual system's

Tableau 1.3: Types of Attention Mechanisms for Image Classification

Type of Attention	Description	Benefits
Self-attention	Allows the model to attend to different parts of the same input feature map, capturing long-range dependencies and relationships between different regions.	- Captures global context and long-range dependencies. - Improves performance on images with complex structures.
Channel Attention	Focuses on the importance of different feature channels, allowing the model to selectively enhance or suppress specific features that are more relevant to the classification task.	- Enhances relevant features and suppresses irrelevant ones. - Improves robustness to noise and variations.
Spatial Attention	Focuses on the spatial location of features, allowing the model to prioritize certain regions of the image based on their spatial context.	- Emphasizes spatially relevant features. - Improves localization of objects or lesions.

ability to selectively focus on specific regions, address these limitations by allowing the model to dynamically weigh the importance of different parts of the input [67].

By selectively focusing on the most relevant regions and features, attention mechanisms can significantly improve classification accuracy, especially for complex or challenging datasets [68]. This feature can help models become more robust to distracting parts of the image ignoring irrelevant or noisy information, improving its robustness to real-world variations

Similar to CNNs, attention mechanisms trained on large datasets can often be effectively transferred and fine-tuned for specific image classification tasks with limited data. Whereas other DNN architectures suffer from the lack of interpretability of their predictions, Attention maps can provide insights into which regions or features the model considers important for a particular classification decision, enhancing the model interpretability [69].

Several types of attention mechanisms have been developed for image classification, Table 1.3 summarize the three main attention mechanism types and their benefits.

1.2.2.2 Evaluation Metrics for Classification Models

Evaluating the performance of classification models is crucial to understanding their effectiveness and reliability. In the context of image classification, metrics provide quantitative measures to assess how well a model distinguishes between different classes. This is especially important in medical imaging, where accurate and reliable classification can have significant implications for diagnosis and treatment.

Choosing the appropriate evaluation metric depends on the specific requirements and constraints of the classification task. In medical imaging, where the consequences of misclassifications are significant, it is crucial to balance between metrics such as precision, recall, and F1-score, and consider metrics like AUC-ROC for a comprehensive evaluation.

A) Confusion Matrix

The confusion matrix summarizes the number of correct and incorrect classified samples break-downing by each class showing the classes where the method makes confused predictions. The confusion matrix is a $K \times K$ matrix where K is the number of classes and S is the total number of tested images in the dataset. Each cell C_{ij} represents the total number of samples that belong to i^{th} class (ground truth) and are misclassified as j^{th} class. All correctly classified samples lie on the diagonal of the matrix $C_{ij} i = j$. The sum of all the elements of the matrix is equal to the total number of the testing samples S such that:

$$S = \sum_{i=1}^{K} \sum_{j=1}^{K} C_{ij} \tag{1.6}$$

B) Accuracy, Precision, Recall, and F1 Score

- Accuracy: is one of the most commonly used evaluation metrics for classification tasks. It represents the proportion of correctly classified instances among the total instances in the dataset as follows:

$$Accuracy = \frac{Number\, of\, Correct\, Predictions}{Total\, Number\, of\, Predictions} \tag{1.7}$$

Higher accuracy indicates that the model correctly classifies most of the images. However, in cases of imbalanced datasets where one class is much more frequent than others, accuracy can be misleading. For example, if 90% of images belong to one class, a model that always predicts this majority class will have 90% accuracy, even though it is ineffective at distinguishing the minority class.

- Precision: measures the proportion of correctly predicted positive instances among all instances predicted as positive. It is a critical metric when the cost of false positives is high. precision can be computer by the following equation:

$$Precision = \frac{True\,Positives\,(TP)}{True\,Positives\,(TP) + False\,Positives\,(FP)} \tag{1.8}$$

Higher precision indicates that the model has a low false positive rate and is reliable in its positive predictions. This precision is useful in medical imaging when confirming a diagnosis is critical, and false positives can lead to unnecessary treatments.

- Recall (Sensitivity or True Positive Rate): measures the proportion of correctly predicted positive instances among all actual positive instances. It is crucial when the cost of false negatives is high. The recall metric is formulated as below:

$$Recall = \frac{True\,Positives\,(TP)}{True\,Positives\,(TP) + False\,Negatives\,(FN)} \tag{1.9}$$

A higher recall value indicates that the model effectively identifies most of the positive instances. This metric is important in scenarios where missing a positive case (e.g., failing to detect a disease) has severe consequences.

- F1-Score: is the harmonic mean of precision and recall, providing a single metric that balances the two. It is particularly useful when the dataset is imbalanced and a balance between precision and recall is needed.

$$F1\,Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{1.10}$$

The high F1-Score indicates a good balance between precision and recall. This metric is valuable in medical imaging where both detecting diseases (recall) and minimizing false alarms (precision) are important.

Specificity (True Negative Rate): measures the proportion of correctly predicted negative instances among all actual negative instances. It is the complement of recall for the negative class.

$$Specificity = \frac{True\,Negatives\,(TN)}{True\,Negatives\,(TN) + False\,Positives\,(FP)} \tag{1.11}$$

Higher specificity indicates that the model is effective at identifying negative instances and has a low false positive rate. The specificity is important in medical imaging to ensure healthy cases are not wrongly classified as diseased

C) Area Under the ROC Curve (AUC-ROC)

AUC-ROC is a performance measurement for classification problems at various threshold settings. ROC (Receiver Operating Characteristic) curve plots the true positive rate (recall) against the false positive rate (specificity). this metric can be computer using the following formula:

$$AUC\ ROC = \int_0^1 ROC\ Curved\,(False\ Positive\ Rate) \tag{1.12}$$

High AUC-ROC indicates that the model is good at distinguishing between the positive and negative classes across different thresholds. It can be used in evaluating the model's performance over a range of decision thresholds, particularly when the classes are imbalanced.

1.2.3 Deep Learning for Image Segmentation

Image segmentation is a fundamental task in computer vision that aims to divide an image into distinct regions based on certain criteria, such as color, texture, or shape. The resulting segmentation map assigns each pixel to a specific class or label, representing different structures or regions of interest allowing for detailed analysis and interpretation of the image. In medical imaging, this task is particularly critical in several applications for identifying anatomical structures, detecting abnormalities, and quantifying disease progression, in which image segmentation can isolate regions such as tumors, organs, or blood vessels, facilitating diagnosis, treatment planning, and monitoring [5].

Three main image segmentation types can be distinguished:

1. **Semantic Segmentation:** Classifies each pixel into a category from a set of predefined classes. All pixels belonging to the same class are labeled identically. It focuses on identifying and classifying different objects or structures within an image, , for instance, in a retinal image, pixels may be classified into categories like lesion, and background.

2. **Instance Segmentation:** Extends semantic segmentation by distinguishing between different instances of the same object class. Each instance is segmented separately, allowing for differentiation between, for example, multiple lesions in a retinal scan.

3. **Panoptic Segmentation:** Combines semantic and instance segmentation by providing a complete scene understanding. It classifies all pixels into both classes and instances, ensuring that every pixel is accounted for in the segmentation process.

1.2.3.1 Deep Learning Techniques for Image Segmentation

A) Fully Convolutional Networks (FCNs)

Fully Convolutional Networks (FCNs) are CNN-based architectures specifically designed for dense prediction tasks like image segmentation. Unlike traditional CNNs, FCNs replace fully connected layers with convolutional layers, preserving the spatial structure of the input image.

As shown in Figure 1.12, the architecture of FCNs consists of a sequence of convolutional and pooling layers followed by up-sampling layers. The up-sampling layers, also known as deconvolutional layers, increase the resolution of the feature maps to produce a segmentation map that matches the input image size. The final output of an FCN is a pixel-wise prediction map, where each pixel is assigned a class label, enabling precise segmentation [5].

B) Mask R-CNN

Mask R-CNN extends the capabilities of the Faster R-CNN architecture for instance segmentation, allowing it to detect and segment each object instance in an image [6]. As illustrated in Figure 1.13, the Mask R-CNN architecture consists of three main blocks:

- **Region Proposal Network (RPN):** Generates candidate object bounding boxes (regions of interest, or RoIs) that likely contain objects.
- **RoIAlign:** Refines the regions to ensure accurate spatial alignment, crucial for precise segmentation.
- **Segmentation Head:** Adds a small FCN to predict a binary mask for each proposed region, segmenting the object instance within its bounding box.

C) U-Net Architecture

The U-Net architecture has become one of the most influential models in medical image segmentation due to its ability to produce precise and reliable segmentations with limited training data.

The U-Net architecture consists of a symmetric encoder-decoder structure with skip connections, forming a "U" shape (see Figure 1.14). This structure allows for efficient and precise segmentation by combining local and global contextual information [7].

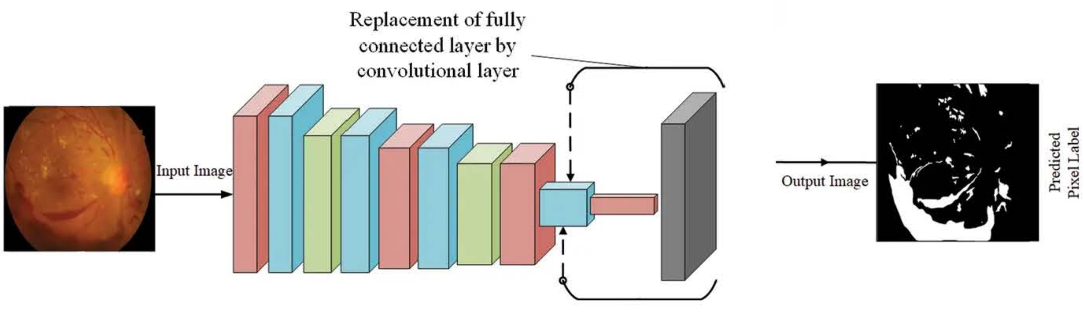

Figure 1.12: Architecture of a fully convolutional network [5].

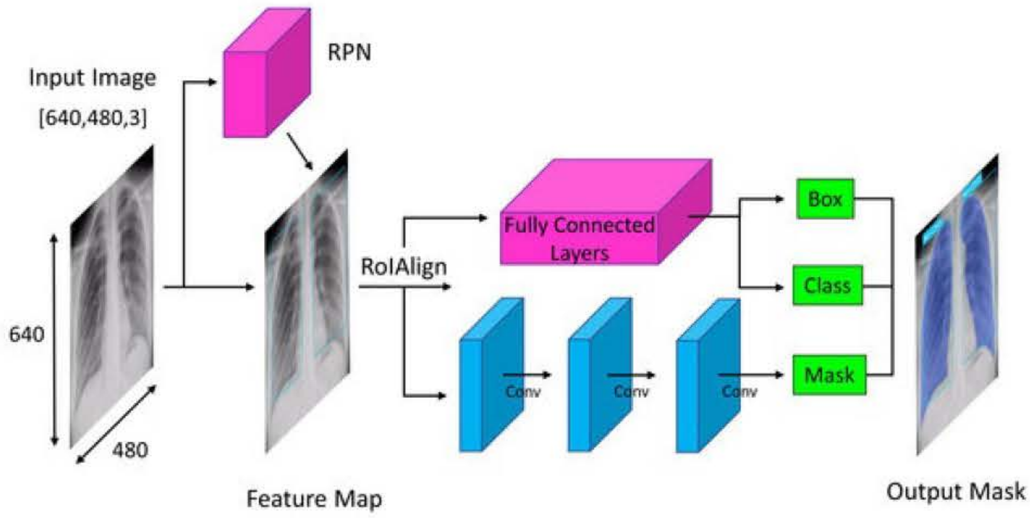

Figure 1.13: Architecture of a Mask R-CNN model [6].

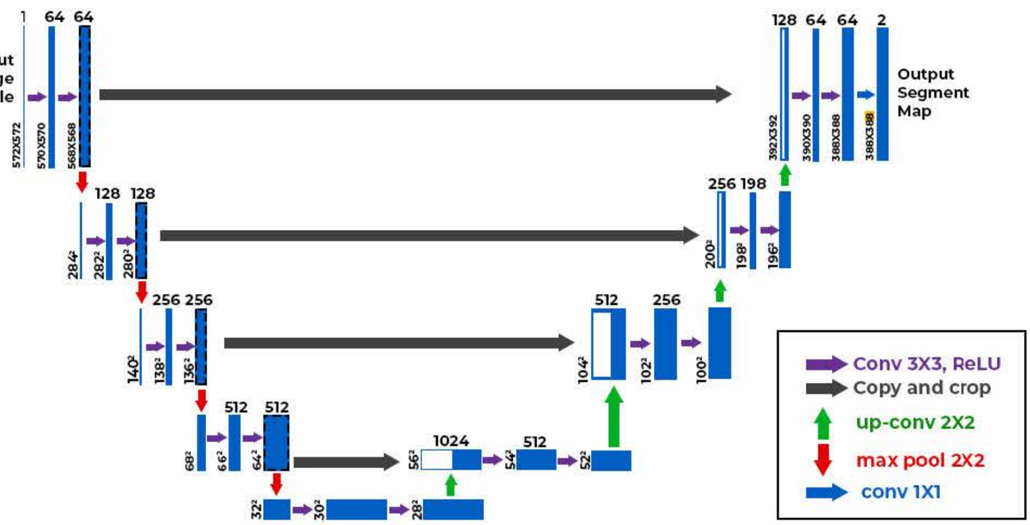

Figure 1.14: Standard U-Net architecture [7].

- **Encoder-Decoder Structure:** U-Net consists of an encoder path that compresses the input image into a compact representation and a decoder path that reconstructs the segmentation map. The encoder captures context, while the decoder refines the segmentation to the original resolution.

- **Skip Connections:** U-Net uses skip connections between corresponding layers in the encoder and decoder. These connections transfer fine-grained details from the encoder to the decoder, enhancing the model's ability to produce detailed and accurate segmentations.

1.2.3.2 Evaluation Metrics for Segmentation Models

Evaluating segmentation models requires specialized metrics that go beyond simple accuracy, as these models predict dense, pixel-level information rather than discrete class labels. The effectiveness of a segmentation model is determined by how well the predicted segments match the ground truth regions in an image. In medical imaging, where precise delineation of structures such as organs, tumors, or lesions is critical, choosing the right evaluation metrics is essential for assessing model performance accurately.

For P represents the set of pixels in the predicted segment and G is the set of pixels in the ground truth segment, $|P|$ and $|G|$ are the number of pixels in the predicted segment and the ground truth segment respectively. Also, $|P \cap G|$ represents the number of pixels common to both P and G where $|P \cup G|$ is the total number of pixels in either P and G. The following metrics are the most commonly used for evaluating the segmentation performance of the predicted segment.

A) Intersection over Union (IoU)

Intersection over Union (IoU), also known as the Jaccard Index, is widely used for evaluating segmentation tasks, providing a balanced measure of accuracy and completeness. It measures the overlap between the predicted segmentation and the ground truth. The IoU metric can be calculated using the below formula where high IoU indicates a strong overlap between the predicted and ground truth segments.

$$IoU = \frac{|P \cap G|}{|P \cup G|} \tag{1.13}$$

B) Dice Coefficient

The Dice Coefficient (F1 Score), is another metric for measuring the overlap between two sets. It is similar to IoU but tends to be more sensitive to the sizes of the segments. This metric is particularly useful in medical imaging where it is crucial to maximize the overlap between predicted and actual regions. The high Dice Coefficient reflects a strong correspondence between the predicted and ground truth segments. It can be formulated as:

$$Dice\,Coefficient = \frac{2\,|P \cap G|}{|P| + |G|} \tag{1.14}$$

C) Precision and Recall for Segmentation

Precision and Recall can be adapted for segmentation tasks to evaluate the correctness of the positive predictions and the model's ability to capture all relevant regions, respectively. Balancing precision and recall is essential in medical imaging to ensure both accurate and complete segmentations. High precision indicates that a high proportion of the predicted segment pixels are correctly labelled, where high recall indicates that a high proportion of the ground truth segment pixels are correctly captured by the prediction. Precision and recall metrics are computed by the following equations

$$Precision = \frac{|P \cap G|}{|P|} \tag{1.15}$$

$$Recall = \frac{|P \cap G|}{|G|} \tag{1.16}$$

Conclusion

This chapter has established the foundational concepts for understanding the subsequent exploration of diabetic retinopathy (DR) detection. Initially, we examined the medical background, detailing the impact of diabetes on global health and its specific impacts on ocular health, leading in DR. We detailed the anatomy of the eye and the mechanisms by which diabetes causes retinal damage, outlining the progression and symptoms of DR, as well as current treatment methodologies.

Subsequently, we transitioned to the technological context, introducing Deep Learning (DL) as a key tool in modern medical image analysis. We underlined the importance of image classification and segmentation tasks in the context of DR detection, illustrating how these techniques can enhance diagnostic accuracy and efficiency.

Having bridged the medical and technological domains, a background is prepared to proceed to the next chapter. In Chapter 2, we explore deeper into the existing literature on deep learning-based diabetic retinopathy detection. By reviewing the most relevant research, we will gain insights into the current state-of-the-art techniques, identify key challenges, and explore potential avenues for future research. This comprehensive review will provide a solid foundation for the subsequent chapters, where we will present our proposed methodology and contribute to the advancement of this critical field.

Chapter 2

State of the Art

Introduction

Diabetic Retinopathy (DR) as detailed in the previous chapter, remains a leading cause of vision impairment globally, necessitating robust screening and early detection methods. Recent advancements in medical image analysis, particularly through deep learning techniques, have advanced the automated diagnosis of DR from retinal fundus images. This chapter presents a comprehensive review of the most relevant research in retinal image segmentation and DR classification, focusing on two pivotal aspects: the delineation of retinal structures and the identification of pathological changes indicative of DR progression.

The first section examines state-of-the-art methods in retinal image segmentation, with particular emphasis on blood vessel extraction and retinal lesions identification. These techniques form the foundation for advanced DR detection systems, enabling the precise localization and quantification of retinal abnormalities. We explore the evolution from traditional fully convolutional networks to more sophisticated architectures, including variants of U-Net and region-based convolutional neural networks, each offering unique advantages in addressing the inherent challenges of retinal image analysis.

The second part of this chapter delves into recent methods for DR classification, highlighting end-to-end learning paradigms and lesion-based methodologies. We review different architectural networks, training strategies, and performance optimizations that have significantly enhanced the accuracy and reliability of automated DR grading systems. This comprehensive overview aims to explain the current landscape of DR detection techniques, providing a solid framework for understanding the field's trajectory and identifying different axes for future research and clinical application.

2.1 Retinal Image Segmentation

The fundus image is the area where arteries, veins, and capillaries can be observed directly. Changes in the retinal vessels in the fundus alert us to the onset of many fundus and cardiovascular diseases, like diabetic retinopathy. Consequently, retinal vessel segmentation is crucial for the diagnosis and screening of various diseases. Retinal image segmentation is the process of partitioning an image into meaningful sections, which enables clinicians and automated systems to isolate and analyze distinct anatomical structures within the retina, such as blood vessels and lesions. This process is crucial for identifying abnormalities and tracking disease progression.

The breaking down of the retinal images into detailed segments gives clearer insights into the state of the eye identifying pathological changes and facilitates the monitoring of disease progression. This section delves into the relevant studies used in retinal image segmentation including retinal blood vessel and retinal lesion segmentation tasks.

2.1.1 Retinal Blood Vessel Segmentation

Many computer-aided diagnostic (CAD) systems begin with the segmentation of retinal blood vessels to diagnose non-ocular conditions like hypertension, cerebrovascular, and cardiovascular diseases, in addition to ocular conditions like diabetic retinopathy (DR). Retinal segmentation involves automatically detecting the boundaries of blood vessels within the retina, predicting whether each pixel in a fundus image is a vessel or not through binary classification. This detection is critical for classifiers to learn essential features like retinal proliferation and retinal detachment, which significantly aid in accurately classifying diabetic retinopathy.

The task is challenging due to complex vessel morphology, lesion confusion, and limited annotated data. However, recent retinal blood vessel segmentation advancements have notably improved the accuracy and reliability of diagnosing diabetic retinopathy and other vision-threatening conditions. Various researchers have developed methods to enhance segmentation performance, often employing Fylly Convolutional Neural Networks (FCNNs), or exploiting U-Net architecture and its variations for semantic segmenting blood vessels from fundus retina images.

2.1.1.1 FCN-based Methods for Retinal Blood Vessel Segmentation

Fully Convolutional Neural Networks (FCNs) have proven to be highly effective in applications related to image segmentation. In retinal blood vessel segmentation, FCNs have been integrated into various studies. In [70], authors have introduced dual-source fusing FCN to improve blood vessel segmentation accuracy. The FCN models receive fused input from dual sources, including retinal grayscale image and edge information obtained by the Sobel filter. FCN has also been used in [71] for retinal patch-level segmentation. The authors in this study have proposed a blood vessel segmentation framework that contains three main steps, patch extraction, FCN-based patch segmentation, and then the obtained FCN segmentation masks are combined using a patch aggregation phase. Their framework is tested using the DRIVE dataset, which shows promising results. Figure 2.1 presents an overview of the semantic segmentation approach based on FCN methods proposed in [8].

An improved multi-path FCN model was proposed in [72] for automatic retinal vessel segmentation. The model receives a multi-scale retinal image input where the final segmentation result is obtained by overlapping the multi-segmentation outputs from the multi-path FCN model using a reconstruction algorithm. The experimental study has been conducted on three retinal image datasets including STARE, DRIVE, and CHASE_DB1 showing a performance improvement compared to basic FCN.

In the same context, a butterfly FCN (BFCN) [73] has been proposed to improve the basic FCN model by using multi-scale inputs. The multi-scale inputs can improve the segmentation quality by reaching powerful special information via the different receptive fields size. This advantage helps the proposed BFCN to better understand the local context information of the retinal image.

Another study [74] has proved the efficacity of using multi-scale inputs for retinal blood vessel segmentation by designing residual multi-scale FCN architecture. The experimental results on CHASE DB1, DRIVE, and STARE have shown the segmentation ability of thin, medium, and large vasculature under different optic disk positions, sizes, and large variations of contrast.

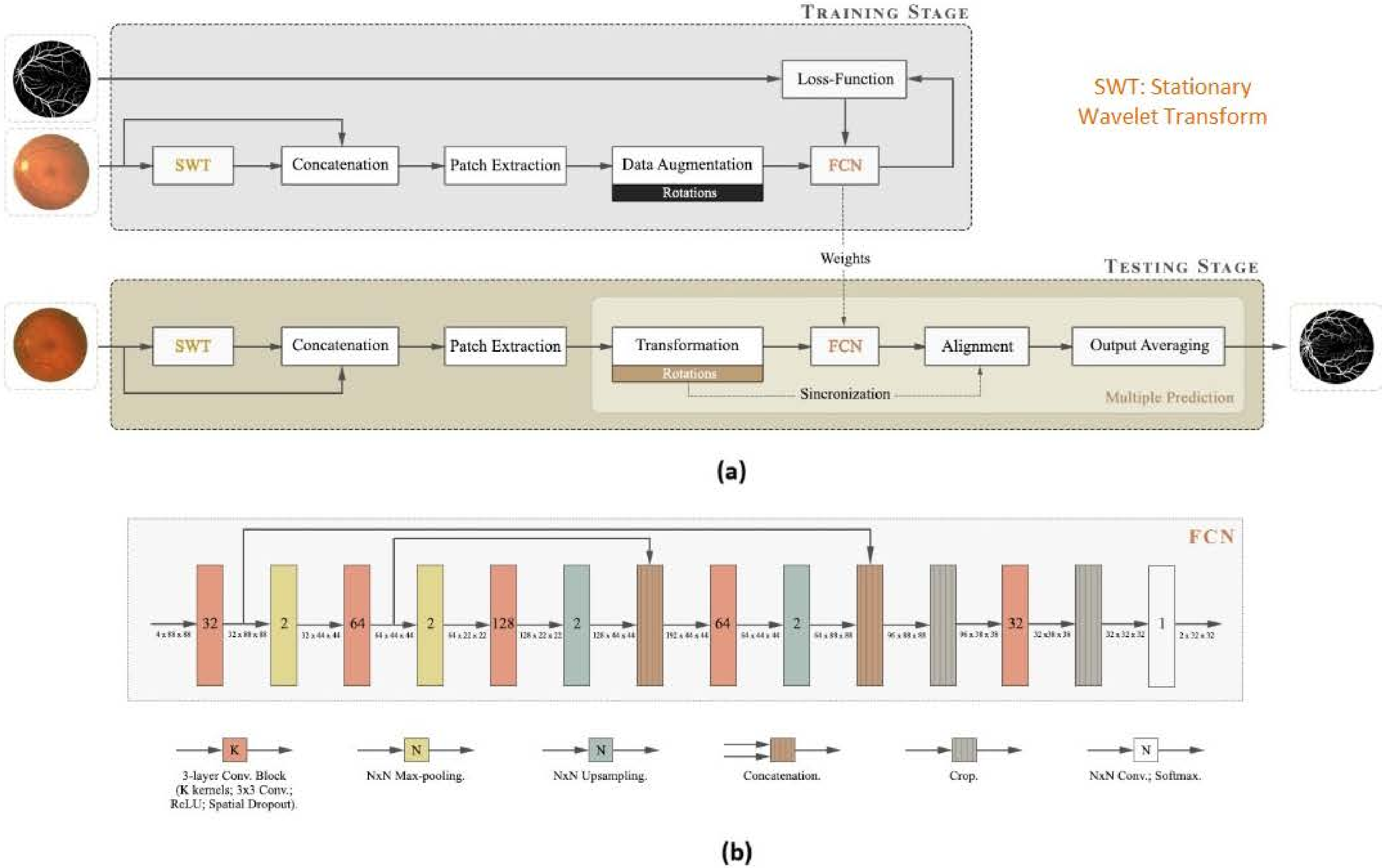

Figure 2.1: Overview of a method for Retinal Blood Vessel Segmentation: (a) Block diagram; (b) FCN architecture [8].

2.1.1.2 U-Net-based Methods for Retinal Blood Vessel Segmentation

The U-Net architecture has been widely used in several medical image analyses including retinal blood vessel segmentation. In [75], U-Net architecture is integrated with region merging. Their approach mitigates the common issue of feature loss during segmentation by recombining segmented regions with the original retinal images, preserving essential details for accurate disease classification. This method achieved impressive accuracy in classifying stages of diabetic retinopathy using Convolutional Neural Networks (CNNs), marking a substantial advancement in diagnostic methodologies. However, due to U-Net's limitations, such as the unsuccessful identification of thin retinal vessels, high computational complexity, insufficient segmentation degree, and poor continuity of microretinal vessels, several methods have been proposed to address these issues.

Authors in [76] have developed Dense U-Net model that advances the capabilities of the U-Net architecture by incorporating dense blocks. This model improves the segmentation accuracy of small blood vessels by ensuring each layer's input derives from the output of all previous layers. Their model demonstrated superior performance on the DRIVE and CHASE_DB1 datasets, showcasing advancements in sensitivity and accuracy for detecting finer vascular details crucial for diagnosing ophthalmic diseases. The authors of [77] have addressed the limited representative capacity of the U-net network by proposing a residual U-Net. The proposed network has a new residual block structure and a dropout layer to alleviate the network's over-fitting problems. The added residual blocks allow the residual U-Net network to extract more representative retinal image features showing promising results for retinal blood vessel segmentation on DRIVE and STARE datasets. In the same context, residual attention and dual-supervision cascaded U-Net (RADCU-Net) model is proposed in [78], using an attention mechanism with a residual unit for image feature representation improvement. The model's performance has been demonstrated on the DRIVE and STARE datasets.

Table 2.1 summarizes the most relevant related works on retinal blood vessel segmentation, highlighting their key contributions, advantages, and limitations providing a comprehensive overview of the state-of-the-art methods for this task.

Tableau 2.1: Synthesis of Retinal Blood Vessel Segmentation Methods

Method	Advantages	Limitations
Dual-Source Fusing FCN [70].	• Improved segmentation accuracy • Uses retinal grayscale image and edge information	• Limited by the quality of edge detection
FCN-Based Patch Segmentation [71].	• Effective patch-level segmentation • Includes patch aggregation phase	• Miss global context information
Improved Multi-Path FCN [72].	• Multi-scale retinal image input • Reconstruction algorithm for final segmentation	• Increased complexity due to multiple paths
Butterfly FCN (BFCN) [73].	• Multi-scale inputs for better local context • Improved understanding of retinal image features	• Potentially higher computational requirements

Continued on next page

Table 2.1 – *Continued from previous page*

Reference & Method	**Advantages**	**Limitations**
Residual Multi-Scale FCN [74].	• Effective for various vessel sizes and optic disk positions • Handles variations in contrast	• Struggle with extremely low-contrast images
U-Net with Region Merging [75].	• Preserves essential details for disease classification • Addresses feature loss issue	• High computational complexity
Dense U-Net [76].	• Improved small blood vessel segmentation • Each layer uses the output from all previous layers	• Increased model complexity and memory usage
Residual U-Net [77].	• New residual block structure • Addresses overfitting with a dropout layer	• Require larger datasets for optimal performance

Continued on next page

Table 2.1 – *Continued from previous page*

Reference & Method	Advantages	Limitations
RADCU-Net [78].	• Uses attention mechanism with residual unit • Dual-supervision for better feature representation	• Complex architecture is challenging to train and optimize

Because retinal vascular segmentation is a challenging and urgent process, several techniques for this task are constantly being developed. Figure 2.2 depicts the development timeline.

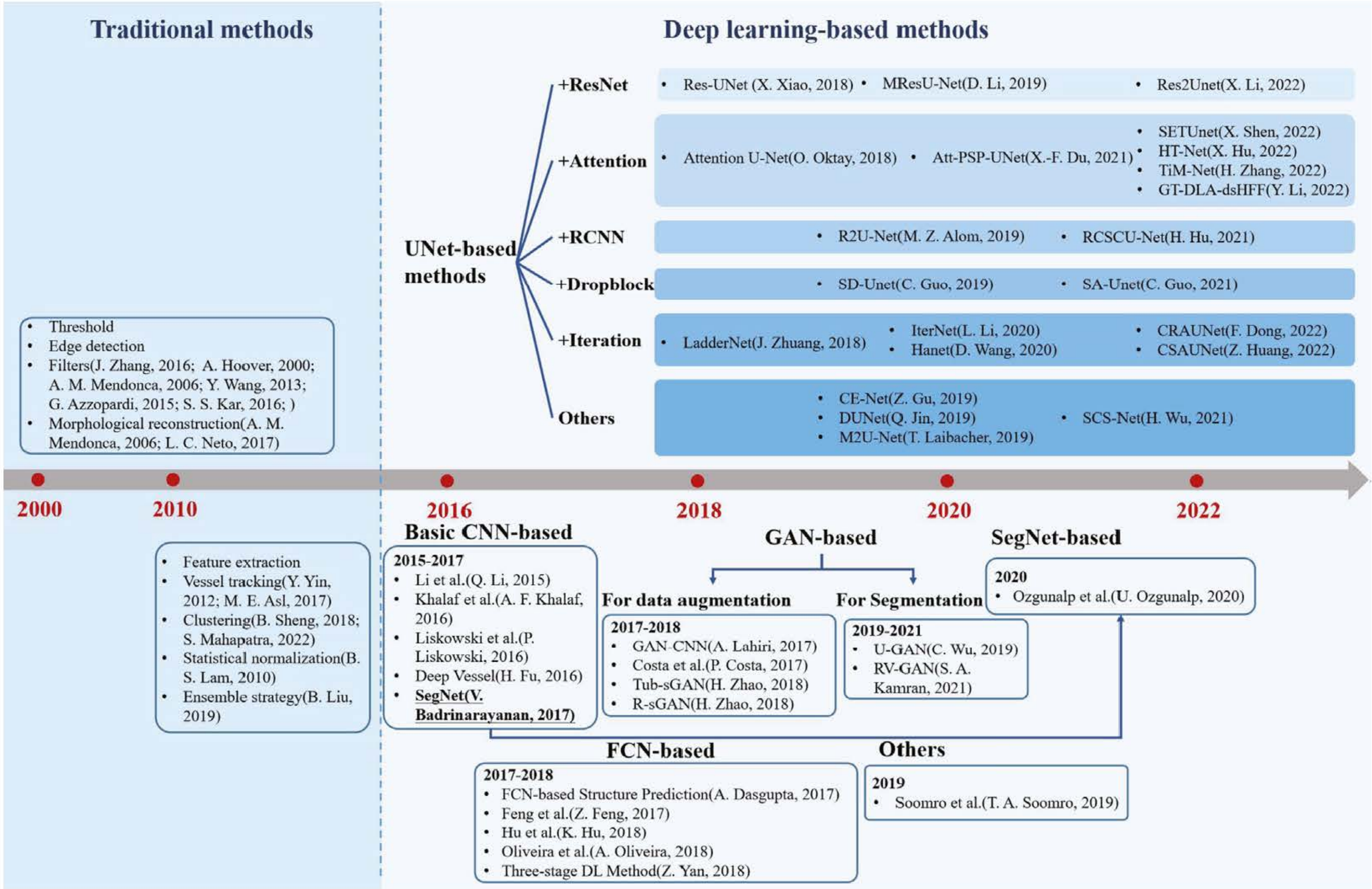

Figure 2.2: History of the automatic segmentation of retinal vessels [9].

2.1.2 Retinal Lesions Segmentation

Retinal lesions are key indicators of eye conditions, especially diabetic retinopathy. These lesions, which include microaneurysms, hemorrhages, hard and soft exudates, mark the stages and severity of the disease. Segmenting these lesions from retinal images is vital for accurate diagnosis and monitoring. By extracting these abnormalities, we can better understand the extent of retinal damage and develop targeted treatments. This section explores the various studies used to segment retinal lesions, highlighting the advantages and the limits of each work.

2.1.2.1 RCNN-based Methods for Retinal Lesions Segmentation

Region-based Convolutional Neural Networks (R-CNN) have emerged as a powerful tool in medical image analysis, offering a robust framework for object detection and segmentation. In the context of retinal lesion segmentation, R-CNNs provide the capability to precisely identify and outline different types of lesions within complex retinal images, addressing challenges such as variability in lesion size, shape, and intensity. Several studies have explored the application of R-CNNs for retinal lesion segmentation.

Faster Region-based Convolutional Neural Network architecture is proposed in [79] to tackle the challenges of DR lesions segmentation. This method was used on two public datasets: DIARETDB1 and Messidor for the localization of four different DR abnormalities including MicroAneurysms (MA), Hemorrhages (HE), Hard Exudates (HE) and Soft Exudates (SE). Authors of [80] addressed the problem of recognizing two types of lesions: Microaneurysms and Exudates, by introducing DRDr model which makes use of R-CNN and Transfer Learning for the segmentation of the lesions found in fundus images. A normalized database was built from e-ophtha MA and e-ophtha EX and customize Mask R-CNN to spot small lesions. To mitigate the limitations of their small dataset, they apply data augmentation strategies and leverage the pre-trained weights of ResNet101.

2.1.2.2 U-Net-based Methods for Retinal Lesions Segmentation

U-Net has become instrumental in segmenting retinal lesions by effectively capturing intricate details from images. Its robust architecture enhances the accuracy of identifying and delineating lesions, contributing significantly to the field of retinal image analysis and clinical diagnosis. A study by [81] has presented a modified U-Net architecture that integrates residual network features, periodic shuffling, and sub-pixel convolution initialized to nearest neighbor resize. This architecture is specifically tailored for the segmentation of microaneurysms and hard exudates for diabetic retinopathy. The study validates the proposed method on two widely used IDRiD and e-ophtha datasets.

U-Net was also integrated into a new architecture [82] featuring a Relation Transformer Block (RTB) that incorporates attention mechanisms at two levels: a self-attention transformer for global lesion feature dependencies and a cross-attention transformer to integrate vascular information, improving lesion detection in complex fundus structures. Additionally, a Global Transformer Block (GTB) captures detailed lesion patterns early in the network. The proposed network segments four types of lesions simultaneously and is validated on IDRiD and DDR datasets.

To tackle the complex structural challenges inherent in lesion segmentation for diabetic retinopathy, authors in [83] introduce Cascade Attentive RefineNet (CARNet). This approach leverages a dual-input strategy, incorporating both global and local information from fundus images. The global image encoder (ResNet50) captures broad context, while the local image encoder (ResNet101) focuses on fine details within patch images. A high-level refinement decoder integrates features from both encoders using dual attention mechanisms, enhancing the model's ability to discern and segment lesions accurately. Evaluation on benchmark datasets (IDRID, E-ophtha, and DDR) demonstrates CARNet's efficacy in overcoming complexities such as diverse lesion sizes and similarity to other fundus tissues.

Feature Fusion U-Net (FFU-Net), an enhanced model tailored for lesion segmentation in diabetic retinopathy is introduced in [84]. The model improves upon traditional U-Net architecture by replacing pooling layers with convolutional layers to better retain spatial information in fundus images. Additionally, FFU-Net incorporates a Multiscale Feature Fusion (MSFF) block to enhance the learning of multiscale features and integrates a Balanced Focal Loss function to mitigate issues related to data imbalance and misclassification. Experimental evaluation on the IDRID dataset validates FFU-Net's efficacy, demonstrating notable advancements in segmentation accuracy compared to conventional U-Net models.

To address the difficulty of segmenting small lesions in fundus images with extensive backgrounds, authors in [85] introduce HACDR-Net. This network includes heterogeneous cross-convolution, modulated deformable convolution, and near-far-aware convolution. An adaptive aggregation module combines heterogeneous feature maps to identify diverse lesion areas. Additionally, the Noise Adjusted Loss (NALoss) function is proposed to balance predictions between background and lesions by incorporating Gaussian noise and hard example mining. The effectiveness of these methods is validated through experiments on the IDRiD and DDR datasets.

Recent studies struggle with accurate model design due to insufficient annotated training data. To address this, authors in [86] propose a semi-supervised multitask learning approach using unlabelled data, like Kaggle-EyePACS, to enhance diabetic retinopathy (DR) segmentation. The model, featuring a multi-decoder architecture, combines unsupervised and supervised learning phases to utilize unlabelled data and improve segmentation performance. This method is rigorously evaluated on FGADR and IDRiD datasets.

Table 2.2 presents a comparative overview of different methods for retinal lesion segmentation. It summarizes the main findings from several investigations, stressing both their benefits and drawbacks. The objective of this comparison analysis is to provide light on the advantages and disadvantages of each work for particular retinal lesion segmentation applications.

Tableau 2.2: Synthesis of Retinal Lesion Segmentation Methods

Reference & Method	**Advantages**	**Limitations**
Faster R-CNN for DR Lesions [79].	• Localizes four different DR abnormalities • Tested on two public datasets	• May struggle with very small lesions
DRDr Model with Mask R-CNN [80].	• Recognizes Microaneurysms and Exudates • Uses transfer learning and data augmentation	• Limited by small dataset size
Modified U-Net [81].	• Integrates residual network features • Uses periodic shuffling and sub-pixel convolution	• May be computationally intensive

Continued on next page

Table 2.2 – *Continued from previous page*

Reference & Method	**Advantages**	**Limitations**
U-Net with Relation Transformer Block [82].	• Incorporates multi-level attention mechanisms • Segments four types of lesions simultaneously	• Complex architecture may require significant training data
Cascade Attentive RefineNet (CARNet) [83].	• Dual-input strategy for global and local information • Uses dual attention mechanisms	• May require high computational resources
FFU-Net (Feature Fusion U-Net) [84].	• Retains spatial information better • Incorporates Multi-scale Feature Fusion block	• Might struggle with extremely small lesions
HACDR-Net [85].	• Uses heterogeneous cross-convolution • Incorporates Noise Adjusted Loss function	• Complex architecture may be difficult to optimize

Continued on next page

Table 2.2 – *Continued from previous page*

Reference & Method	Advantages	Limitations
Semi-Supervised Multi-task Learning [86].	• Utilizes unlabelled data • Combines unsupervised and supervised learning	• Performance may depend on quality of unlabelled data

2.2 Diabetic Retinopathy Classification

2.2.1 End-to-End Learning

Recent years have seen the introduction of a few unique CNN-based techniques for DR classification, such as IDRiD, EyePacs, APTOS, DDR, and Messidor, which use end-to-end training on small datasets. To enhance the amount of data needed to train their suggested architectures, these architectures frequently use augmentation. Works provided in [87], [88], [89], [90], [91], [92], [93], [94], [95], [96], [97], and [98] have examples of similar buildings. Despite being trained from scratch, these architectures exhibit strong learning logic and achieve excellent results in the detection and grading of DR. To enhance the architecture or create a new architecture centred around CNN encoders, each of these architectures incorporates a novel feature into the established deep CNN techniques.

Traditional CNN-based techniques have been improved, as demonstrated in the works [88], [89], [90], [91], [93], [94], and [96]. Although the work in [91] builds its suggested method around the Inception architecture, the authors included long-range global dependency units to link feature maps from several convolutional layers. This prevents overfitting and greatly enhances the model's learning on tiny datasets. In a comparable manner, the multiclass classification task in [93] is approached as two binary and one ternary classification tasks. By splitting the detection tasks that identify images as DR and No-DR, NPDR and PDR, and then lastly as mild, moderate, and severe NPDR, three networks have been employed to achieve DR grading. Because each network had a unique architecture and preprocessing method, the model was able to address the class imbalance present in the datasets. The work in [90] presents the chimp optimization approach for training a high-layer DenseNet variation. The paper also introduced a new CNN module for multiclass classification tasks in DR grading, called Spiking neural networks. In studies [99], [88], [94], and [96], other techniques involving residual block integration, entropy improvement, and discrete wavelet transform have been combined with conventional CNNs to allow them to be trained from scratch instead of utilizing pre-trained models.

The following publications described a few unconventional CNN-based DL architectures: [87], [92], [95], [97], and [98]. These architectures deviate from the conventional pooling layer stacking method. Instead, those efforts improved the activation and loss functions and included more sophisticated and novel methods of extracting feature maps, attention layers, and cascading skip connections. In [87], skip connections are integrated into dilated convolutional blocks, and feature maps are produced from these convolutional layers by hierarchical feature extraction. A unique approach that enhances current capsule networks and adapts them to DR detection and grading tasks was presented in the work of [92]. Newer and shorter variants of the ResNet are provided in the works [97] and [98], which are appropriate for learning from datasets with little to no data. When compared to typical ResNets with 50 layers or more, these architectures exhibit reduced fluctuations on the learning curve and superior generalization capabilities.

Table 2.3 provides a comprehensive comparison of various end-to-end learning methods for diabetic retinopathy classification. It encapsulates key findings from multiple studies, highlighting both the strengths and limitations of each approach. This comparative analysis aims to illuminate the merits and drawbacks of different techniques, facilitating informed decision-making when selecting appropriate methods for specific diabetic retinopathy classification tasks. By presenting this information in a structured format, researchers and practitioners can readily assess the suitability of each method for their particular applications.

Tableau 2.3: Summary of End-to-End Learning Techniques for DR Classification

Method	**Advantages**	**Limitations**
[87]	• Integrates skip connections into dilated convolutional blocks • Hierarchical feature extraction	• May be complex to implement • Requires significant computational resources

Continued on next page

Table 2.3 – *Continued from previous page*

Method	Advantages	Limitations
[88]	• Uses augmentation to enhance training • Incorporates residual block integration	• Performance highly dependent on augmentation techniques • Potential overfitting if not carefully managed
[89]	• Uses augmentation to enhance training	• Performance highly dependent on augmentation techniques • Potential overfitting if not carefully managed
[90]	• Introduces Spiking neural networks for multiclass classification • Uses DenseNet variation with chimp optimization	• Complex architecture may be difficult to optimize • Requires high computational resources
[91]	• Incorporates long-range global dependency units • Prevents overfitting on small datasets	• May require extensive hyperparameter tuning • Computationally intensive

Continued on next page

Table 2.3 – *Continued from previous page*

Method	Advantages	Limitations
[92]	• Enhances capsule networks for DR tasks • Novel approach to feature extraction and attention layers	• May require specialized knowledge to implement • Potential scalability issues
[93]	• Addresses class imbalance with unique preprocessing methods • Splits detection tasks for more accurate classification	• Complex system with three networks may be hard to manage • High computational requirements
[94]	• Uses augmentation to enhance training • Integrates residual blocks for improved performance	• Performance highly dependent on augmentation techniques • Potential overfitting if not carefully managed
[95]	• Improves activation and loss functions • Uses sophisticated feature map extraction techniques	• Complex architecture may be difficult to implement • Requires high computational resources

Continued on next page

Table 2.3 – *Continued from previous page*

Method	Advantages	Limitations
[96]	• Uses augmentation to enhance training • Incorporates entropy improvement and discrete wavelet transform	• Performance highly dependent on augmentation techniques • Potential overfitting if not carefully managed
[97]	• Suitable for small datasets • Reduces learning curve fluctuations	• May not perform as well on larger datasets • Limited by the depth of the network
[98]	• Superior generalization capabilities • Suitable for small datasets	• May not perform as well on larger datasets • Limited by the depth of the network

2.2.2 Retinal Lesions based Methods

This section summarizes the most common techniques for DR classification based on the detection of single and multiple ocular lesions. The most popular method consists of two steps: first, a set of features relating to DR-related lesions are extracted, and subsequently, a model utilizes these features to classify or grade the DR.

The proposed framework in [100] for ocular lesion detection and DR classification includes three stages: preprocessing fundus images, lesion detection using transfer learning with pre-trained models (DenseNet121, Xception, ResNet50, and MobileNet). Then, classifier training on the Kaggle EyePACS dataset. The best classifier's lesion predictions are then used for DR classification, and evaluated on the Messidor-2 dataset.

Another work presented in [101] introduces a CNN-based algorithm for diabetic retinopathy (DR) classification, focusing initially on preprocessing fundus images and segmenting branching blood vessels using maximum principal curvature from the Hessian matrix. Although effective, relying solely on blood vessel characteristics for classification poses limitations. While enhancing images with adaptive histogram equalization and morphological opening, classification is carried out by two CNN-based sub-networks: an excitation sub-network and a bottleneck sub-network. Their agreement determines the final label. Experimental validation was performed on DIARETDB1 dataset. However, the method's reliance on blood vessel features alone may overlook other crucial indicators of DR progression and severity, suggesting a need for broader feature inclusion to enhance diagnostic accuracy.

In [102], a novel directed acyclic graph (DAG) network based on multi-feature fusion of fundus pictures is proposed for the multi-classification of diabetic retinopathy. First, three characteristics—microaneurysms, soft exudate, and neovascularization spots—are extracted using several algorithms from diabetic retinas of varying grades. These attributes are then incorporated into the innovative DAG network that has been presented. Additionally, distinct grades of diabetic retinas are learned by the combination of the multi-feature fusion method. Lastly, the clinical multi-classification of diabetic retina is carried out using the optimal classification model. The IDRiD and clinical datasets from Dalian NO.3 People's Hospital are utilized to assess how well the suggested approach performs.

Apart from these techniques, Authors in [103] have presented an innovative methodology that makes significant progress in the field. In order to detect retinal lesions, their system uses a hybrid classifier that goes through multiple stages: preprocessing to remove background pixels and extract the optic disc and blood vessels; filter banks are used to detect potential lesions; and a feature set is created using descriptors like shape, intensity, and statistics. In order to improve classification accuracy, this method expands on the m-Mediods based modeling technique by integrating it with an ensemble Gaussian Mixture Model. DIARETDB1, STARE, MESSIDOR, and DRIVE—standard fundus image databases were used to illustrate the efficacy of this approach.

Table 2.4 offers a detailed overview of different retinal lesion-based methods for diabetic retinopathy classification. It synthesizes results from various research efforts, emphasizing both the advantages and potential shortcomings of each methodology. The purpose of this comparative summary is to shed light on the strengths and weaknesses of each study. This tabular presentation allows for easy comparison and evaluation of different lesion-based techniques.

Tableau 2.4: Summary of Retinal Lesion-Based Methods for DR Classification

Model	Advantages	Limitations	Datasets Used
Transfer Learning Model [100]	• Utilizes pre-trained models • Multi-stage approach for improved accuracy	• May require significant computational resources • Dependence on quality of pre-trained models	• Kaggle • EyePACS • Messidor-2
CNN-based Blood Vessel Model [101]	• Focuses on blood vessel characteristics • Uses adaptive histogram equalization	• Overlooks other DR indicators • May miss non-vascular lesions	• DIARETDB1
Deep Feature Extraction Model [104]	• Combines deep learning with traditional classifiers • Uses data augmentation	• Complexity in model selection • Potential overfitting with augmented data	• IDRID Dataset

Continued on next page

Table 2.4 – *Continued from previous page*

Model	Advantages	Limitations	Datasets Used
DAG Network Model [102]	• Multi-feature fusion approach • Considers various lesion types	• Complex network structure • May require extensive training data	• IDRiD • Dalian NO.3 People's Hospital dataset
Hybrid Classifier Model [103]	• Comprehensive lesion detection • Combines multiple classification techniques	• High computational complexity • May be sensitive to image quality	• DIARETDB • STARE • MESSIDOR • DRIVE

Conclusion

This chapter has examined current methods in retinal image analysis for diabetic retinopathy (DR) classification. We focused on three key areas where recent progress has been significant: retinal blood vessel segmentation, retinal lesion segmentation, and diabetic retinopathy classification.

For retinal blood vessel segmentation, we reviewed several network architectures, including Fully Convolutional Networks (FCNs) and U-Net variants. These approaches have improved the accuracy of retinal vasculature delineation, crucial for early DR detection. In retinal lesion segmentation, we discussed advancements using Region-based Convolutional Neural Networks (R-CNN) and modified U-Net designs. These techniques have enhanced our ability to identify and outline various DR-related lesions, such as microaneurysms, hemorrhages, and exudates.

Regarding diabetic retinopathy classification, we explored two main strategies: end-to-end learning and lesion-based methods. End-to-end approaches classify DR directly from fundus images, while lesion-based methods use segmented lesions as classification features. Both approaches represent important developments in automated DR diagnosis. We have analyzed the strengths and weaknesses of each method, providing a balanced overview of current research. This review highlights challenges and opportunities in DR detection and classification using computational techniques.

The next chapter will introduce our proposed framework for retinal lesions-based DR classification. This approach builds on the strengths of the methods discussed here while addressing some of their limitations. Integrating an advanced segmentation model with a recent classification network architecture to improve DR diagnosis accuracy. The following chapter will outline the proposed framwork with the experimental study showing its contribution to automated DR classification.

Chapter 3

Contributions and Results

Introduction

This chapter outlines the suggested framework for analysing retinal images, with an emphasis on segmentation and the identification o f D iabetic R etinopathy (DR). The chapter is organised as follows: we start by outlining the suggested framework and going over the segmentation and detection phases of the retinal images. The experimental setup, including the datasets and evaluation metrics, is then described. We wrap up with a summary of the experimental outcomes and give the blood vessel image segmentation, retinal lesions image segmentation, and DR detection.

3.1 Proposed Framework

For more information please contact us on

- amel.bouchemha@univ-tebessa.dz
- yaakoub.boualleg@univ-tebessa.dz

General Conclusion

In this project, we proposed and realized a realtime diabetic retinopathy detection using advanced deep-learning techniques. Experimental results prove that the proposed approach has proven its effectiveness and robustness, both absolutely and in comparison with other similar state-of-the-art methods, using several well-known and publicly available datasets.

The proposed method includes many elements that contribute fully to its success. First, the proposed approach combines an innovative image segmentation method against state-of-the-art classification models to achieve high accuracy in DR diagnosis. The U-Mamba architecture, leveraging the power of structured state-space models, demonstrated robust performance in segmenting various retinal lesions and blood vessels. Then, the Swin Transformer-based classification model showed encouraging performance across several datasets, building on earlier segmentation results. The high sensitivity and specificity values across all datasets further underscore the model's reliability in correctly identifying both positive and negative cases. Our method proved to be robust to the training set and to the inter-rater variability, which shows its potential for real-world application in screening and diagnostic systems.

The proposed approach gave high performance for the different levels of retinopathy severity. All metric validations demonstrated the effectiveness of the developed system for early and real-time detection of DR. Nevertheless, while these results are promising, there is still room for further improvement and exploration. Future work could focus on improving model performance for less common severity levels, investigating the interpretability of model decisions, and validating the approach on other diversified datasets. In conclusion, this chapter demonstrates the potential of combining advanced segmentation and classification techniques for accurate DR detection. The proposed framework holds promise for aiding the early diagnosis and management of diabetic retinopathy, which could contribute to improved patient outcomes in the clinical setting.

Bibliography

[1] Yuyi Han, Xiaoqian Zhang, Zhiyin Zhou, Mengxi Yu, Qiuhong Wang, Jin Yao, and Yi Han. Hydrogen sulfide serves as a biomarker in the anterior segment of patients with diabetic retinopathy. *Int. Ophthalmol.*, 40(4):891–899, April 2020.

[2] Jinming Zou, Yi Han, and Sung-Sau So. *Overview of Artificial Neural Networks*, pages 14–22. Humana Press, Totowa, NJ, 2009.

[3] Rudolf Kruse, Sanaz Mostaghim, Christian Borgelt, Christian Braune, and Matthias Steinbrecher. *Multi-layer Perceptrons*, pages 53–124. Springer International Publishing, Cham, 2022.

[4] Kit Yan Chan, Bilal Abu-Salih, Raneem Qaddoura, Ala' M. Al-Zoubi, Vasile Palade, Duc-Son Pham, Javier Del Ser, and Khan Muhammad. Deep neural networks in the cloud: Review, applications, challenges and research directions. *Neurocomputing*, 545:126327, 2023.

[5] Yu Li Anas Bilal Malik Abdul Manan Abdul Qadir Khan, Guangmin Sun. Optimizing fully convolutional encoder-decoder network for segmentation of diabetic eye disease. *Computers, Materials & Continua*, 77(2):2481–2504, 2023.

[6] Soumyajit Podder, Somnath Bhattacharjee, and Arijit Roy. An efficient method of detection of covid-19 using mask r-cnn on chest x-ray images. *AIMS Biophysics*, 8(3):281–290, 2021.

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[8] Américo Oliveira, Sérgio Pereira, and Carlos A. Silva. Retinal vessel segmentation based on fully convolutional neural networks. *Expert Systems with Applications*, 112:229–242, 2018.

[9] Qing Qin and Yuanyuan Chen. A review of retinal vessel segmentation for fundus image analysis. *Engineering Applications of Artificial Intelligence*, 128:107454, 2024.

[10] Martin M Nentwich and Michael W Ulbig. Diabetic retinopathy - ocular complications of diabetes mellitus. *World Journal of Diabetes*, 6(3):99–489, 2015.

[11] Gurmessa Nugussu Gelcho and Firomsa Shewa Gari. Time to diabetic retinopathy and its risk factors among diabetes mellitus patients in jimma university medical center, jimma, southwest ethiopia. *Ethiopian Journal of Health Sciences*, 32(5):937–946, 2022.

[12] J.E. Shaw, R.A. Sicree, and P.Z. Zimmet. Global estimates of the prevalence of diabetes for 2010 and 2030. *Diabetes Research and Clinical Practice*, 87(1):4–14, January 2010.

[13] Sameer P Leley, Thomas A Ciulla, and Ashay Bhatwadekar. Diabetic retinopathy in the aging population: A perspective of pathogenesis and treatment. *Clinical Interventions in Aging*, Volume 16:1367–1378, July 2021.

[14] Borja Corcóstegui, Santiago Durán, María Olga González-Albarrán, Cristina Hernández, José María Ruiz-Moreno, Javier Salvador, Patricia Udaondo, and Rafael Simó. Update on diagnosis and treatment of diabetic retinopathy: A consensus guideline of the working group of ocular health (spanish society of diabetes and spanish vitreous and retina society). *Journal of Ophthalmology*, 2017(1):8234186, 2017.

[15] Silke Aumann, Sabine Donner, Jörg Fischer, and Frank Müller. *Optical Coherence Tomography (OCT): Principle and Technical Realization*, pages 59–85. Springer International Publishing, Cham, 2019.

[16] Biswajit Jena, Sanjay Saxena, Gopal K. Nayak, Luca Saba, Neeraj Sharma, and Jasjit S. Suri. Artificial intelligence-based hybrid deep learning models for image classification: The first narrative review. *Computers in Biology and Medicine*, 137:104803, 2021.

[17] R.I.G. Holt and A. Flyvbjerg. *Textbook of Diabetes*. Wiley, 2024.

[18] Mark A Atkinson, George S Eisenbarth, and Aaron W Michels. Type 1 diabetes. *The Lancet*, 383(9911):69–82, January 2014.

[19] Sudesna Chatterjee, Kamlesh Khunti, and Melanie J Davies. Type 2 diabetes. *The Lancet*, 389(10085):2239–2251, June 2017.

[20] Michael J. Fowler. Microvascular and Macrovascular Complications of Diabetes. *Clinical Diabetes*, 26(2):77–82, 04 2008.

[21] Anjali D Deshpande, Marcie Harris-Hayes, and Mario Schootman. Epidemiology of diabetes and diabetes-related complications. *Physical Therapy*, 88(11):1254–1264, 11 2008.

[22] American Diabetes Association. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2020. *Diabetes Care*, 43(Supplement$_1$) : $S14 - -S31, 122019$.

[23] Thomas A. Buchanan and Anny H. Xiang. Gestational diabetes mellitus. *The Journal of Clinical Investigation*, 115(3):485–491, 3 2005.

[24] Nicolai Lohse Peter Damm Moshe Hod Aliya Jiwani, Elliot Marseille and James G. Kahn. Gestational diabetes mellitus: results from a survey of country prevalence and practices. *The Journal of Maternal-Fetal & Neonatal Medicine*, 25(6):600–610, 2012.

[25] N Unwin, J Shaw, P Zimmet, and K G M M Alberti. Impaired glucose tolerance and impaired fasting glycaemia: the current status on definition and intervention. *Diabetic medicine : a journal of the British Diabetic Association*, 19(9):708—723, September 2002.

[26] Hong Sun, Pouya Saeedi, Suvi Karuranga, Moritz Pinkepank, Katherine Ogurtsova, Bruce B. Duncan, Caroline Stein, Abdul Basit, Juliana C.N. Chan, Jean Claude Mbanya, Meda E. Pavkov, Ambady Ramachandaran, Sarah H. Wild, Steven James, William H. Herman, Ping Zhang, Christian Bommer, Shihchen Kuo, Edward J. Boyko, and Dianna J. Magliano. Idf diabetes atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Research and Clinical Practice*, 183:109119, 2022.

[27] Christopher C. Patterson, Valma Harjutsalo, Joachim Rosenbauer, Andreas Neu, Ondrej Cinek, Torild Skrivarhaug, Birgit Rami-Merhar, Gyula Soltesz, Jannet Svensson, Roger C. Parslow, Conxa Castell, Eugen J. Schoenle, Polly J. Bingley, Gisela Dahlquist, Przemysława K. Jarosz-Chobot, Dalė Marčiulionytė, Edna F. Roche, Ulrike Rothe, Natasa Bratina, Constantin Ionescu-Tirgoviste, Ilse Weets, Mirjana Kocova, Valentino Cherubini, Natasa Rojnic Putarek, Carine E. deBeaufort, Mira Samardzic, and Anders Green. Trends and cyclical variation in the incidence of childhood type 1 diabetes in 26 european centres in the 25 year period 1989–2013: a multicentre prospective registration study. *Diabetologia*, 62(3):408–417, 2019.

[28] Bin Zhou, Yuan Lu, Kaveh Hajifathalian, James Bentham, Mariachiara Di Cesare, Goodarz Danaei, Honor Bixby, Melanie J Cowan, Mohammed K Ali, Cristina Taddei, Wei-Cheng Lo, Bruno Reis-Santos, Gretchen A Stevens, Leanne M Riley, J Jaime Miranda, Peter Bjerregaard, Juan A Rivera, Heba M Fouad, Guansheng Ma, Jean-Claude

Mbanya, Stephen T McGarvey, Viswanathan Mohan, Altan Onat, Aida Pilav, Ambady Ramachandran, Habiba Ben Romdhane, Christopher J Paciorek, James E Bennett, and Majid Ezzati. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *The Lancet*, 387(10027):1513–1530, 2016.

[29] Kamelia Zahra Yala, Bouchra Naima Yahia Cherif, and Abdellatif Yousfi. Estimation des coûts des complications macrovasculaires du diabète. *JOURNAL ALGÉRIEN DE PHARMACIE*, 5(3):1–8, janv. 2024.

[30] Michal Shani, Tali Eviatar, Doron Komaneshter, and Shlomo Vinker. Diabetic retinopathy –incidence and risk factors in a community setting- a longitudinal study. *Scand. J. Prim. Health Care*, 36(3):237–241, July 2018.

[31] Helga Kolb. How the retina works: Much of the construction of an image takes place in the retina itself through the use of specialized neural circuits. *American Scientist*, 91(1):28–35, 2003.

[32] Kishore Cholkar, Supriya Reddy Dasari, Dhananjay Pal, and Ashim K. Mitra. 1 - eye: anatomy, physiology and barriers to drug delivery. In Ashim K. Mitra, editor, *Ocular Transporters and Receptors*, Woodhead Publishing Series in Biomedicine, pages 1–36. Woodhead Publishing, 2013.

[33] Michael Brownlee, Lloyd P. Aiello, Mark E. Cooper, Aaron I. Vinik, Richard W. Nesto, and Andrew J.M. Boulton. *Complications of Diabetes Mellitus*, pages 1462–1551. Elsevier, Netherlands, 12th edition, January 2011.

[34] Lee Ann Remington and Denise Goodwin. *Clinical Anatomy and Physiology of the Visual System E-Book*. Elsevier Health Sciences, 2021.

[35] Charumathi Sabanayagam, Riswana Banu, Miao Li Chee, Ryan Lee, Ya Xing Wang, Gavin Tan, Jost B Jonas, Ecosse L Lamoureux, Ching-Yu Cheng, Barbara EK Klein, et al. Incidence and progression of diabetic retinopathy: a systematic review. *The lancet Diabetes & endocrinology*, 7(2):140–149, 2019.

[36] Amir Mahdjoubi, Youcef Bousnina, Gaelle Barrande, Faïza Bensmaine, Sadri Chahed, and Amina Ghezzaz. Features of cotton wool spots in diabetic retinopathy: a spectral-domain optical coherence tomography angiography study. *International Ophthalmology*, 40(7):1625–1640, March 2020.

[37] Shweta Chaudhary, Jill Zaveri, and Norbert Becker. Proliferative diabetic retinopathy (pdr). *Disease-a-Month*, 67(5):101140, 2021.

[38] Nishtha Panwar, Philemon Huang, Jiaying Lee, Pearse A. Keane, Tjin Swee Chuan, Ashutosh Richhariya, Stephen Teoh, Tock Han Lim, and Rupesh Agrawal. Fundus photography in the 21st century—a review of recent technological advances and their implications for worldwide healthcare. *Telemedicine and e-Health*, 22(3):198–208, 2016.

[39] Stephen H. Tsang and Tarun Sharma. *Fluorescein Angiography*, pages 7–10. Springer International Publishing, Cham, 2018.

[40] Silke Aumann, Sabine Donner, Jörg Fischer, and Frank Müller. *Optical Coherence Tomography (OCT): Principle and Technical Realization*, page 59–85. Springer International Publishing, 2019.

[41] KV Chalam and Kumar Sambhav. Optical coherence tomography angiography in retinal diseases. *Journal of Ophthalmic and Vision Research*, 11(1):84, 2016.

[42] Ibrahim E. Mohamed, Mona A. Mohamed, Mohamed Yousef, Mustafa Z. Mahmoud, and Batil Alonazi. Use of ophthalmic b-scan ultrasonography in determining the causes of low vision in patients with diabetic retinopathy. *European Journal of Radiology Open*, 5:79–86, 2018.

[43] Steffen Schmitz-Valckenberg, Maximilian Pfau, Monika Fleckenstein, Giovanni Staurenghi, Janet R. Sparrow, Almut Bindewald-Wittich, Richard F. Spaide, Sebastian Wolf, Srinivas R. Sadda, and Frank G. Holz. Fundus autofluorescence imaging. *Progress in Retinal and Eye Research*, 81:100893, 2021.

[44] Shivani V. Reddy and Deeba Husain. Panretinal photocoagulation: A review of complications. *Seminars in Ophthalmology*, 33(1):83–88, 2018.

[45] Miho Nozaki, Ryota Ando, Toshiya Kimura, Fusae Kato, and Tsutomu Yasukawa. The role of laser photocoagulation in treating diabetic macular edema in the era of intravitreal drug administration: A descriptive review. *Medicina*, 59(7), 2023.

[46] Rishi P. Singh, Michael J. Elman, Simran K. Singh, Anne E. Fung, and Ivaylo Stoilov. Advances in the treatment of diabetic retinopathy. *Journal of Diabetes and its Complications*, 33(12):107417, 2019.

[47] Saagar Pandit, Yasha Modi, and Nitish Mehta. *Advances in Vitrectomy*, pages 313–324. Springer Nature Singapore, Singapore, 2023.

[48] Igor Kozak and Jeffrey K. Luttrull. Modern retinal laser therapy. *Saudi Journal of Ophthalmology*, 29(2):137–146, 2015.

[49] Donald Olding Hebb. *The organization of behavior*, volume 65. Wiley New York, 1949.

[50] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958.

[51] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[52] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255. IEEE, 2009.

[53] John Nickolls, Ian Buck, Michael Garland, and Kevin Skadron. Scalable parallel programming with CUDA. *Queue*, 6(2):40–53, 2008.

[54] Erik Lindholm, John Nickolls, Stuart Oberman, and John Montrym. Nvidia tesla: A unified graphics and computing architecture. *IEEE micro*, 28(2):39–55, 2008.

[55] Rami Al-Ruzouq, Mohamed Barakat A. Gibril, Abdallah Shanableh, Abubakir Kais, Osman Hamed, Saeed Al-Mansoori, and Mohamad Ali Khalil. Sensors, features, and machine learning for oil spill detection and monitoring: A review. *Remote Sensing*, 12(20), 2020.

[56] Abraham Yosipof, Rita C. Guedes, and Alfonso T. García-Sosa. Data mining and machine learning models for predicting drug likeness and their disease or organ category. *Frontiers in Chemistry*, 6, May 2018.

[57] Pam Mantri and John Thomas. Design of neural networks. *IOP Conference Series: Materials Science and Engineering*, 1174(1):012023, aug 2021.

[58] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2(4):303–314, December 1989.

[59] J. Snyman. *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms*. Applied Optimization. Springer, 2005.

[60] HENRY J. KELLEY. Gradient theory of optimal flight paths. *ARS Journal*, 30(10):947–954, October 1960.

[61] Pérez-Enciso and Zingaretti. A guide for using deep learning for complex trait genomic prediction. *Genes*, 10(7):553, July 2019.

[62] Mariusz Bojarski, Philip Yeres, Anna Choromanska, Krzysztof Choromanski, Bernhard Firner, Lawrence Jackel, and Urs Muller. Explaining how a deep neural network trained with end-to-end learning steers a car. 2017.

[63] Asifullah Khan, Anabia Sohail, Umme Zahoora, and Aqsa Saeed Qureshi. A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, 53(8):5455–5516, 2020.

[64] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. 2015.

[65] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778. IEEE, 2016.

[66] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9. IEEE, 2015.

[67] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. October 2020.

[68] Meng-Hao Guo, Tian-Xing Xu, Jiang-Jiang Liu, Zheng-Ning Liu, Peng-Tao Jiang, Tai-Jiang Mu, Song-Hai Zhang, Ralph R Martin, Ming-Ming Cheng, and Shi-Min Hu. Attention mechanisms in computer vision: A survey. *Comput. Vis. Media (Beijing)*, 8(3):331–368, September 2022.

[69] Zhaoyang Niu, Guoqiang Zhong, and Hui Yu. A review on the attention mechanism of deep learning. *Neurocomputing*, 452:48–62, 2021.

[70] Jianqiang Li, Qidong Hu, Azhar Imran, Li Zhang, Ji-jiang Yang, and Qing Wang. Vessel recognition of retinal fundus images based on fully convolutional network. In *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, volume 02, pages 413–418, 2018.

[71] Taibou Birgui Sekou, Moncef Hidane, Julien Olivier, and Hubert Cardot. Retinal blood vessel segmentation using a fully convolutional network – transfer learning from patch- to image-level. In Yinghuan Shi, Heung-Il Suk, and Mingxia Liu, editors, *Machine Learning in Medical Imaging*, pages 170–178, Cham, 2018. Springer International Publishing.

[72] Yun Jiang, Hai Zhang, Ning Tan, and Li Chen. Automatic retinal blood vessel segmentation based on fully convolutional neural networks. *Symmetry*, 11(9), 2019.

[73] Yun Jiang, Falin Wang, Jing Gao, and Wenhuan Liu. Efficient bfcn for automatic retinal vessel segmentation. *Journal of Ophthalmology*, 2020(1):6439407, 2020.

[74] Tariq M. Khan, Antonio Robles-Kelly, Syed S. Naqvi, and Muhammad Arsalan. Residual multiscale full convolutional network (rm-fcn) for high resolution semantic segmentation of retinal vasculature. In Andrea Torsello, Luca Rossi, Marcello Pelillo, Battista Biggio, and Antonio Robles-Kelly, editors, *Structural, Syntactic, and Statistical Pattern Recognition*, pages 324–333, Cham, 2021. Springer International Publishing.

[75] Sairaj Burewar, Anil Balaji Gonde, and Santosh Kumar Vipparthi. Diabetic retinopathy detection by retinal segmentation with region merging using cnn. In *2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS)*, pages 136–142, 2018.

[76] Yinlin Cheng, Mengnan Ma, Liangjun Zhang, ChenJin Jin, Li Ma, and Yi Zhou. Retinal blood vessel segmentation based on densely connected u-net. *Mathematical Biosciences and Engineering*, 17(4):3088–3108, 2020.

[77] Di Li, Dhimas Arief Dharmawan, Boon Poh Ng, and Susanto Rahardja. Residual u-net for retinal vessel segmentation. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1425–1429, 2019.

[78] Yong Yang, Weiguo Wan, Shuying Huang, Xin Zhong, and Xiangkai Kong. Radcu-net: residual attention and dual-supervision cascaded u-net for retinal blood vessel segmentation. *International Journal of Machine Learning and Cybernetics*, 14(5):1605–1620, November 2022.

[79] Tahira Nazir, Aun Irtaza, Junaid Rashid, Marriam Nawaz, and Toqeer Mehmood. Diabetic retinopathy lesions detection using faster-rcnn from retinal images. In *2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH)*, pages 38–42, 2020.

[80] Farzan Shenavarmasouleh and Hamid R. Arabnia. Drdr: Automatic masking of exudates and microaneurysms caused by diabetic retinopathy using mask r-cnn and transfer learning. In Hamid R. Arabnia, Leonidas Deligiannidis, Hayaru Shouno, Fernando G. Tinetti, and Quoc-Nam Tran, editors, *Advances in Computer Vision and Computational Biology*, pages 307–318, Cham, 2021. Springer International Publishing.

[81] Nitigya Sambyal, Poonam Saini, Rupali Syal, and Varun Gupta. Modified u-net architecture for semantic segmentation of diabetic retinopathy images. *Biocybernetics and Biomedical Engineering*, 40(3):1094–1109, 2020.

[82] Shiqi Huang, Jianan Li, Yuze Xiao, Ning Shen, and Tingfa Xu. Rtnet: Relation transformer network for diabetic retinopathy multi-lesion segmentation. *IEEE Transactions on Medical Imaging*, 41(6):1596–1607, 2022.

[83] Yanfei Guo and Yanjun Peng. CARNet: Cascade attentive RefineNet for multi-lesion segmentation of diabetic retinopathy images. *Complex Intell. Syst.*, 8(2):1681–1701, April 2022.

[84] Yifei Xu, Zhuming Zhou, Xiao Li, Nuo Zhang, Meizi Zhang, and Pingping Wei. Ffu-net: Feature fusion u-net for lesion segmentation of diabetic retinopathy. *BioMed Research International*, 2021(1):6644071, 2021.

[85] QiHao Xu, Xiaoling Luo, Chao Huang, Chengliang Liu, Jie Wen, Jialei Wang, and Yong Xu. Hacdr-net: Heterogeneous-aware convolutional network for diabetic retinopathy multi-lesion segmentation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(6):6342–6350, Mar. 2024.

[86] Zahid Ullah, Muhammad Usman, Siddique Latif, Asifullah Khan, and Jeonghwan Gwak. Ssmd-unet: semi-supervised multi-task decoders network for diabetic retinopathy segmentation. *Scientific Reports*, 13(1), June 2023.

[87] Ufaq Khan, Mustaqeem Khan, Abdulmotaleb Elsaddik, and Wail Gueaieb. Ddnet: Diabetic retinopathy detection system using skip connection-based upgraded feature block. In *2023 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–6, 2023.

[88] K Ashwini and Ratnakar Dash. Grading diabetic retinopathy using multiresolution based cnn. *Biomedical Signal Processing and Control*, 86:105210, 2023.

[89] Rajasekhar Kommaraju and M.S. Anbarasi. Diabetic retinopathy detection using convolutional neural network with residual blocks. *Biomedical Signal Processing and Control*, 87:105494, 2024.

[90] Mahmoud Ragab, Wajdi H. Aljedaibi, Alaa F. Nahhas, and Ibrahim R. Alzahrani. Computer aided diagnosis of diabetic retinopathy grading using spiking neural network. *Computers and Electrical Engineering*, 101:108014, 2022.

[91] Xiaoling Luo, Wei Wang, Yong Xu, Zhihui Lai, Xiaopeng Jin, Bob Zhang, and David Zhang. A deep convolutional neural network for diabetic retinopathy detection via mining local and long-range dependence. *CAAI Transactions on Intelligence Technology*, 9(1):153–166, 2024.

[92] G. Kalyani, B. Janakiramaiah, A. Karuna, and L. V. Narasimha Prasad. Diabetic retinopathy detection and classification using capsule networks. *Complex amp; Intelligent Systems*, 9(3):2651–2664, March 2021.

[93] Muwei Jian, Hongyu Chen, Chen Tao, Xiaoguang Li, and Gaige Wang. Triple-drnet: A triple-cascade convolution neural network for diabetic retinopathy grading using fundus images. *Computers in Biology and Medicine*, 155:106631, 2023.

[94] Xiang Li, Yuchen Jiang, Jiusi Zhang, Minglei Li, Hao Luo, and Shen Yin. Lesion-attention pyramid network for diabetic retinopathy grading. *Artificial Intelligence in Medicine*, 126:102259, 2022.

[95] Guanghui Yue, Yuan Li, Tianwei Zhou, Xiaoyan Zhou, Yun Liu, and Tianfu Wang. Attention-driven cascaded network for diabetic retinopathy grading from fundus images. *Biomedical Signal Processing and Control*, 80:104370, 2023.

[96] Fatima, Muhammad Imran, Anayat Ullah, Muhammad Arif, and Rida Noor. A unified technique for entropy enhancement based diabetic retinopathy detection using hybrid neural network. *Computers in Biology and Medicine*, 145:105424, 2022.

[97] Serena Sunkari, Ashish Sangam, Venkata Sreeram P., Suchetha M., Rajiv Raman, Ramachandran Rajalakshmi, and Tamilselvi S. A refined resnet18 architecture with swish activation function for diabetic retinopathy classification. *Biomedical Signal Processing and Control*, 88:105630, 2024.

[98] Mohaimenul Azam Khan Raiaan, Kaniz Fatema, Inam Ullah Khan, Sami Azam, Md. Rafi Ur Rashid, Md. Saddam Hossain Mukta, Mirjam Jonkman, and Friso De Boer. A lightweight robust deep learning model gained high accuracy in classifying a wide range of diabetic retinopathy images. *IEEE Access*, 11:42361–42388, 2023.

[99] Prasanna Porwal, Samiksha Pachade, Ravi Kamble, Manesh Kokare, Girish Deshmukh, Vivek Sahasrabuddhe, and Fabrice Meriaudeau. Indian diabetic retinopathy image dataset (idrid): A database for diabetic retinopathy screening research. *Data*, 3(3), 2018.

[100] Melissa delaPava, Hernan Rios, Francisco J. Rodríguez, Oscar J. Perdomo, and Fabio A. González. A deep learning model for classification of diabetic retinopathy in eye fundus images based on retinal lesion detection. In Letícia Rittner, Eduardo Romero Castro M.D., Natasha Lepore, Jorge Brieva, and Marius George Linguraru, editors, *17th International Symposium on Medical Information Processing and Analysis*, volume 12088, page 120880T. International Society for Optics and Photonics, SPIE, 2021.

[101] Sraddha Das, Krity Kharbanda, Suchetha M, Rajiv Raman, and Edwin Dhas D. Deep learning architecture based on segmented fundus image features for classification of diabetic retinopathy. *Biomedical Signal Processing and Control*, 68:102600, 2021.

[102] Lingling Fang and Huan Qiao. A novel dag network based on multi-feature fusion of fundus images for multi-classification of diabetic retinopathy. *Multimedia Tools and Applications*, 82(30):47669–47693, May 2023.

[103] M. Usman Akram, Shehzad Khalid, Anam Tariq, Shoab A. Khan, and Farooque Azam. Detection and classification of retinal lesions for grading of diabetic retinopathy. *Computers in Biology and Medicine*, 45:161–171, 2014.

[104] Doaa Elsawah, Ahmed Elnakib, and Hossam El-Din Moustafa. Deep learning grading system for diabetic retinopathy using fundus images. (dept. e). *MEJ. Mansoura Engineering Journal*, 45(4):1–8, October 2020.

[105] Kamble R Kokare M Deshmukh G Sahasrabuddhe V Meriaudeau F Porwal P, Pachade S. Indian diabetic retinopathy image dataset (idrid): A database for diabetic retinopathy screening research., 2018.

[106] J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever, and B. van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4):501–509, 2004.

[107] Sohier Dane Karthik, Maggie. Aptos 2019 blindness detection, 2019.

[108] Jorge Will Cukierski Emma Dugas, Jared. Diabetic retinopathy detection, 2015.

[109] Jun Ma, Feifei Li, and Bo Wang. U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722*, 2024.

[110] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Proceedings of the IEEE international conference on computer vision*, pages 1395–1403, 2015.

[111] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.

[112] Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, and Pheng-Ann Heng. H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. *IEEE transactions on medical imaging*, 37(12):2663–2674, 2018.

[113] Hongxing Peng, Chao Xue, Yuanyuan Shao, Keyin Chen, Juntao Xiong, Zhihua Xie, and Liuhong Zhang. Semantic segmentation of litchi branches using deeplabv3+ model. *Ieee Access*, 8:164546–164555, 2020.

[114] Minghao Yin, Zhuliang Yao, Yue Cao, Xiu Li, Zheng Zhang, Stephen Lin, and Han Hu. Disentangled non-local neural networks. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV 16*, pages 191–207. Springer, 2020.

[115] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, et al. Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 43(10):3349–3364, 2020.

[116] Xiangxiang Chu, Zhi Tian, Yuqing Wang, Bo Zhang, Haibing Ren, Xiaolin Wei, Huaxia Xia, and Chunhua Shen. Twins: Revisiting the design of spatial attention in vision transformers. *Advances in neural information processing systems*, 34:9355–9366, 2021.

[117] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.

[118] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer, 2022.

[119] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, et al. Swin transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12009–12019, 2022.

[120] Qing Liu, Haotian Liu, Wei Ke, and Yixiong Liang. Automated lesion segmentation in fundus images with many-to-many reassembly of features. *Pattern Recognition*, 136:109191, 2023.

[121] QiHao Xu, Xiaoling Luo, Chao Huang, Chengliang Liu, Jie Wen, Jialei Wang, and Yong Xu. Hacdr-net: Heterogeneous-aware convolutional network for diabetic retinopathy multi-lesion segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 6342–6350, 2024.

[122] Jiong Zhang, Behdad Dashtbozorg, Erik Bekkers, Josien PW Pluim, Remco Duits, and Bart M ter Haar Romeny. Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores. *IEEE transactions on medical imaging*, 35(12):2631–2644, 2016.

[123] Zhongwei Feng, Jie Yang, and Lixiu Yao. Patch-based fully convolutional neural network with skip connections for retinal blood vessel segmentation. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 1742–1746. IEEE, 2017.

[124] Lin Gu, Xiaowei Zhang, He Zhao, Huiqi Li, and Li Cheng. Segment 2d and 3d filaments by learning structured and contextual features. *IEEE transactions on medical imaging*, 36(2):596–606, 2016.

[125] Song Guo, Kai Wang, Hong Kang, Yujun Zhang, Yingqi Gao, and Tao Li. Bts-dsn: Deeply supervised neural network with short connections for retinal vessel segmentation. *International journal of medical informatics*, 126:105–113, 2019.

[126] Yi Lin, Honggang Zhang, and Guang Hu. Automatic retinal vessel segmentation via deeply supervised and smoothly regularized network. *IEEE Access*, 7:57717–57724, 2018.

[127] Di Li, Dhimas Arief Dharmawan, Boon Poh Ng, and Susanto Rahardja. Residual u-net for retinal vessel segmentation. In *2019 IEEE international conference on image processing (ICIP)*, pages 1425–1429. IEEE, 2019.

[128] Alex Krizhevsky. One weird trick for parallelizing convolutional neural networks. *CoRR*, abs/1404.5997, 2014.

[129] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[130] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[131] Ningning Ma, Xiangyu Zhang, Hai-Tao Zheng, and Jian Sun. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In *Proceedings of the European conference on computer vision (ECCV)*, pages 116–131, 2018.

[132] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[133] Xiaoling Luo, Wei Wang, Yong Xu, Zhihui Lai, Xiaopeng Jin, Bob Zhang, and David Zhang. A deep convolutional neural network for diabetic retinopathy detection via mining local and long-range dependence. *CAAI Transactions on Intelligence Technology*, 9(1):153–166, 2024.

[134] Sara Hosseinzadeh Kassani, Peyman Hosseinzadeh Kassani, Reza Khazaeinezhad, Michal J Wesolowski, Kevin A Schneider, and Ralph Deters. Diabetic retinopathy classification using a modified xception architecture. In *2019 IEEE international symposium on signal processing and information technology (ISSPIT)*, pages 1–6. IEEE, 2019.

[135] Debjyoti Sinha and Mohamed El-Sharkawy. Thin mobilenet: An enhanced mobilenet architecture. In *2019 IEEE 10th annual ubiquitous computing, electronics & mobile communication conference (UEMCON)*, pages 0280–0285. IEEE, 2019.

[136] Hossein Shakibania, Sina Raoufi, Behnam Pourafkham, Hassan Khotanlou, and Muharram Mansoorizadeh. Dual branch deep learning network for detection and stage grading of diabetic retinopathy. *Biomedical Signal Processing and Control*, 93:106168, 2024.