REPUBLIC OF TÜRKİYE

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Information Technologies

RETINAL EYE DISEASE DETECTION USING DEEP LEARNING

Saja Salman Ali AL HAMEEDAWI

Master's Thesis

Supervisor

Asst. Prof. Dr. Muhammad ILYAS

İstanbul, 2024

RETINAL EYE DISEASE DETECTION USING DEEP LEARNING

Saja Salman Ali AL HAMEEDAWI

Information Technologies

Master's Thesis

ALTINBAŞ UNIVERSITY

2024

The thesis titled RETINAL EYE DISEASE DETECTION USING DEEP LEARNING prepared by SAJA SALMAN ALI AL HAMEEDAW and submitted on 18/12/2024 has been **accepted unanimously** for the degree of Master of Science in Information Technologies.

Asst. Prof. Dr. Muhammad ILYAS

Supervisor

Thesis Defense Committee Members:

Asst. Prof. Dr. Muhammad ILYAS	Department of Computer Engineering, Altınbaş University	_____
Asst. Prof. Dr Abdullahi Abdu IBRAHIM	Department of Computer Engineering, Altınbaş University	_____
Assoc. Prof. Dr. Jawad RASHEED	Department of Computer Engineering İstanbul Sabahattin Zaim University	_____

I hereby declare that this thesis meets all format and submission requirements of a master's thesis.

I hereby declare that all information/data presented in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare all unoriginal materials and conclusions have been cited in the text and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the abovementioned rules and conduct.

Saja Salman Ali AL HAMEEDAWI

Signature

ABSTRACT

RETINAL EYE DISEASE DETECTION USING DEEP LEARNING

AL HAMEEDAWI, Saja

M.Sc., Information Technologies, Altınbaş University

Supervisor: Asst. Prof. Dr. Muhammad ILYAS

Date: 12/2024

Pages: 88

The global prevalence of retinal abnormalities affects millions of individuals, highlighting the urgency of early detection and intervention to prevent the advancement of these conditions, ultimately mitigating the risk of avoidable blindness. In this thesis, we delve into the critical realm of retinal disease detection using deep learning techniques. Leveraging a diverse ensemble of state-of-the-art neural network architectures, including MobileNetV2, ResNet50, InceptionV3, and DenseNet, we conduct a comprehensive evaluation of their performance in classifying retinal scans. Our meticulous preprocessing steps, resize images, convert grayscale to RGB and rigorous training cycles lay the foundation for an advanced model. Notably, our results reveal that ResNet50 outperforms other models, achieving an accuracy of 0.89, setting a new benchmark in retinal scan analysis. This research contributes to the vital field of early retinal disease detection, offering the potential to enhance clinical diagnosis and patient outcomes.

Keywords: Retinal Abnormalities, Disease Detection, Deep Learning, ResNet50, MobileNetV2, InceptionV3, DenseNet.

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

ABBREVIATIONS

AI	:	Artifical Intelligence
DL	:	Deep Learning
ML	:	Machie Learnig
CNN	:	Convolution Neural Network
VGG	:	Visual Geometry Group
ACC	:	Accuracy
PREC	:	Precision
REC	:	Recall
F1-S		F1-Score

1. INTRODUCTION

1.1 BACKGROUND

Diabetic retinopathy (DR) stands as a prominent consequence of the worldwide diabetes epidemic, emerging as the primary instigator of adult blindness in the United States. It materializes due to the insidious effects of diabetes mellitus, which precipitates a myriad of complications. The propensity for an individual to develop DR amplifies with prolonged diabetes duration, elevated blood sugar concentrations, and arterial hypertension [1][2]. Initially, DR targets the intricate network of blood vessels situated within the retina, a light-sensitive tissue nestled in the posterior section of the eye. Prolonged exposure to heightened glucose levels compromises these vessels, culminating in capillary endothelial damage. This devastation prompts adjacent capillary cells to become permeable, or "leaky". The lethal synergy of escalated glucose concentrations and endothelial injury converges to induce capillary occlusion, subsequently curtailing blood circulation [3].

Diabetic retinopathy is a severe visual impairment worldwide affecting a wide spectrum of the global population which emerges from the greater spectrum of diabetic complications. Ophthalmic examinations are fundamental in allowing clinicians to identify the presence of this pathology. Still, given the comparatively large number of retinal images, it is difficult to scrutinize them well enough for a precise diagnosis. In this digital age, computer-aided decision-support systems are a viable option, allowing medical workers to perform an exact diagnosis, thus increasing the time-effectiveness of the process. This paper aims to examine current computer-aided research on DR. While its effects are only limited to the retina, similar to this organ, the heart or kidneys are vital organs [4] [5] [6]. Once again, the International Diabetes Federation presents compelling evidence that out of over 537 million people with diabetes, a remarkable 90 million people find it difficult to cope with the nuances of DR [7]. When one disentangles the pathophysiology surrounding DR, the frequency at which it occurs indicates that DR disrupts the retina by instigating vascular inflammation and intraocular fluid imbalance. This pathology doesn't merely hinder vision but escalates to irreversible blindness. To elucidate, DR accounts for an estimated 2.6% of global blindness incidences [8] [9].

Diabetic retinopathy occupies a complex category, which represents a variety of damaging diabetes consequences focusing on the retina as a site for possible visual impairments and sight loss (including blindness) in the most serious cases [10]. Therefore, diagnosing the condition from the early stages, not only allows us to have the ability to save the eye from its damage but also will be useful in protecting against the possible blindness [11]. The venerated way of uncovering the signs of diabetic retinopathy goes around with different crisscrosses of challenges. The manual diagnostics that include the unbearable volume of data and the weaknesses in classification being two of the most significant ones [12] consume the prodigious amount of time, and it is not clear if decisions are sound. On one hand, through the help of computer technologies, machine learning algorithms, and artificial intelligence, the nearness of healthcare tools elevated the ability of deciding the maximum retinopathy complications from anterior locations and with great accuracy [13]. This population group found its counterpart in an alarming number of impoverished countries - the statistics show that 75% of the burden currently falls on them [14]. No longer having the possible diagnostic apparels and tribal, these surroundings are insistently calling for rational solutions. These forerunners of decision-support systems which are designed for the diagnosis of diabetic retinopathy serve as the symbol of hope for the early diagnosis that could actually turn the disease into a thing of successfully managed past. Besides the specific criteria of diagnosing the disease, the retinal imaging transforms into the scene on which the key signs of the disease are laid to exposed. Distinctive lesions, i.e., hemorrhages (HM), microaneurysms (MA), together with soft and hard exudates (EX) play a role as an intrinsic warning signal of diabetes retinopathy, pointing to the disease's progress, the stage of development [15].

In the era preceding 1961, the quest to detect diabetic retinopathy navigated through a mosaic of diagnostic avenues. Primarily, a visual acuity examination charted the landscape of an individual's vision, assessing the clarity with which they perceived objects across varying distances. Furthermore, to attain an intricate view of the retina and the optic nerve, a dilated eye examination was orchestrated, scouring for any vestiges of damage. Complementing these examinations, tonometry played a pivotal role, meticulously gauging the intraocular pressure and identifying any deviations from the norm, specifically readings falling below 12 mmHg or soaring above 22 mmHg [16].

However, the diagnostic medicine tapestry changed beyond all recognition with the expanding arsenal of avant-garde imaging techniques. The technique that stood out among the rest was fluorescein angiography, with no doubt an example of careful and inventive thinking. It worked by introducing a dye that was luminescent under a camera to the eye, thus illuminating areas of vascular leakage and ischemia, thus opening up unprecedented opportunities for examination of the retinal state [17] [18]. Despite its unprecedented efficacy as a gold standard of diabetic retinopathy diagnostics, the invention of the optical coherence tomography had merely overlooked the field in the dustbin of history. The optical coherence tomography was like a star of technical brilliance; it enjoyed a lack of need for pupil dilation and integrated fully with a telemedicine field [19] [20] [21]. Therefore, rapid and accurate diagnosis on a constant and repeated basis saved not only lives but also resources for the developed and the underdeveloped countries . For the first time, retina anatomy had been visualized at such an ultra-detailed level, and the thinnest retinal lesions could be brought to the light and identified.

The modern medical scenario rarely bears the tide of compromise and instead conducts an incessant search for perfection, resulting in the emergence of new fundus photography modalities. New methodologies such as the wide-field fluorescein angiography are currently under rigorous observation, while some have already found their way into the diagnostic armamentarium . A salient instance of the latter is the ultra-widefield imaging system Optos 200 Tx, a marvel crafted by Optos plc. In comparison with contemporary imaging techniques, the UWF devices provide a view of 200 degrees. Such a scope emplaces the anatomical surroundings of the examined area into the hands of a diagnostician enhancing the acuity of one’s judgment. In addition, the new devices promise quick image processing and provide a user-friendly environment for cloning and upon it manipulation. The field of medicine is currently encountering another exponential: the machine learning scenario. These computational methodologies, renowned for their superior classification efficiency, ventured into the diagnostic realm . A subdivision of machine learning, known as deep learning, employs multi-level architectures to analyze intricately abstracted images . In this approach, each level integrates the knowledge of the previous level autonomously and builds a novel large-scale understanding of image features. Deep learning models excel when abundant data are involved. Hence, as long as the neural network is sufficient, the models can learn and accurately select the features of an image with few to no prior descriptions of

training. Eminent scholars such as LeCun et al. [22], Liu et al. [23], and Litjens et al. [24] have endeavored to unravel the potential of deep learning in medical image analysis.

In the world of Optical coherence tomography images, past research has relied on marking important features, such as vessels, and line up the diabetic retinopathy stage using classifiers. However, a new era of ophthalmic research has begun: pioneers like Pratt et al. [25], Sangeethaa et al. [26] have crossed the new frontier of diagnosing diabetic retinopathy by using the Convolution Neural Network (CNN). This complexity of computations is visualized via three key layers: the input, hidden, and output layers. The hidden layer is a "black box" with complicated convolution and pooling operations designed to reduce computation. CNNs, the crown jewel of deep learning, outperform past methods by avoiding handcrafting features and offering an adaptable network for classifying on a range of tasks. Despite the fact that CNNs have demonstrated their ability to mine OCT images to study diabetic retinopathy I stage, how well these networks would perform in ultra-widefield fundus images is not well understood, especially on different diabetic retinopathy stages.

1.2 PLANNING THE PROBLEM

Retinal diseases include a plethora of conditions including Diabetic Retinopathy, Age-Related Macular Degeneration, and Glaucoma ranking high on the list of the most feared eye diseases if they are not diagnosed and treated early. The retina, a sophisticated, light-responsive coating located at the eye's rear, is the major organ that integrates vision. Vision is the most precious sense that health directly determines an individual's capability to discern visual world. Although retina examination is very significant, the early discovery and accurate diagnosis of retinal diseases is the most problematic area of the ophthalmology. While manual analysis of fundus photos coupled with traditional diagnostic approaches have had some positive results, the issue of scalability when dealing with individuals at high risk of these conditions, especially in resource constrained zones has also posed great challenges. In addition, these traditional methods that usually depend on the expert physician experience and their own mind can sometimes not see the first signs of the disease, or fail to provide consistent results from different doctors. Against the background of the emergent diabetes pandemic and of an aging population that is concurrently becoming high-risk patients for retinal diseases, we hereby argue for the need for more rapid, reliable and precision diagnostic approaches. The technological adoption in medical imaging, which uses broadly

deep learning techniques, is the one that shows the way forward. However, some smaller scale studies have shown that with adequate preparation, these algorithms are able to be either at the same or better level as the professionals with extensive experience. Although the thorough research is dictated to understand the undisclosed effectiveness, challenges, limitations, and chances of utilizing deep learning for retinal eye disease diagnostic, nonetheless. The lack of efficient screening facilities can hinder the diagnosis of retinal diseases. The proposed thesis aims to address this gap by developing, validating, and optimizing the deep learning models specifically for this difficult task. Hence, the future of eye care can be influenced by the transformative model of screening.

1.3 THESIS MOTIVATION

The human eye, a part of anatomy traditionally regarded as a 'gateway to the soul', is, however, so much more than just a mere metaphor; instead, the eye is a portal to one's entire health. Systemic diseases show themselves on the retina canvas, like diabetes, hypertension, and also certain infectious diseases, prior to the development of full-blown condition. Yet a sizable number of the world population is unfortunately unaware of retinal conditions as they do not realize the threat to their eyes until the diseases are too advanced and treatment becomes difficult and prognosis becomes grim. The imperative for early and accurate detection of retinal diseases has never been as clear as it is today, particularly in the era where high prevalence of conditions such as diabetes is rocketing. Although advanced and innovative techniques and equipment have undoubtedly brought about major improvements in accuracy, consistency and scalability, the traditional diagnostic methodologies still retain some their disadvantages. Not withstanding the inequalities in the healthcare provision and the lack of access to expert ophthalmological assessments for many individuals in resource-constrained regions globally. Deep learning in medical diagnostics is the future, which portends a new horizon of such deep transformations. These algorithms can be successfully leveraged to make diseases of the retina more accessible to all, by introducing accurate and scalable detection solutions which overcome economic and geographic barriers. The prospect of raising the accuracy of a diagnosis, speeding up the process of patient care and eventually limiting a global vision loss epidemic becomes our thesis driving force. By means of the research, we are going to bring technology in diagnostics of eye diseases to a new level and together we can reduce preventable vision loss to a significantly lower level.

1.4 THESIS OBJECTIVES

Our research focuses to provide the creation of a highly effective deep learning model employed for identifying of the retinal eye illnesses represented by the most advanced methods as its main objective. In the light of the necessity from big, heterogeneous and correctly annotated retinal datasets, and the immense need to use a transfer learning method that has been pre-trained on big, complex images from the Internet to save a lot of computational power and resources as we need to train the deep neural networks carefully from scratch, our thesis will specifically recommend integrating transfer learning methods. Transfer learning, a methodology where pre-trained convolutional models are fine-tuned for specific tasks using smaller datasets, has demonstrated success in different domains due to its efficiency and the benefits of using learned characteristics from one task to facilitate performance on another. Among various data-centric strategies, we focus upon channelizing this methodology to gain more by using the knowledge that is present within the brain of a neural network and fit them for the complicated task of retina disease detection. The core objectives include: analyzing the system architecture of different models that are to architecture that are well fitted to the retinal datasets to carry out the disease-specific features, and that the model is well tested compare it to the traditional methods. The models created in the course of our research not only guarantee high accuracy and speed but they also contribute to establishing a platform where people regardless of their place of residence can get a screening without much hassle.

1.5 THESIS CONTRIBUTION

The contribution of this research is:

a. Novel Application of Transfer Learning: The research presented here specifies the usage of transfer learning specifically for retinal eye disease recognition. In the literature it constitutes a negative gap which is now filled by this case study, and as a result, it provides a general plan for the follow-up activities in this area.

b. Benchmarking and Performance Metrics: Due to the exploration of multiple pre-trained neural network architecture models, our thesis offers a fundamental index of retinopathy

detection, the index could be useful in future model design and optimization as metrics can be used as reference.

c. Optimized Fine-tuning Strategy: Through the process of systemic testing our efforts offer the best practices for on- boarding pre-trained models onto datasets related to retina so that diagnostics with the highest accuracy and the least expensive resources can be achieved.

d. Enhanced Accessibility and Scalability: The research presents the transfer learning model as the solution for having scalability across multiple datasets, thus it paves the way for the emergence of more standard diagnostic tools which can serve a wider population from developed settings and the resource-constrained regions.

e. In-depth Analysis of Model Robustness: This work provides a thus far unreported look at the disadvantages and the obstacles that current deep learning model design in imaging for medical purposes has, and also the strategies for fixing it to enable it to perform better in real use.

f. Interactive Dataset and Model Repository: This is part of the extension to our research where we are going to develop and publish an annotated retinal dataset, and the trained model weights, as a side effect. This will provide to achieve higher precision of the models, also helping on collaborative enhancements.

g. Comprehensive Clinical Integration Blueprint: Beyond the development of a diagnostic model, our thesis provides guidelines and recommendations for integrating the proposed solution into clinical workflows, ensuring that the research translates to tangible real-world impact.

h. Insights on Model Interpretability: In the realm of medical diagnostics, understanding the 'why' behind model predictions is crucial. Our research delves into methods that enhance the interpretability of our deep learning model, ensuring clinicians can trust and understand the diagnostic outputs.

1.6 THESIS ORGANIZATION

The research blueprint for this thesis unfolds in a structured manner.

Literature Review: This chapter delves into retinal eye diseases, their prevalence, and diagnostic challenges. It transitions into machine and deep learning, emphasizing their relevance in healthcare. Key concepts like transfer learning are introduced, with a focus on its role in enhancing deep learning models. The chapter concludes by highlighting previous research in machine and deep learning for retinal disease detection.

Methodology: This chapter introduces the research methodology, giving a description of the dataset and the attributes that can make it challenging during the course of the analysis. It tells us first about data preparation procedures and then jumps to details of the main research, a proposed model. The construction, design, and transfer learning setting where the custom pre-trained models are used for retinal imaging are the main components.

Results and Discussion: The research is going to start by presenting the empirical evidence obtained from a literature search. Following through with evaluation metrics and methods, the proposed model subsequently is presented. Thusly, the comparative analysis with other benchmarks then ensues into a critical discussion on findings, their implications and thus their alignment with what current literature states.

Conclusion and Future Work: Chapter highlights the research's influence on diagnostic eye problems, reflects on the constraints and gives the possible directions for further exploration of deep learning and retinal imaging.

2. LITERATURE REVIEW

2.1 INTRODUCTION

This chapter of the paper is focused on the multifaceted picture of Diabetic Eye Disease (DED) and its existing issues. AI is vital and our work explores machine and deep learning algorithms which have broken new ground in current medical image analysis framework. Besides, we also explore the use of transfer learning methods and their abilities to capture DED-detection complexities with different nuances. We will perform an intricate dissection of previous works related to ours, which will showcase the changes and discoveries, which helped create the framework of our proposed model.

2.2 DIABETIC EYE DISEASE

2.2.1 Structure of the Eye

The human eye, an intricate organ with a multifaceted structure, plays a pivotal role in the process of vision. At a cursory glance, the external components of the eye that are readily discernible include the sclera, a tough white outer covering; the cornea, a transparent front surface; the iris, which gives our eyes their distinctive color; and the pupil, a black circular opening at the center of the iris. Delving deeper into the eye's internal anatomy, one encounters the retina, a thin layer of tissue lining the back of the eye. Within the retina lies the macula, a small central region that facilitates sharp, detailed vision, and the fovea, a tiny pit located in the macula that provides the clearest vision of all. Adjacent to these is the optic disc, the point of exit for ganglion cell axons leaving the eye, and the posterior pole, an area encompassing the central retina. As depicted in Figure 2.1, the process of vision is initiated when light strikes the cornea. This semi-focused light then traverses the pupil, reaching the lens, which further refines the focus. Post this fine-tuning, the image travels through the jelly-like vitreous humor, eventually projecting onto the macula [30]. The macula, due to its specialized cells, enables intricate visual tasks such as reading, writing, and discerning colors. In contrast, the peripheral region of the retina aids in peripheral or side vision. Functionally, the retina acts as a dynamic converter, transforming incoming light into neural signals. These signals, carried by the optic nerve, are relayed to the brain, where they are interpreted as images [31]. It's intriguing to note that the retina, given its neural origins

during embryonic development, is essentially an extension of the brain. To maintain its functionality, the retina is nourished by an intricate network of blood vessels. However, systemic conditions like diabetes can compromise these vessels, undermining the retina's capabilities and potentially leading to visual impairment.

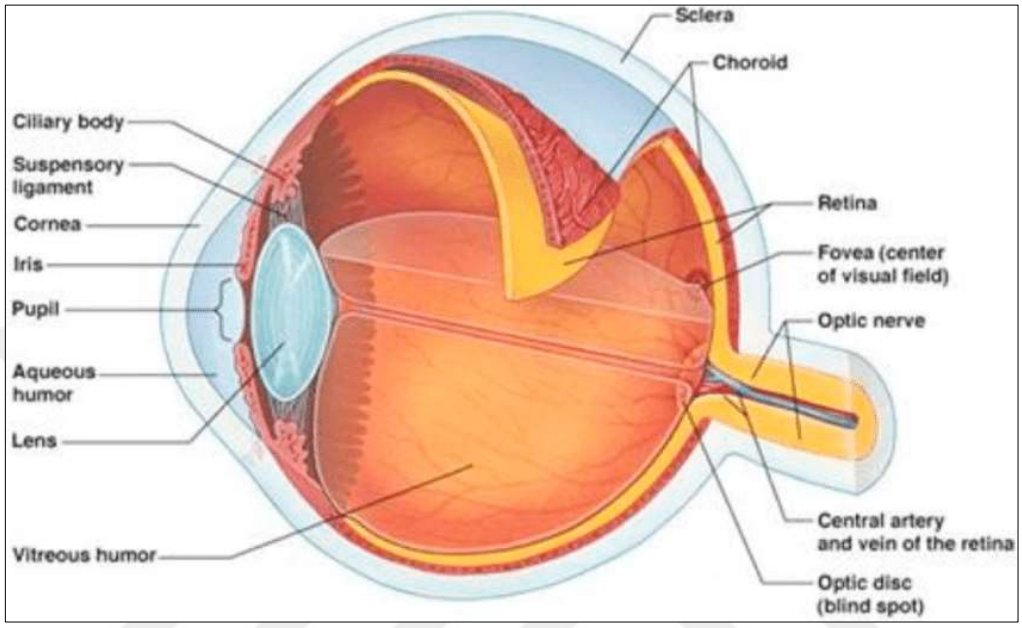

Figure 2.1: Structure of The Eye [28]**.**

2.2.2 Prevalence and Types of Diabetic Eye Disease

Diabetic Eye Disease (DED) encompasses a spectrum of ocular conditions intrinsically linked to diabetes, characterized by their potential to compromise vision, often leading to irreversible blindness. The umbrella of DED covers a gamut of conditions, namely, Diabetic Retinopathy, Diabetic Macular Edema, Glaucoma, and Cataract [32]. It's worth noting that individuals aged between 20 and 74 are particularly susceptible to these ocular manifestations of diabetes, emphasizing the significant span of the affected demographic. As per data from the International Diabetes Federation (IDF), the global prevalence of diabetes in 2017 was alarmingly high, affecting approximately 425 million people. This number is predicted to experience a steep incline, potentially reaching an estimated 692 million by 2045 [33]. Such staggering figures underscore the profound public health impact of diabetes, not only from a medical perspective but also considering its broad social and economic repercussions. In fact, the toll of diabetes on global mortality is so profound that it ranks as the fourth leading cause of death worldwide [34].

The multifaceted manifestations of diabetes extend beyond systemic symptoms to impact specific organs, notably the eye, with the retina being a primary site of diabetic complications. To provide a clearer context, one can envision the retina's normal anatomy as depicted in Figure 2.2. However, in the throes of DED, as illustrated in Figure 2.3, the retinal landscape undergoes detrimental changes. The inception of severe DED is often marked by aberrant blood vessel growth, consequential optic nerve damage, and the appearance of hard exudates concentrated in the macula, which is the central region of the retina responsible for sharp, central vision. While there are four predominant types of DED, each with its distinct pathophysiology, what unites them is their shared propensity to jeopardize visual acuity. Detailed elucidation of these specific DED types and their respective impacts on the eye will be furnished in the subsequent subsections.

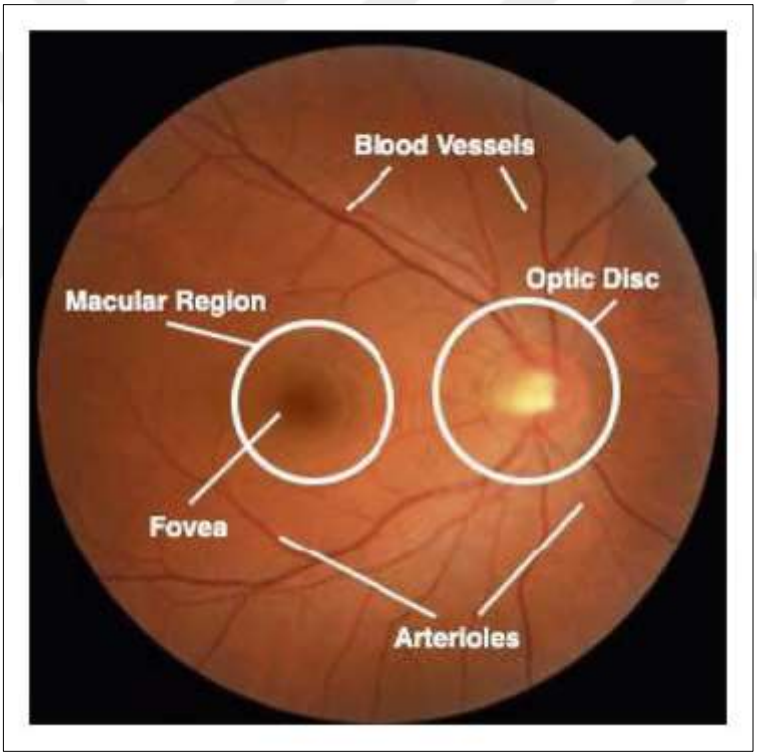

Figure 2.2: Normal Anatomical Structures of the Retina [27].

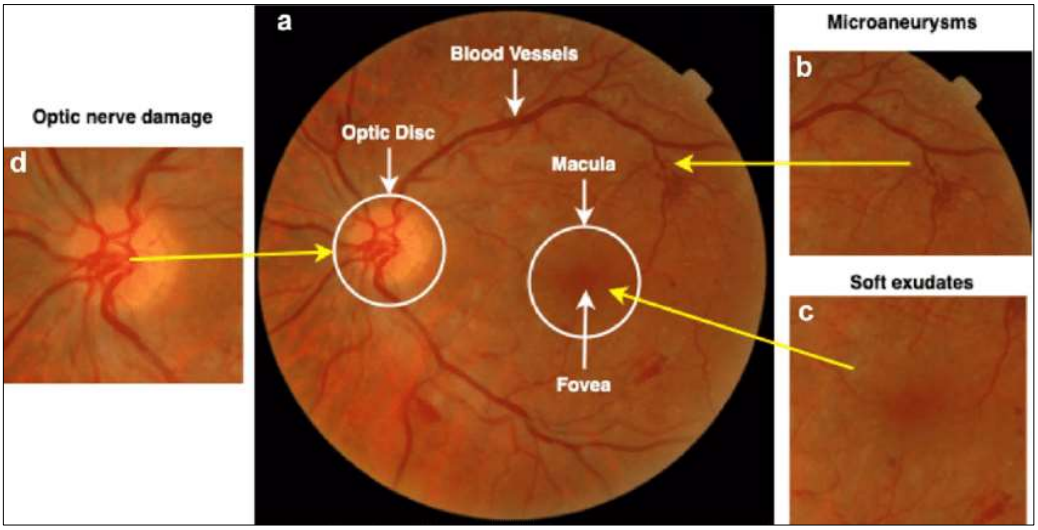

Figure 2.3: Complications of Diabetic Eye Disease in the Retina [28].

2.2.3 Diabetic Retinopathy

Diabetic Retinopathy (DR) emerges as a complication of diabetes, primarily manifesting as a result of sustained damage to the delicate blood vessels that feed the retina. Located at the eye's posterior, the retina is a critically significant layer of light-sensitive cells. Its primary function is to perceive incoming light and, in turn, transmit these light signals to the brain. It is the brain's task to then interpret these signals, enabling us to construct and recognize the visual world around us [35]. DR can be broadly categorized into two pivotal stages: the early stage and the advanced stage.

In its early stage, also known as non-proliferative diabetic retinopathy (NPDR), DR exhibits a conspicuous absence of proliferating or newly developing blood vessels. This stage witnesses the blood vessel walls within the retina undergoing a weakening process. Consequently, small, narrowed protrusions, referred to as microaneurysms, begin to emerge from the weakened vessel walls. These microaneurysms often leak, spilling blood and other fluids into the retina's domain. Concurrently, some of the broader retinal vessels may undergo dilation and manifest inconsistencies in their diameter. As more vessels fall victim to blockages, NPDR's intensity escalates, ranging from mild to severe stages. Depending on the progression's severity, one may observe nerve fibers within the retina experiencing swelling. Particularly concerning is when this swelling invades the macula, the retina's central region instrumental for focused, central vision. This condition, termed macular

edema, necessitates immediate medical intervention. To offer a clearer gradation, NPDR can be subdivided into three distinct severity levels: mild, moderate, and severe [36].

Transitioning to the advanced stage of DR, we encounter what is termed proliferative diabetic retinopathy (PDR). At this juncture, the previously damaged vessels begin to leak into the vitreous - the gel-like substance that fills the eye's core. This leakage fosters an environment conducive for the growth of abnormal blood vessels within the retina. A significant concern arises when these newly sprouted vessels impede the vitreous's typical fluid flow, leading to an undesirable pressure buildup within the eye. Such an elevated intraocular pressure poses a direct threat to the optic nerve, responsible for ferrying visual data from the eye to the brain. If left unchecked, this can culminate in glaucoma, a condition characterized by irreversible damage to the optic nerve, often culminating in blindness.

2.2.4 Diabetic Macular Edema

Diabetic Macular Edema is a major vision-related complication of diabetes. It is primarily a result of the accumulation of fluid in the macula, which is centrally located in the retina. The macula is a highly sensitive, central and priority region of the retina. Its main function involves providing sharp and clear central forward vision; for instance, it is the region used in reading, driving, or identifying people’s faces. In the case where the developing diabetes affects the blood vessels, the vessels will become highly susceptible to damage. Among the worst affected vessels are those responsible for feeding the retina. The damage often causes leaking, allowing fluids to penetrate the macula. The fluids will then cause the macula to swell and thicken, thus distorting the vision. As such, the distorted vision, even the DME, can present as mild to severe blurriness, varied with the degree of interstitial fluid accumulation [35].

To offer a structured understanding of DME's progression, it has been classified into stages, namely: mild, moderate, and severe. Each stage is characterized by specific retinal changes, which serve as indicators of the edema's intensity [37]. In the mild stage, there's a noticeable thickening of the retina, particularly in the fovea, either equivalent to or less than 500 μ, or up to a third of the disc's diameter. As DME transitions to its moderate phase, hard exudates begin to surface. These are lipid residues that often accompany retinal thickening and can be observed within 500 μ of the fovea. The severe stage of DME is discerned when the retinal

thickening exceeds one disc diameter, approximately 1500 μ, but remains within one diameter's distance from the fovea. Each of these stages, while distinct, underlines the progressive nature of DME and the escalating threat it poses to clear vision.

2.2.5 Glaucoma

Glaucoma (Gl) stands as one of the most insidious ocular conditions, primarily owing to its ability to progressively and irreversibly damage the optic nerve, a critical conduit that bridges the eye and the brain. The optic nerve plays a pivotal role in the visual process, transmitting visual signals from the retina to the brain, where these signals are interpreted as the images we see. The integrity of this nerve can be compromised by elevated fluid pressure within the eye, commonly referred to as intraocular pressure (IOP). Alarmingly, heightened blood sugar levels, a hallmark of diabetes, can exponentially increase an individual's susceptibility to glaucoma. The repercussions of unchecked glaucoma are dire, often culminating in irreversible blindness and a gradual loss of peripheral, then central vision. Given its stealthy progression, early detection and intervention are crucial to mitigating its debilitating impact.

Regarding the spectrum of glaucoma categorization, classification is based on morphological changes seen in the eye which targets a special structurally distinctive feature – the optic nerve head, also called the optic disc. The core measurement in this aspect is the Cup-to-Disc Ratio, which indicates how much cupping, or concavity, is formed in the disc. According to this feature, the condition has at least three types, while glaucoma, similarly to many other eye conditions, has progression, which means it can develop in three stages: mild, moderate, and severe [38]. Each severity stage corresponds to the stage of the optic nerve damage, eye vision deterioration, and treatment difficulties.

2.2.5.1 Cataracts

Cataract (Ca) emerges as a condition characterized by the clouding of the eye's natural lens, a structure pivotal for focusing light onto the retina to produce clear images. The genesis of this clouding can be attributed to the degeneration of lens proteins, a process that can be exacerbated by elevated sugar levels in the bloodstream. This degeneration manifests as the formation of opaque or blurry areas on the lens, consequently leading to visual distortions like blurriness, glare, and a reduction in color intensity. Notably, individuals with diabetes

face an amplified risk in this context. Their susceptibility to developing cataracts is not only heightened, but the onset of the condition also tends to occur earlier in life compared to their non-diabetic counterparts. To delineate the progression of this lens opacity, cataracts have been categorized into a spectrum of severity ranging from non-cataractous (or clear lens) to mild, moderate, and culminating in the severe stage, where the lens becomes significantly opaque and vision is drastically compromised [39].

The link between diabetes and ocular complications is underscored by the fact that diabetic patients manifest a heightened vulnerability to Diabetic Eye Disease (DED). Given the potential implications of DED, such as irreversible vision loss, the emphasis on its early detection has intensified, encompassing both adults and children within its purview. Pertinent research underscores the gravity of the situation, with findings indicating that an overwhelming 90% of diabetic patients can potentially circumvent the advancement of DED provided it's detected and intervened upon at an early stage [40]. However, the conventional methods of detecting DED, which rely heavily on manual examinations without the leverage of computer-assisted technologies, present several challenges. Not only do they extend the time interval between diagnosis and therapeutic intervention, but they also grapple with the subtlety of DED's initial manifestations. Such nuanced changes, often imperceptible in their early stages, can easily elude even the seasoned eyes of experts, underscoring the imperative need for more advanced and precise diagnostic tools.

2.2.6 Overview of Retinal Imaging Modalities

Over the span of many decades, marked by relentless research and innovation, retinal imaging has crystallized as the linchpin in the clinical evaluation, management, and treatment of patients plagued by both retinal and systemic maladies [31]. The ceaseless endeavors in the domain of retinal imaging have borne fruit in the form of superior visualization capabilities, allowing clinicians to delve deeper into the intricate landscape of the retina. This enhanced visibility into retinal pathophysiology has invariably ushered in a new era where early detection, precise diagnosis, and efficacious management of multifarious chorio-retinal anomalies have become the norm rather than the exception. Among the plethora of imaging techniques birthed by this discipline, Fundus photography and Optical Coherence Tomography (OCT), along with their respective variants, reign supreme [41]. They have solidified their positions as the gold standards in retinal imaging,

largely owing to their unparalleled accuracy and versatility. In the subsequent discourse, we embark on a detailed exploration of these two seminal modalities, shedding light on their intrinsic methodologies, their advantages, and their indomitable role in diagnosing an array of retinal afflictions.

2.2.7 Fundus Imaging

Fundus photography being a leading tool available at the present level of laser technology, this opens up a door towards an observation of the minute details of retina organization inside the eyeball. As analysed by [23], this modality is capable of complexities in 3-D tissues of retina and can be generalised into 2-dimensional (2-D) projection. Method of operation is based on the theory of helium reflection principle. Light rays are sent to the retina and reflected off it. The image that is obtained in the end is a result of the variation in intensities, for each individual reflection have been captured with unmatched accuracy. The intensities of shades gradients cope with the intricacies of retinal tissue organization that reflect the structure and correspond to the function of the part, providing a detailed, layered view. Nevertheless, the fish bowl of fundus imaging is neither a one dimensional physicality. Technological innovations and intensive research that have taken place for quite some time now have enriched the nutritional offers and thereby bred a multitude of types of the product. Mostly reputable among them are scanning laser ophthalmoscopy which makes use of laser as a more accurate source of information and adaptive optics SLO which brings the dynamism of adaptive optics to achieve unprecedented resolution. The other two are color fundus photography which provides rich vibrant pictures of which all retinal structures are captured and hyperspectral imaging which goes beyond the traditional imaging by capturing Through these modalities, the doctors have much better understanding of the disease process and visualize an entirely a different entity. Here, we will focus specifically on these techniques, namely their mechanisms, the way they are used, and more importantly, their imprints on ocular diagnostics.

2.2.7.1 Fundus Autofluorescence

Fundus Autofluorescence (FAF) stands out as a sophisticated retinal imaging technique that facilitates a vivid visualization of the intricacies embedded within the retinal pigment epithelium and the photoreceptor layer, a phenomenon meticulously documented by [42].

At its core, FAF employs the principles of fluorescence; by subjecting certain molecules to specific wavelengths of light, they are excited and, in turn, emit a luminescent glow. The brilliance of FAF lies in its non-invasive nature. While some imaging modalities require the use of exogenous agents, FAF exploits the autofluorescent nature of intrinsic retinal molecules. Therefore, the procedure does not call for the injection of any foreign contrasting or illuminative agents. Throughout the years and after numerous empirical studies, FAF has not only established prominence in generating high-definition retinal images but has also deepened the knowledge of a plethora of pathophysiological pathways responsible for various retinal diseases. It has thus become an efficient tool in the detection of biomarkers, making early detections possible and providing the basis for comprehensive management. However, FAF extends beyond the purview of common imaging, and its uses are rare, but highly specialized. These are for the diagnosis of toxic retinopathies, Age-related Macular Degeneration, as well as a multitude of retinal tumors, and several other ocular abnormalities.

2.2.7.2 Adaptive optic scanning laser ophthalmoscopy (AO-SLO)

An innovative AO-SLO imaging technique, which lies at the heart of this technique, constitutes a breakthrough in the field of ophthalmological imaging that is primarily given its unrivaled AO subsystem. Rubbing the technical heart of this SVO system is a highly advanced liquid crystal spatial light modulator, which plays the central role in the operating of the device. The plot of this cutting-edge modulator is the instigator of the system which is leadingly successful in neutralizing ocular optics aberrations. AO-SLO efficiently prevents such irregularities on images background, offering imagery with stunningly clear visuals and detail, unmatched by any other resource thus far. In terms of its light sources, the system adopts a two-pronged approach: a 780 nm diode laser serves as the navigation tool for wavefront sensing, but the illuminating SLD is the one with the 840 nm wavelength. What, beyond technical features, makes AO-SLO, more than any other imaging device, is the application of a customized software to take control and refine the imaging process. Thus, ensures that not the only exceptionally detailed are the resulted images but as well the resolution. Strikingly, the AO-SLO exam's impressive scanning property is one of its biggest strengths [43,44]. Different from other modalities during that time, AO-SLO displays the

ability of viewing wider spots of the retina, thereby, this modality results in large images that provide insight to entirety retinal health for clinicians.

2.2.7.3 Fundus fluorescein angiography

Fundus Fluorescein Angiography (FFA) stands as an integral pillar in the retinal imaging spectrum, possessing an established legacy in the arena of diagnosing an array of retinal afflictions. Specifically, its prowess has been recognized in discerning conditions such as Age-related Macular Degeneration (AMD) and Diabetic Retinopathy (DR). The inherent high-resolution capabilities of FFA empower clinicians with the ability to identify and monitor minuscule lesions, exemplified by microaneurysms (MAs), in their nascent stages. Detecting these anomalies at such an early juncture is pivotal, as it opens the avenue for timely interventions, thus potentially mitigating the progression towards more severe, vision-jeopardizing complications [45]. However, like any medical modality, FFA does not come without its set of challenges. Foremost among these is its invasive character, which, although rare, carries with it a slight risk of inducing an anaphylactic reaction in the patient. This element of invasiveness, combined with certain operational intricacies and constraints, can sometimes result in procedural delays. Such postponements in FFA administration may inadvertently cascade into subsequent delays in treatment initiation, potentially casting a shadow on the visual prognosis of the patient.

2.2.7.4 Indocyanine green angiography (ICGA)

Indocyanine Green Angiography (ICGA) is a sophisticated retinal imaging technique that primarily employs the indocyanine green contrast dye. Upon administration, this dye is systematically introduced into the circulatory system of the eye. Subsequent to this injection, the ICGA procedure is executed, harnessing a specialized laser light source paired with a high-resolution charge-coupled camera. This synergy enables ICGA to offer invaluable insights, encapsulating real-time assessments of blood flow, or perfusion, within the retinal tissues. A hallmark of ICGA lies in its exceptional resolution capabilities coupled with heightened diagnostic sensitivities, allowing for a nuanced visualization of even minute vascular anomalies. However, a prominent caveat associated with ICGA is its intravenous nature, which, though effective, casts concerns regarding the patient's safety profile. The introduction of an external agent into the body invariably comes with an array of potential risks and complications. Consequently, despite the potential diagnostic prowess of ICGA,

its application in the clinical setting witnesses certain reservations. Medical professionals, gravitating towards a more patient-centric approach, often express a preference for modalities that offer diagnostic insights without breaching the non-invasive paradigm [46].

2.2.7.5 Limitations and evolution of fundus imaging

Fundus imaging, while transformative in the realm of ophthalmology, comes with its set of intrinsic challenges. The anatomical nuances of the human eye, specifically the relatively diminutive size of the pupil and the constricted diameter of the iris, pose a significant constraint. These attributes can hinder the efficient passage of externally illuminated light directed towards the retina as well as the subsequent retina-reflected light trying to find its way out. In the design of fundus imaging equipment, it's paramount to ensure these light pathways remain distinct to prevent the obliteration of vital image contrast. Such intricate requirements, ensuring the separation and proper management of these light pathways, augment the technical complexity of fundus imaging systems. Consequently, this elevates the cost of the equipment and necessitates its operation by individuals steeped in considerable expertise and experience [41].

Therefore, it is important to praise the progress which fundus imaging has made within the last years by becoming available to a wider range of healthcare professionals. The most salient milestone was the shift from analogue photography to the digital universe. Moreover, such an approach does not just bring about new level of clarity and accessibility of the images but also stimulates the user orientation. Through adoption of digital techniques and the use of standardized imaging protocols, the fundus imaging process has turned from being daunting to one in which even non- eye care personnel can participate. Still, along with the advancements in fundoscopy technologies, large-scale population-based screening for vital ocular diseases such as diabetic retinopathy (DR), glaucoma, and age-related macular degeneration (AMD) remains its main application [41].

2.2.7.6 Optical Coherence Tomography (OCT)

Optical Coherence Tomography (OCT) stands out as a revolutionary, non-invasive imaging methodology, tailor-made for capturing detailed three-dimensional volumetric depictions of the intricate retinal structure [47]. This modality harnesses the principles of light waves, allowing it to meticulously visualize and document the cross-sectional imagery of the retina.

A salient advantage of using OCT is its capacity to delineate each distinct layer of the retina, empowering clinicians with the capability to precisely gauge their respective thicknesses. The underlying mechanism of OCT lies in its ability to estimate the origin depth of backscatter by calculating its light travel time. It's fascinating to note that the light's travel time is protracted for backscatters originating from deeper tissue layers compared to their shallower counterparts. Given the incredibly minuscule differences in these travel times, especially considering the retina's slender thickness range of 300–500 μm, OCT relies on interferometry as its trusted tool for flight time measurements [31]. Over the years, researchers and scientists have fine-tuned the OCT technique, giving rise to three primary methods to accomplish the coveted A-scan for the requisite depth range within the tissue:

2.2.7.7 Swept Source Encoded Frequency OCT

This variant of OCT refrains from maneuvering the reference arm. Instead, it rapidly modulates the light source around a central wavelength. For each central wavelength, a photo center comes into play, aiming to ascertain the correlogram. The depth positioning of tissue scatterings at the imaged location is then decoded by employing the Fourier transform technique.

2.2.7.8 Time domain OCT

Time Domain OCT offers a unique approach by mechanically transporting the reference arm to distinct spatial points. Each such relocation results in varying light travel time delays attributed to the reference arm. However, an inherent challenge associated with Time Domain OCT is the mechanical constraints of the arm movements, which restrict the number of A-scans to a few thousand every second.

2.2.7.9 Domain OCT

Spectrum Domain OCT exhibits operational parallels with the Swept Source OCT. The fundamental distinction lies in the choice of light source; the Spectrum Domain OCT utilizes a broadband spectrum. The modality employs diffraction combined with a CMOS linear sensor to carry out spectral decomposition of the interferogram. Yet, in sync with its OCT counterparts, the Fourier transform remains an instrumental technique to pinpoint the exact depth origin of each scattering signal.

2.3 ARTIFICIAL INTELLIGENCE

Artificial Intelligence (AI), rooted in computer science, aims to emulate the cognitive faculties commonly associated with humans and animals. Established as an academic discipline in 1955, AI faced periods of stagnation, often referred to as "AI winters". However, recent years have seen a resurgence in its popularity and application, primarily attributed to the convergence of significant computational capacities, abundant data availability, and advancements in theoretical frameworks that strive to mimic human cognition. AI spans a broad spectrum of subdomains: knowledge representation ensures that machines comprehend information in a structured manner, while reasoning enables the computational deduction of information from these representations. The natural language processing makes machines capable of speech comprehension as well as understanding, interpretation and generation of human language. Machine perception space deals with mirroring machines' sensory interpretation to man-like one. Here we are talking about computer vision for visual scenes understanding, machine hearing for audio data, as well as machine touch for tactile feelings. Besides AI, we have the staple of Machine Learning (ML) which is one of the most celebrated creation. The majority of ML is centered around teaching computers how to spot patterns within huge data collections by using a number of algorithms and statistical approaches. As a result, computers can solve problems they've never been instructed to without explicit programming of the task.

2.4 MACHINE LEARNING

Machine learning (ML) is a successful technology of Artificial Intelligence (AI) to give machines the ability to learn and make correct decisions from data and does not require the imposition of explicit rules of programming. The core idea of machine learning is in its algorithm's capability to identify distinctive patterns in enormous data volumes and then to provide predictions or inferences on the collected information. Classic ML algorithms consist of items including trees, likes K-nearest Neighbors, Naive Bayes, neural networks and advanced kernel techniques. For example, SVMs (support vector machines) fit into this category. The algorithms in question are aimed at continuously adjusting themselves and refining themselves by means of data primed knowledge.

ML algorithms differs from one to another, nevertheless, the majority of the algorithms are consistent in that they tend to learn from a given experience (E) and their performance is

typically evaluated by the specified metrics (P). In the words of an expert it is evident that the more the algorithm will have performed at the beginning the faster will be its learning process (X) thanks to experience (E) through which its effectiveness (Y) is assessed by accuracy or the relative error, and so on. In the area of different tasks, classification, regression, and anomaly detection could be indicated stil performed.

One pivotal aspect of ML is the nature of the learning experience it draws from: supervised, unsupervised, or reinforcement. Supervised learning capitalizes on labeled datasets, guiding the algorithm towards desired outcomes. In contrast, unsupervised learning delves into unlabeled data, uncovering hidden structures and relationships. Reinforcement learning adopts a more exploratory approach, wherein the model interacts with its environment, receiving feedback, and adjusting its strategies to optimize outcomes.

2.5 DEEP LEARNING

Deep Learning, nestled within the vast domain of Machine Learning, predominantly operates through structures known as Deep Neural Networks (DNN). Illustratively, Figure 2.4 elucidates the intertwined relationships between Artificial Intelligence, Machine Learning, and the more specialized Deep Learning and Transfer Learning. Rooted in the foundational concepts of the 1940s and 1950s, Deep Neural Networks draw inspiration from our very own biological neural configurations found within the brain [78].

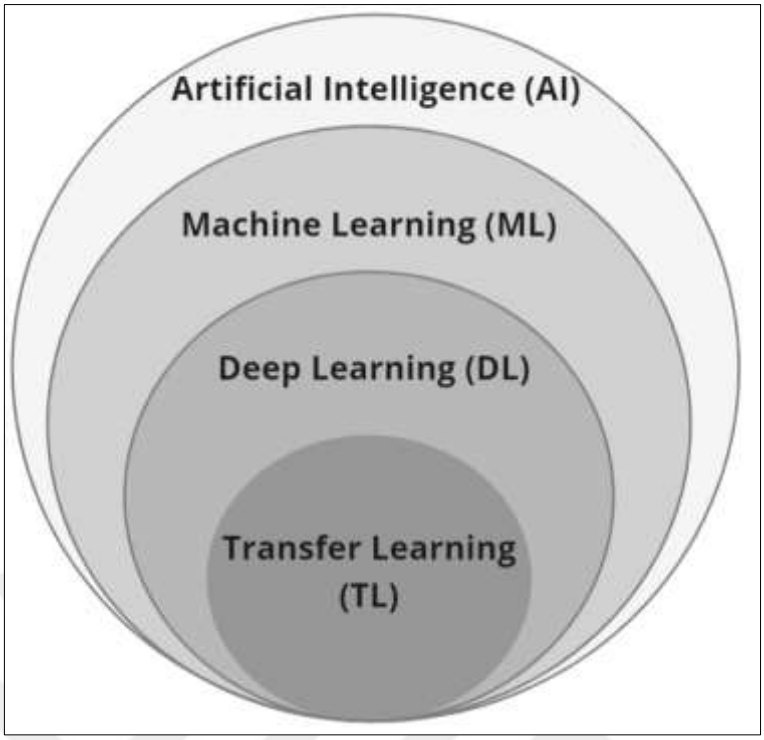

Figure 2.4: Relationships Between AI, ML, DL and TL**.**

Venturing into the anatomy of a standard neural network, it primarily features three layers: the input layer, a series of hidden layers, and the concluding output layer. The input layer serves as the initial receptor, channeling data into the network. This data is then intricately processed through the hidden layers, which engage in complex, non-linear transformations derived from the preceding layers, be it input or other hidden layers. Each of these stratified layers is populated with entities called neurons. Intriguingly, these neurons mirror the structural nuances of the biological neurons nestled within the human brain. Serving as the fundamental building blocks of neural networks, the specifics of a neuron are further visualized in Figure 2.5.

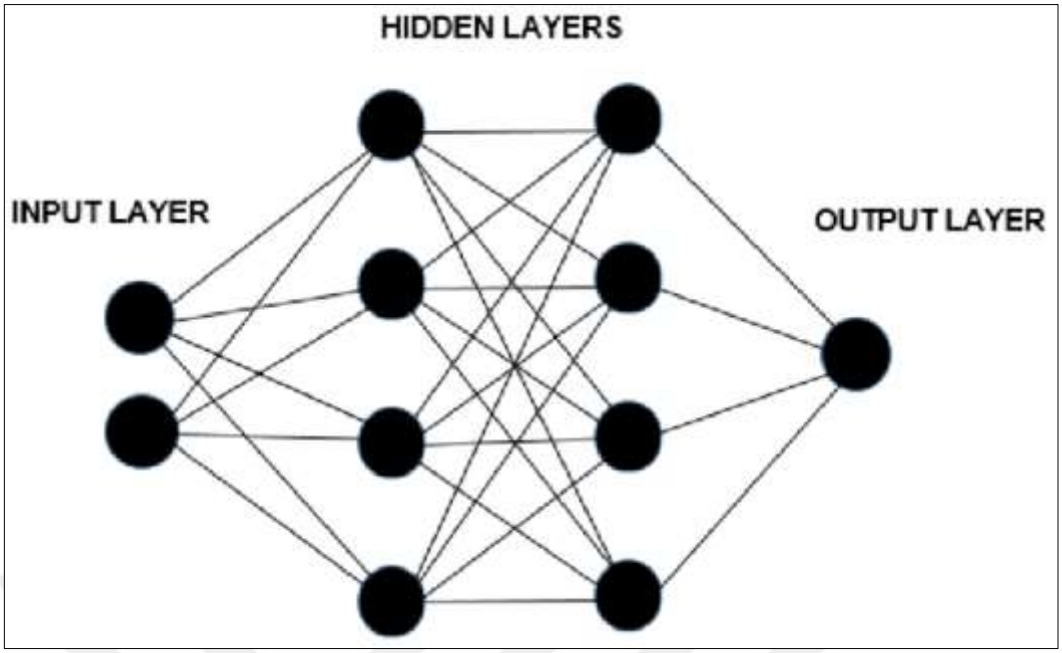

Figure 2.5: The basic Structure of a Neural Network. [77].

Neural networks, often conceptualized as the backbone of modern deep learning, are intricate computational models inspired by the intricate web of neurons in the human brain. These systems are designed to recognize patterns, assimilate data, and, most importantly, make decisions, just as humans do but at unparalleled speeds and efficiency.

A basic neural network comprises three primary layers: the input layer, one or more hidden layers, and the output layer. The input layer receives raw data for the network to process, much like our senses gather information from the environment. The hidden layers are where the magic happens; they process this information through a series of weighted connections and activation functions, allowing the network to "learn" from data and make sophisticated inferences. The output layer then provides the final result or decision based on the processed information.

Each node or neuron within these layers can be visualized as a mini decision-making box. It takes multiple inputs, processes them (often weighted by importance), and produces a single output. The activation function within each neuron determines whether and to what extent that information proceeds to the next layer, much like a gatekeeper.
Neural network training is basically a process of feeding lots of diverse data to it and calibrating the connections between neurons in terms of their weight (importance) to

minimize the difference between the forecasted outcomes and exact results. This process that is perpetual and is based on algorithms such as backpropagation and which is called gradient descent to errors progressively shrinks the network's volatility. As it learns, the neural network becomes so accurate that it can even process unseen data as a result and is therefore a potent tool for tasks such as image recognition, language processing, and even diagnosis.

2.5.1 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) shown in figure 2.6 were firstly developed by LeCun et al. in 1989, to extract patterns from handwritten digit images. In this paper, convolutional neural networks were greatly engaged in the classification and analysis of the diabetic retinopathy images. Originally developed for the processing of grid data in the form of images, CNNs emerged as extraordinarily task-oriented systems which readily spot spatial patterns or structures. Such networks, however, are hierarchical in their workings starting from the first layer where simple and local patterns are created to subsequently combining different units with one another to report complex and global patterns in the deeper layers. We implement this structured technique to sift and transform the important features in an image into the shape of others; we do so that the image can either be used to classify or regress others.

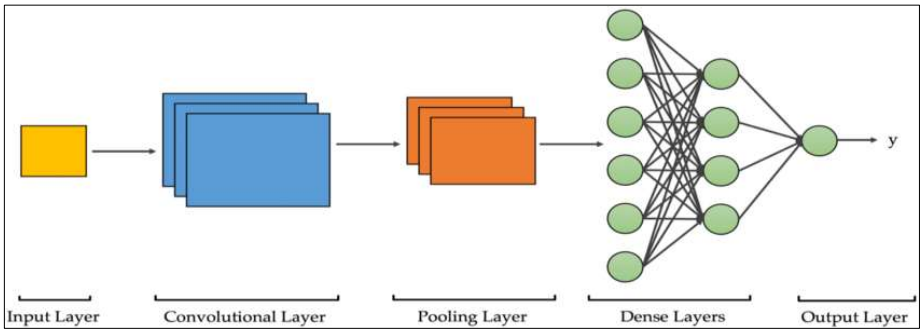

Figure 2.6: Convolutional Neural Networks [79].

Similar to standard neural networks, CNNs are made up of a series of layers. In each layer a convolutional operation converts an input tensor which is a multi-dimensional array into an output tensor. This output after that is a candidate for additional transformations within the same layer, including batch normalization, activation functions or dimensionality-reducing

operations like max-pooling or average pooling. First of all, theses transformations have parameters that can be trained and fine-tuned during training to increase the accuracy of the prediction. The main component that makes the CNNs stand out and gives them their special abilities is the convolution operation.

a. Convolution Operation: The convolution operation, as depicted in figure 2.7, is methodically executed across various locations of the input data. This operation is characterized by several parameters: a tensor that accounts for the number of input features, convolution width, convolution height, and number of output values; as well as both horizontal and vertical padding and stride values.

Padding refers to the supplementary pixels appended to the input's width and height before the convolution operation is applied. For instance, if the original size of the input is defined by its width and height, then after padding, the adjusted size for the convolution operation becomes (width + 2 x padding width) by (height + 2 x padding height).

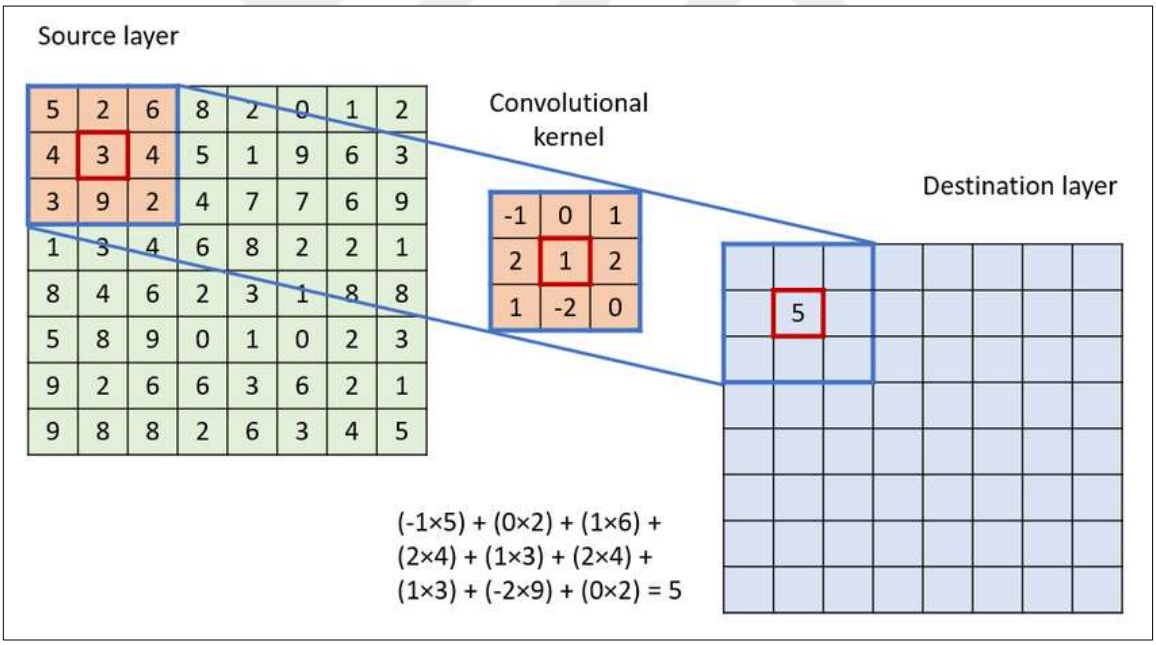

Figure 2.7: Convolution Operation [80].

The convolution process is initiated at various points on the input, and the starting points for each application are separated by a set distance defined by the horizontal and vertical strides. The resulting output value from each convolution operation is derived from the linear amalgamation of the convolution parameters with the values from the input.

b. Pooling Operation: Pooling operations, as illustrated in figure. 2.8, are fundamental components frequently employed in CNNs, primarily to downsize feature maps. Max-pooling specifically chooses the highest value within a designated window, ignoring the rest. Conversely, average-pooling computes the average of the values within the window. Both methods are widely utilized techniques for diminishing the dimensions of feature maps.

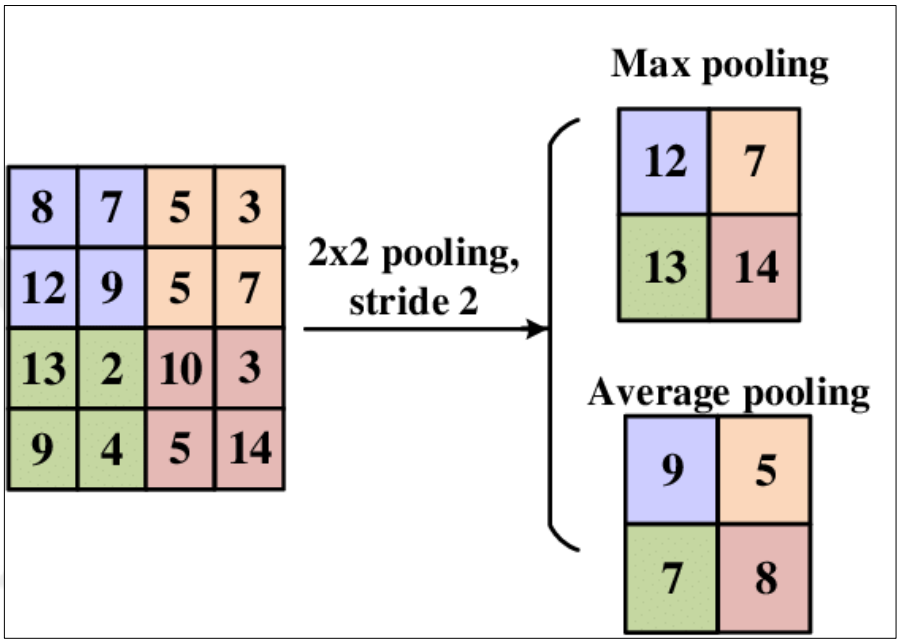

Figure 2.8: Pooling operations [82].

c. Batch normalization: Batch normalization is a commonly utilized component within CNN layers. This technique normalizes the layer inputs, aiding in mitigating covariate shift. This not only enhances the model's stability but also serves as a form of regularization [81].

d. Activation functions: Activation functions, as depicted in figure. 2.9, play a pivotal role in the functionality of neural networks. Among them, ReLU stands out as the most versatile and commonly used. Historically, the sigmoid function dominated the 1980s and is currently employed primarily in specific output layers, largely due to challenges associated with its gradient descent vanishing. Absent these activation functions, deep learning models would essentially act as linear functions.

In the past, the sigmoid function was favored for its smooth, continuously differentiable nature. However, it presents a significant drawback known as the gradient vanishing problem [84], whereby its first derivative flattens out not far from its origin. This characteristic hinders network loss optimization due to nearly zero gradients. In a landmark study [85], ReLUs were introduced as an enhancement to Restricted Boltzmann Machines (RBMs), serving as approximations to stepped sigmoid units. Further research [86] comparing the performance of Sigmoid, Tanh, and ReLU found that, while Sigmoid might have closer biological parallels, Tanh and ReLU were more effective in training multi-layer perceptrons.

ReLU networks generally outperform others, even with their inherent non-differentiability at zero. Their capacity to lead to sparse representations offers dual advantages: a robust information representation and significant computational efficiency. Moreover, ReLU's simple function and derivative expedite calculations, a crucial advantage for large networks. Its consistent gradient value also bypasses the gradient vanishing problem, paving the way for deeper network designs. Since these discoveries, ReLU has been cemented as the go-to activation function in deep learning. While other activation functions, such as LeakyReLU, have been introduced over time, they haven't brought substantial performance improvements.

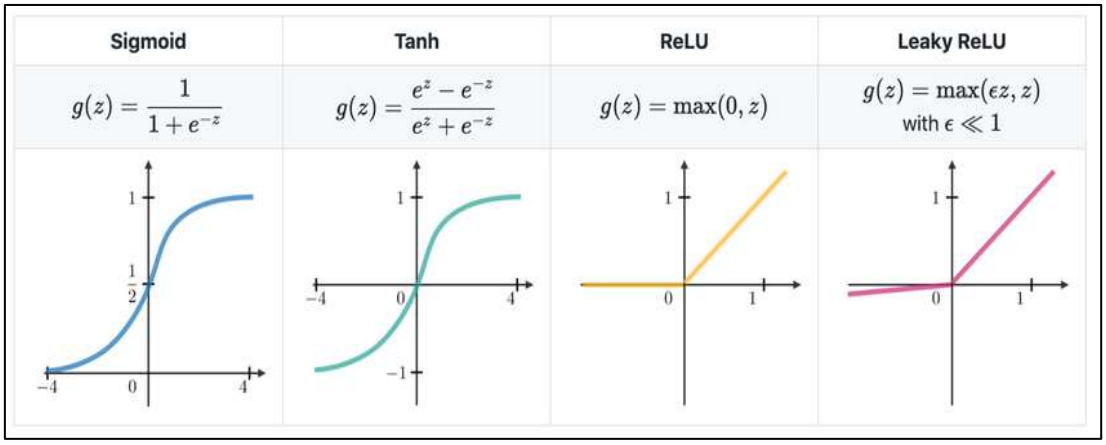

Figure 2.9: Typical Activation Functions [83].

2.6 TRANSFER LEARNING

Transfer learning is a modern method in the ML and DL field that concentrates on reusing knowledge that you acquired first and utilizing it to a problem that is similar but different than the first problem. As compared to starting with a blank slate, the data patterns that have previously been learned can be picked up by the transfer learning model, leading to faster training and better results, particularly when it comes to datasets of limited size.

The core essence of transfer learning is based on the fact that the features acquired by models during one task can have a role of a basic starting point when working on more complex tasks. DL also brings this illustration to live in particular cases like solving image recognition problems with CNNs. This example demonstrated that some of the functions in early layers in models pretrained in large datasets like ImageNet, such as scene understanding and texture perception, are universal and are present across species. These pretrained models are transferred, although the task involved maybe different categories, notch changes are often observed around the earlier layers, if at all. Precisely, the deeper neurons of the network are dedicated to the special sensory functions

Additionally, this one rule shows how transfer learning can take an advantage of a small amount of labeled data. Generally, ML chooses deep neural networks, and these networks need a large amount of data in order to run effectively without getting stuck in overfitting during the process of training. Using supervised learning methods, though, in the case of data collection, there is often the issue of acquiring labeled data which may be hard, lengthy, or too costly. With the help of transfer learning, the model can cross this gap by simplyly using the existing knowledge thereby, data won't be a must in vast amount.

Harnessing pre-trained models as a starting point, transfer learning ensures faster convergence and enhanced model generalization. Among the myriad of architectures available, some have stood out for their effectiveness and adaptability in a plethora of applications. Notably, models such as MobileNet, ResNet50, InceptionV3, and DenseNet have become the bedrock of many advanced deep learning solutions. Each of these architectures, with their unique design principles and strengths, offer a rich tapestry of features that can be adapted and fine-tuned for various tasks. As we delve deeper, we will

explore the nuances, capabilities, and specific use cases where these models shine, and discuss their potential when leveraged via transfer learning

2.6.1 MobileNetV2

In the ever-evolving landscape of deep learning, the significance of efficient and lightweight architectures has grown, especially given the increasing computational demands of most conventional models. Within this context, MobileNetV2 emerges as a beacon of efficiency, particularly designed for mobile devices, making it an ideal choice for this thesis. While popular convolutional neural networks (CNNs) like DenseNets and VGG exhibit impressive performance, their extensive computational requirements make them less suitable for resource-constrained applications. MobileNetV2, building on the legacy of its precursor MobileNetV1, introduces advancements that make it both more efficient and powerful.

A distinguishing feature of MobileNetV1 is its utilization of depthwise separable convolutions. Instead of the traditional convolution process, it cleverly divides the convolution layer into two sequential operations. Initially, a depthwise convolution layer filters the input, which is subsequently processed by a 1x1 pointwise convolution to amalgamate these filtered inputs, producing new features. These two layers, in tandem, constitute what is termed as the 'depthwise separable convolution block'. This innovative architecture achieves the equivalent of a standard CNN's functionality but operates at nearly nine times the speed, while maintaining comparable accuracy [88].

MobileNetV1 is further enhanced by incorporating batch normalization and deploying the ReLU6 activation function an upgrade over the standard ReLU, known for its superior performance. The network concludes with a global average pooling layer, succeeded by either a fully connected layer or a 1x1 convolution, and culminates in a softmax. An intriguing aspect of the MobileNet architecture is the depth multiplier, often referred to as the width multiplier. This hyperparameter allows for the tuning of the number of channels in each layer, providing a mechanism to tailor the network's complexity based on specific needs [87].

2.6.2 ResNet50

Deep learning is particularly benefiting from the numerous innovations that have occurred during the thriving period of this technique with an outstanding example being ResNet50, which is a variant of ResNet architecture. ResNet, which is noteworthy paper by He et al., 2015 demonstrated the way of thinking that was coupled with deep neural networks' structure. Before its adoption, it was held by the people that whenever network is deep, then it has better performance. But nevertheless as networks went much deeper into learning stage there challenge increased to train them due to the factor of the disappearing and exploding gradients which sometimes resulting into saturation or degradation.

ResNet, being the first network to adopt "skip connections", also known as "residual connections", creates a path that does not include one (or several) layers during the training phase. Such linkages form upturns in the network and make it possible to bypass specific layers allowing the gradient to move straight through to the output layer where it declines directly without vanishing. The word "residual" comes from the structure (this is the link where it learns the residual mapping, which will then will update stack of layers). Apparently, this intends to shift the responsibility of the layers stacked together to learn the essential underlying mapping, as opposed, to their work on residual mapping which is always easy.

ResNet50 is a sequential variant of the ResNet model with 50 layers. It is composed of a range of convolutional layers with each one followed by batch normalization and then a ReLU activation. Structurally, the model is composed of several stages with convolutional layers increasing in depth while decreasing the spatial dimension followed by residual connections . The model’s final layers include a global average pooling layer followed by fully connected layer with a softmax activation bridging the links with input classifications. The depth and residual connectivity are the strengths of ResNet50. The depth makes the model capable of learning extensive and intricate patterns in input data output. The model, under this scenario performed excellently in image classification tasks due to its exceptional acquired power. On the other hand, residual switching ensures that ResNet50 is trainable even with its depth. These two aspects have made ResNet50 an enabling model to achieve specific state of the art performance bounds on several benchmarks.

2.6.3 InceptionV3

Inception modules are responsible for the variation in the depth and width network's topology. InceptionV3 is a variant of the Inception architecture and perhaps one of the greatest masterpieces in DL for a computer vision context. The original Inception architecture, or "GoogLeNet" as it was often called in honor of its creators at Google, was developed to increase the depth and width of the network, while keeping computational costs efficient. It was achieved using a new type of module-inceptions. The Inception module allowed the network to select the optimal size of convolutional filters within it. The innovation of the Inception module lies in the fact that while the previous layers of the network were forced to choose convolutions with only a single kernel size, this module performs convolutions with many different kernel sizes at the same time. For example, it can conduct 1x1, 3x3, and 5x5 convolution, and pooling operation, in one layer. Therefore, it can capture spatial hierarchies between features at different scales. After that, the results of those are concatenated and provided as an input to the next layer. Due to this structure, the network can learn multilayered representation of features out of the data without the necessity of an exponential increase in costs.

InceptionV3 can be seen as a refined form of this architecture building upon the ideas sketched by its predecessors. InceptionV3 has multiple refinements to its predecessors for better performance and efficient training. One of the major innovations in InceptionV3 is the use of factorized convolutions. Instead of directly using larger convolutional filters like 5x5, InceptionV3 factorized them into a combination of two smaller convolutions, which is 3x3. Factorization will reduce the number of parameters and computations, making it more efficient. Another interesting development is 'label smoothening' to improve model regularization. It makes the network slightly less confident about its predictions making the ground truth less likely to be in the right classification. It is counter-intuitive because a less confident network can often generalize better on unseen data. Finally, the use of auxiliary classification that is placed in the middle of the network also appears to help the gradient to propagate during training, addressing the vanishing gradient problem with such a deep network. Structurally, InceptionV3 has 48 layers with auxiliary classification and batch normalization excluded. The significant design and computational efficiency make InceptionV3 a preferred option for a wide range of computer vision tasks. It often performs

VGGs or close in the latest benchmarks. The depth and width with innovative modules make InceptionV3 a perfect demonstration of the field and advancement.

2.6.4 DenseNet

DenseNet, which stands for Densely Connected Convolutional Networks, has fundamentally altered the design and concept of convolutional neural networks by abandoning the conventional sequential arrangement of layers. Although numerous versions have been developed since then, one of the most popular is the DenseNet121 model; it encapsulates the principles of the DenseNet method while retaining moderate depth and versatility of usage cases. The most crucial and self-explanatory feature of DenseNet networks is the dense connection of the layers. In conventional networks, each layer gets data solely from the previous one; conversely, each layer in the DenseNet gets extra input from all layers that came before it in the stack and share its data with all following layers. Thus, the nth layer accepts the feature maps of all n-1 preceding layers. This architectural solution enables the most effective information flow through the network.

Having such dense connectivity has several benefits. Firstly, it promotes feature reuse across the network, which is extremely useful in cases when multi-scale feature representations are required. Secondly, it leads to a very small number of parameters and computations since each layer has access to the feature maps of all preceding layers, which means it has to learn fewer filters. In this way, it forms if a solution to the vanishing-gradient problem, making training very deep networks feasible and stable. The other two key components of DenseNet121 are that it behaves as the bottleneck layer, which helps keep the number of features constant and makes the model computation efficient. The other critical part is the use of a transition layer that consists of a convolution and pooling operation to change the size of feature maps. Nearly all of these properties can be actualized in the other DenseNet variations. Another significant advantage of DenseNet is its inherent capacity for incorporating multi-scale data. Since each layer obtains features from all previous layers, it has access to information that is both detailed and abstract. It is particularly useful for tasks like segmentation or object detection. Lastly, due to the aforementioned properties of the DenseNet architecture and particularly the DenseNet121 variant, the model is very regularized. Thus, it generalizes better and is less prone to overfitting, even if the dataset and data augmentation scheme are relatively small.

2.7 LITERATURE REVIEW

Retinal eye diseases are one of the major leading causes of vision loss and blindness across the globe. Although a high percentage of visual loss can be prevented through early detection and treatment, the existence of deep learning methods has introduced new ways of fast and accurate retinal disease identification. Consequently, the subsequent studies elaborate on related works of deep learning areas that use various techniques in identifying and analyzing the retinal eye disease and how they advanced over the years.

2.7.1 Fundus Image Databases

Fundus image databases form the cornerstone of enhancing and revolutionizing research and practices in ophthalmology. With the growing impact of the digital era on medical imaging, the need for datasets of high quality maintained and made available to researchers and practitioners has significantly increased. These datasets fall under two main categories: private and public databases. Public fundus image databases are open-access and availed to the larger research community, while private databases are maintained by institutions for internal research purposes. Fundus image databases contain a broad range of retinal images presenting different conditions and pathologies. The following literature review examines fundus image databases and their contribution in the field.

2.7.2 Public Database

In the constantly evolving field of medical imaging, particularly retinal imaging, there is a growing demand to develop and fine-tune algorithms that can accurately interpret and analyze images. To meet this requirement, researchers often need extensive data to validate or train their models. Recognizing this pressing demand, several research groups have taken the initiative to compile and develop their own comprehensive databases. To foster community collaboration and further advancement in the field, many of these databases have been made publicly accessible. Among these, the DRIVE and STARE databases stand out prominently. These open-access repositories are frequently cited and utilized by researchers and professionals worldwide. One of the primary reasons for their widespread adoption is the exceptionally high resolution of their fundus images, which offers detailed insights into the retina's structure and any present abnormalities [48,49].

Table 2.1 provides a comprehensive overview of various public databases that have been established to aid in the research and development of retinal imaging techniques and analysis. These databases, made accessible to the public by various research groups and institutions, contain invaluable data that can be utilized for model validation, training, and comparative studies. The compilation ensures that researchers have a diverse set of images and cases to work with, further advancing the field of retinal studies.

Table 2.1: Examples of Public Databases**.**

Dataset Name	**Description**
DRIVE	A database created for blood vessel segmentation studies consisting of 40 retinal photos from 400 human subjects aged between 25 to 90. It contains both DR and non-DR images [50].
STARE	A database with 400 retinal images with 40 having manually segmented blood vessels and artery/vein labeling. It also includes algorithms for optic nerve head (ONH) identification [51].
ARIA	A JPEG format database containing 450 images, categorized into healthy, AMD, and DR groups [50].
CHASEDB	Contains 28 manually segmented monochrome ground-truth images to establish the link between ocular vessel tortuosity and cardiovascular disease risk factors [48] [52].
IMAGERET	Consists of two parts: DIARETBD0 and DIARETDB1 saved in PNG format, with images showing varying degrees of DR [53].
MESSIDOR	A TIFF format database created for evaluation and comparison of segmentation algorithms for retinal lesions. Contains 1200 color fundus images labeled for DR grade [54].
MESSIDOR-2	Contains 1200 high-quality images with varying classifications according to the ICDR scale [55].
e-Ophtha	Designed for DR screening and manually annotated by specialists. Contains two subsets for microneurisms and exudates [56].
DIARET DB1	An 89-color fundus image standard database used for benchmarking DR detection algorithms [57].
APTOS	A Kaggle dataset with 3662 images of varying sizes captured with different cameras 58].
AREDS	A longitudinal study database following up patients' AMD conditions [59].
ORIGA	Contains 650 retinal images, with sections of the images depicting glaucoma and others depicting healthy retinas [60] [61].
ACRIMA	Contains 705 images captured using a Topcon TRC retinal camera, with images representing glaucomatous and normal conditions [62].
RIM-ONE	Comprises 159 fundus images, with sections of the images representing glaucoma and others depicting healthy conditions [63].

Table 2.1: Examples of Public Databases "Table Continoued".

LAG DB	A large database with 11,760 images from 10,147 subjects, with images depicting glaucoma and normal conditions [64].
OCT2017	A high-quality TIFF-format public database for detecting retinal diseases in a multi-class problem [65].
SERI DB	Contains 32-volume spectral domain OCT images for DME classification [66].
ODIR Dataset	Contains multilabel annotations for eight retinal diseases from 5000 patients [67].
OIA-ODIR	Contains 5000 images captured by different cameras from different regions of China [68].
ROC	Has three different image types with varying resolutions taken by various camera systems in different settings [69].
IOSTAR	Contains 30 retinal photos taken by a laser fundus camera and annotated by specialists [70].
REVIEW	A compound directory of four datasets designed for blood vessel segmentation model evaluation [48].
DR HAGIS Database	Created as part of a DR screening campaign in the UK with varying resolutions [71].
VAMPIRE	Captured by a fundus camera and software designed for the identification of retinal vessels.
KAGGLE Database	Provided by EyePACS, contains 88,702 high-quality images taken under different conditions [72].
RET-TORT	Contains 60 retinal images from patients with hypertension and healthy patients, providing information about their estimated tortuosity [73].

2.7.3 Private Database

Within the realm of retinal disease detection, many researchers have turned to private databases to assess the efficacy of their algorithms. Given the sensitive nature of medical data, these databases prioritize the confidentiality and privacy of the subjects. Thus, However, these images are subjected to an extensive anonymization process before they can be used to develop or evaluate models to ensure that there are no personal identifiers in such images. Since these databases are inherently private, some datasets can be accessed by an outsider researcher once they obtain special permission to use the images in these databases. This permission is commonly granted by the original authors or the medical institutions that own these databases under strict conditions, including the signing of a research agreement

to not use the data outside the permitted use. This section goes deeper into some of the most notable private databases used for retinal disease research. Table 2.2 describes some of these private databases and their characteristics and potential uses. Going through these datasets enables one to understand the wide range of resources available in the field and how a variety of data is used to develop our understanding of retinal disease detection.

Table 2.2: Examples of Private Databases.

Dataset Name	**Description**
The RetCam3 Dataset	A private database stemming from a premature infant screening program, containing 80 images with a resolution of 640 × 480 pixels, captured using a RetCam3 camera [74].
SCES	Created from the Singapore Chinese Eye Study Ophthalmology (SCES) screening study, this database has 1676 images (1630 normal and 46 glaucomatous). The images have resolutions of either 3888 × 2592 or 3504 × 2336 pixels [75].
TROPIC	Features retinal images from 41 premature infants, totaling 130 images taken with a RetCam130 wide-angle camera at a resolution of 640 × 480 pixels. The dataset includes various gradations of retinopathy of prematurity (ROP) among its images [76].

2.7.4 Previous study of Diabetic Eye Disease

Researchers in this paper [89] proposed an automated classification system for Diabetic Retinopathy (DR), a prevalent eye condition linked with chronic diabetes. Their system evaluates fundus images, accommodating diverse illuminations and fields of view. They utilize machine learning models, notably CNN, VGG-16, and VGG-19, to assign a severity grade for DR. The proposed model demonstrates performance metrics of 80% sensitivity, 82% accuracy, 82% specificity, and an AUC of 0.904. It categorizes images into five grades, spanning from 0 (indicating no DR) to 4 (indicative of proliferative DR).

Researchers in this paper [90] introduced an automatic deep-learning technique for diabetic retinopathy stage detection using a single human fundus photograph. They also presented a multistage transfer learning approach that leverages datasets with variant labeling. Their

method, suitable for early diabetic retinopathy screening, boasts a sensitivity and specificity of 0.99. It achieved a notable ranking of 54 out of 2943 methods on the APTOS 2019 Blindness Detection Dataset, with a quadratic weighted kappa score of 0.925466.

Researchers in this paper [91] introduced a novel diabetic retinopathy monitoring model. The model first employs the Contrast Limited Adaptive Histogram Equalization technique for image quality enhancement during pre-processing. Subsequently, the EfficientNet-B5 architecture is adopted for classification, notable for its uniform scaling capabilities. The model, when trained on a combined dataset of Messidor-2 and IDRiD and evaluated on Messidor, improved the AUC from a previous best of 0.936 to 0.945. Furthermore, when trained on a combined dataset of Messidor-2 and Messidor and evaluated on IDRiD, the AUC advanced from a prior peak of 0.796 to 0.932. This model outperforms others in AUC metrics as per the comparison.

Researchers in this paper [92] introduced a convolutional neural network that leverages the VGG-16 model, pre-trained, for refined classification of Diabetic Retinopathy (DR) severity. The model incorporates advanced deep learning strategies such as data augmentation, batch normalization, dropout layers, and learn-rate scheduling on high-resolution images to optimize accuracy. The outcomes include an average class accuracy of 74%, a sensitivity of 80% with a specificity of 65%, and an AUC of 0.80. These results surpass performance metrics of prior studies utilizing other pre-trained networks or models. In [93], the authors proposed a deep learning algorithm based on the Densely Connected Convolutional Network – DenseNet-169 for early diabetic retinopathy detection. The method sorts fundus images according to the level of severity: from No DR to Proliferative DR. The datasets were taken from Diabetic Retinopathy Detection 2015 and Aptos 2019 Blindness Detection from Kaggle. Our research involved such steps: data collection process, preprocessing, augmentation, and modeling. Our findings proposed a model with 90% accuracy, and an additional regression model achieved 78% accuracy. Our overall aim was to create a powerful system for DR detection.

The authors in [94] proposed a deep learning model for assessing Diabetic Retinopathy's severity, which is a diabetes complication that causes vision loss. The deep learning model was based on various versions of the EfficientNet and used to develop a classifier for DR images to diagnose their absence and presence from their increasing severity to the

proliferative phase. The model was trained on diffeerent training and validation datasets, and while the most efficient one demonstrated a score of 0.924377 for the APTOS test dataset, the others achieved high kappa and low score, meaning that they are not suitable for a particular dataset. This success demonstrates that this model has the potential to be used for quick DR assessment and the early stage identification of this diminishign disease.

A new deep learning model is proposed by these researchers in this paper [95], called Hinge Attention Network. HA-Net employs multiple attention stages, using a pre-trained VGG16 base for initial spatial representation from retinal scans, a spatial attention autoencoder for lesion-specific latent representation, and a channel attention-based hinge neural network to classify the truth of retinopathy. HA-Net also contains a Convolutional LSTM layer to highlight significant spatial maps and strengthens predictions for new data. Using Kaggle APTOS 2019 and ISBI IDRiD benchmark datasets to validate HA-Net, HA-Net outperforms other models, achieving accuracies of 85.54% on Kaggle APTOS and 66.41% on IDRiD.
Researchers in this paper [96] introduced an automatic diabetic retinopathy detection methodology rooted in deep learning. The process comprises two key phases: an initial pre-processing step utilizing deformable registration on the retina to maximize image occupation and reduce background influence, followed by a classification step where four CNN models (DenseNet-121, Xception, Inception-v3, Resnet-50) are trained to determine the diabetic retinopathy stage. Given the APTOS 2019 dataset's limited size, a transfer learning approach was implemented, pre-training the CNNs on ImageNet and fine-tuning on APTOS. During testing, a voting system based on the four CNNs' outputs determines the final prediction. This model achieved an 85.28% accuracy in the testing phase.

Researchers in this paper [97] evaluated 28 hybrid models against seven deep learning architectures for diabetic retinopathy classification. These hybrids combined seven DL techniques with four classifiers. Tested on the APTOS, Kaggle DR, and Messidor-2 datasets, results highlighted the benefits of merging deep learning feature extraction with traditional machine learning classification. Notably, a hybrid of SVM and MobileNet_V2 recorded accuracy levels up to 88.80%, competitive with top deep learning models. Despite DenseNet201 and MobileNet_V2 standalone models outperforming hybrids, the SVM-MobileNet_V2 combination is preferred for its efficiency and simpler tuning.

Table 2.3: Related Works.

Ref	Methodology	Dataset	Results
Nguyen et al. 2020 [89]	Transfer learning of VGG16 and VGG19	Kaggle competition dataset 2015	Accuracies: 71% (VGG16), 73% (VGG19), Improved to 83% after modification
Tymchenko et al. 2020 [90]	Three-head CNN	APTOS 2019 Blindness Detection Dataset	Sensitivity: 99%
Pour et al. 2020 [91]	EfficientNet B5	MESSIDOR, MESSIDOR-2, IDRiD	AUC: 0.94 (MESSIDOR), 0.93 (IDRiD)
Thota and Reddy 2020 [92]	VGG-16	Kaggle EyePACS	Accuracy: 74%, Sensitivity: 80%, Specificity: 65%
Mushtaq and Siddiqui 2021 [93]	DenseNet-169	Diabetic Retinopathy Detection 2015, Aptos 2019 Blindness	Accuracy: 90%
Karki and Kulkarni 2021 [94]	Ensemble of EfficientNet models	Kaggle APTOS	EfficientNet-B3 performed better than the ensemble and other models
Shaik and Cherukuri 2022 [95]	Hinge Attention Network (HANet)	IDRid	Accuracy: 66.4%
Oulhadj et al. 2022 [96]	DenseNet-121, Xception, InceptionV3, ResNet-50	Kaggle APTOS	Best accuracy: 85.28%
Lahmar & Idri. 2023 [97]	SVM-MobileNet_V2	APTOS, Kaggle DR, and Messidor-2	Accuracy: 88.80%

3. METHODOLOGY

3.1 OVERVIEW

This chapter will thoroughly explain the architectural framework and designation of our proposed model. The complexity of retinal image analysis demands a model that is deep and accurate enough to give a comprehensive diagnosis. We, therefore, propose to modify the latest deep learning architectures with their inherent large learning power to appropriate our specific variations in the dataset. We will detail out each sub-component, and the rationale behind selecting it will be provided. These will include the main architecture, custom layers, and training parameters to mention a few. We will provide a clear and straightforward explanation to act as a framework for transparency and repeatability of our research process, thus establishing the innovativeness of the research.

3.1.1 Environment

Our proposed system was facilitated by using computational resources from Colab pro, a cloud-based platform from Google. Significantly, Colab pro provides a high-performance computing environment, creating a scalable infrastructure. The training and validation of deep learning models from large datasets were made possible through the provision of Colab Pro. Being able to integrate Colab Pro also enhanced our entire design; we were able to easily get GPU and TPU acceleration, which reduced the time wasted during the training and validation process. Moreover, using a shared work environment offered by Colab, all our research team worked concurrently; they were able to share code, work on the same project, which was more productive than when one person worked on it. To sum up, the Colab platform is important in our proposed deep learning model. It enabled collaborative effort in the development of real time.

Figure 3.1: Google Colab Logo.

3.1.2 The Technique of Transfer Learning

Transfer learning is a fundamental concept in deep learning and AI and plays a significant role in the development of highly competitive models for a wide range of tasks, ranging from image processing and classification to natural language to name a few. Transfer learning is a showcase of how efficient and effective learning from one domain or task can improve performance in another. At its core, transfer learning allows a neural network to benefit not only from the knowledge it has acquired throughout the training process on one dataset but also to use this knowledge on a different dataset related domain with limited or no labeled data. The guiding principle underpinning transfer learning is that a neural network can transfer learned information from a source domain to perform better in a target domain. This is practically achieved by using a pre-trained neural network, which has learned feature representations from a large dataset, as a basis for a new task. A pre-trained model explicitly learns to identify low-level features and more complicated patterns in its original domain. This often encompasses learning the semantics behind shapes and textures and object hierarchy among many other aspects.

Transfer learning can take several forms, with the most common paradigms being fine-tuning and feature extraction. In the first approach, the weights of the pre-trained base model are adjusted while the model is being trained on the new task's dataset. This enables the network to fine-tune its knowledge to the specifics of the target domain and dataset. Fine-tuning usually involves a smaller learning rate, allowing the weights to be adjusted slightly without erasing much of the features learned from the source domain. Feature extraction, in turn, fixes the weights of the pre-trained base model, using it as a fixed feature extractor. Instead of adjusting the weights, extra layers, called the head layers, are added on top of the base model, which are then trained on the new task's dataset only. Feature extraction helps the network concentrate on learning the task-related features instead of adjusting all of the weights, which can be suspended. The advantages of transfer learning are numerous. Due to having the ability to re-use the pre-trained components of the model's architecture, it can be implemented with less computational resources and annotated data, making it an affordable option for researchers and developers just starting. Additionally, it may have a faster convergence on the target domain's limited dataset and overall improve the performance of the model. In this thesis, we exploit transfer learning as one of the core methods for

developing the proposed retinal disease classification neural network model. By using the existing neural network architectures such as EfficientNet, VGG16, VGG19 and adapting them to our individual domain, we can easily train the neural network, and our model can correctly and efficiently classify retinal disease images, therefore contributing to early detect and treat those diseases.

3.2 PROPOSED METHODOLOGY DESIGN

The Retinal Optical Coherence Tomography dataset has been used in the proposed model. It is an extensive collection of retinal scans that enables one to study the ocular anatomical structures. The data of the HD-OCT images taken at high resolution acts as a center of gravity upon which the deep-learning architecture is built and validated. The raw data is subjected to a series of preprocessing steps to get it into a state where machine learning and deep learning models can use it to perform the desired tasks and deliver accurate results. The first step is resizing all the images in the folder to the same dimensions for the ease of the deep-learning frameworks to be applied. Since the color intensity of the pixels in an image file is a critical aspect of image recognition by machine learning and deep-learning models, even if the image files are loaded into the framework in grayscale, they would be converted to the RGB format to enable the frameworks to recognize even the most subtle variations. This also standardizes the input format for the deep-learning model. We also standardized the normalization such that the pixel values would fall within a certain specific range within which the values should be. This standardization is important as it helps the model to converge faster during the training phase and also avoid getting stuck in local optima. After preprocessing the data, the model is prepared to make predictions. The proposed model uses an ensemble for the training of the various state-of-the-art neural network architectures including MobileNetV2, ResNet50, InceptionV3, and DenseNet. Each of these is trained a ‘train set’ folder, and then using this models predication is carried out in the ‘test set’ folder. By leveraging the unique characteristics and strengths of the different neural network architectures on the OCT dataset, the proposed model hopes to push the frontiers on the analysis of retinal scan images and disease detection.

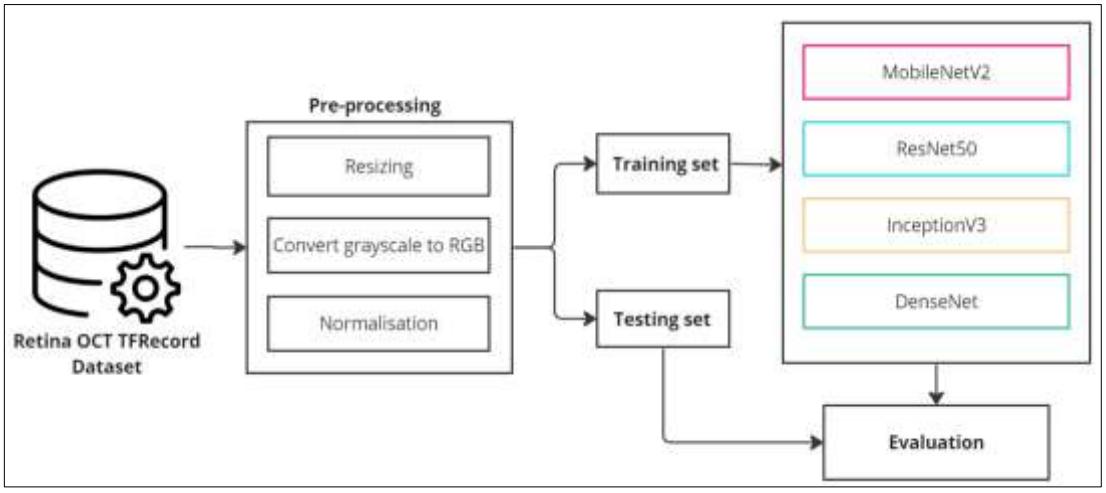

Figure 3.2: Proposed Retinal Eye Disease Model.

3.2.1 Dataset Description

Retinal Optical Coherence Tomography is a high-resolution technique for cross-sectional living patient retinas imaging, with more than 30 million scans delivered per year. This data set[1] brings together an extensive collection of classifications of the OCT images, from choroidal neovascularization, diabetic macular edema, drusen indicative of early age-related macular degeneration , and healthy retina. Specifically, it includes around 84,000 scans, and it is meticulously split into train, validation, test. Every image captured between 2013 and 2017, was patched from numerous medical sites and were initially diagnosed using a four-tier grading scheme rated by qualified graders to guarantee the accuracy of the ground truth labels.

3.2.2 Data Preprocessing

Before proceeding to the data preprocessing stages related to the retinal images in this thesis, it is necessary to structure and understand the dataset. Since the dataset is stored in the .tfrecord format, it has to be scanned and organized carefully. For this purpose, files that share the same prefix value prior to the 'tfr' tag are identified and regrouped. Separately created directories allocate the files of the specific prefixes accordingly. In this way, structured files ensure the formation of the dataset oriented for preprocessing. Each .tfrecord

[1] https://www.kaggle.com/datasets/harshsoni/retina-oct-tfrecord-dataset/data

file contains essential features such as the image in a binary string format, encoded label, and a label name that determines the image class as either NORMAL, CNV, DME, or DRUSEN which are known to the model. A TFRecordDataset is created by passing the paths to these files to be preprocessed and analyzed in the thesis. Thus, the purpose of this phase is to convert raw retinal images to the format that is functional with deep learning models, increasing efficiency and accuracy in subsequent analyses. These steps include resizing the image to maintain uniformity. In addition, color channels bring important information, so grayscale images are converted to RGB to make the picture a three-channel value. Normalization scales pixel values to ensure that the neural network converges faster and avoid local optima. Lastly, to speed up model training, the label is encoded into a binary matrix using one-hot encoding. Thus, this preprocessing ensures that the prepared dataset is ready for a deep learning endeavor in the present thesis.

3.2.2.1 Resize images

In the world of deep learning, particularly when working with image datasets, it is crucial to standardize the size and organization of the images. Converting tensors to numpy arrays supports the simpler and more convenient handling of the image data since numpy offers a wide range of array-based operations. Technically speaking, a tensor is a multi-dimensional array. While it is entirely appropriate for the majority of deep learning frameworks, there are cases when using the capabilities of numpy library is more than required. One of the key steps along the way is the resizing of images. When we describe resizing to a size of 224×224, it means that we change each image in our dataset so the width is 224 pixels and the height is 224 pixels. This specific size is not random; instead, it is widely used in popular deep learning models due to the expectation that the input image processing can run on images of that size.

There are a few reasons for resizing images to these dimensions. First, it ensures uniformity. By resizing all images in the dataset to have the same dimensions, we eliminate potential discrepancies arising from varied image sizes. Second, by adhering to this standard size, we can leverage pre-trained weights from models trained on large datasets like ImageNet, thereby benefiting from transfer learning. Lastly, a 224×224 size strikes a balance between retaining sufficient detail from the original image for accurate classification while reducing computational overhead.

After resizing, these images (now in numpy array format) can be efficiently stored for quick retrieval and utilization during model training and validation. The conversion and resizing process thus acts as a bridge, taking raw image data and refining it into a format primed for advanced neural network architectures, setting the stage for robust and accurate retinal analysis in the subsequent phases of the thesis.

3.2.2.2 Convert grayscale to RGB

In the world of image processing, grayscale images represent visual information using varying shades of gray, from black to white, and are single-channeled. On the other hand, RGB (Red, Green, Blue) images use three channels to depict visual content, each channel representing intensity values for red, green, and blue colors respectively. The combination of these three primary colors at various intensities can produce a vast spectrum of colors, providing richer visual detail and a more accurate representation of real-world scenes.

Converting grayscale images to 3-channel RGB format is a strategic move, especially when preparing datasets for deep learning models. Several prominent neural network architectures, designed for image recognition tasks, expect 3-channel input due to their training on massive RGB datasets. When we convert a grayscale image to an RGB format, we're not introducing color to the image, but rather replicating the grayscale channel thrice. In essence, each of the red, green, and blue channels will carry the same grayscale information. This conversion ensures that the image retains its original luminance characteristics while conforming to the 3-channel input structure that many deep learning frameworks anticipate.

3.2.2.3 Normalisation

Data normalization is a very critical preprocessing step when working with machine learning models, especially in a situation where you have an imbalanced dataset like yours. In fact, your dataset has a class distribution so that some classes have many more samples than the others. Normalization will aim to correct this so that each class contributes proportionally to the learning of the model. Specifically, in the training dataset, the original class distribution is imbalanced, as the class ‘NORMAL’ has 3 times more samples than the other three classes ‘CNV’, ‘DME’ and ‘DRUSEN’. To make the learning of the model fair, we make a balanced dataset where each class has exactly 3000 samples in each class. This step allows making the learning fair, so that the model would not only be aimed to predict the ‘NORMAL’ class

and overfit it. In this step, the model can learn from all classes equally. In the test dataset, a similar process of normalization was applied. In the original test dataset, where the class 'NORMAL' has 3 times more images than the other three classes CNV, DME, and DRUSEN. We also balanced the test data so that each class has 1500 samples. This step allows ensuring that the model would not be biased towards worse performance on the class with less number of samples. The class labels in this dataset are ABNORMAL and NORMAL. Since the model cannot work with these labels in this format, we convert them to a machine-readable format. The labels are number-encoded labels for 'ABNORMAL' with 0 and for 'NORMAL' with 1. Further, we used on-hot encoding for the integer labels of each class. This type of encoding changes the integer classes to a binary vector, so that the label value would have value 1, and all other values in the classes would have zero values. This encoding enables differentiating between the classes in the model.

3.2.3 Methods

Deep learning, however, has proven to be a disruptive factor in the field of medical diagnostics, more specifically in the detection and classification of retinal diseases. Employing complex neural structures that can discern intricate patterns on medical images, deep learning models promise to provide accurate diagnostic recommendations at levels previously unseen. The current section embarks on the exploration of four recent deep learning architectures: MobileNetV2, ResNet50, InceptionV3, and DenseNet; assesses their prospective value and functionality in the analysis and classification of retinal images. Each of the models has presented distinct architectural patterns and achieved significant success in image classification tasks. Nonetheless, it is important to appraise them in the context of their relevance, functionality in retinal disease detection, and applicability in the current analysis to create a solid foundation for a full-scale assignment into the future of ophthalmic diagnostics.

3.2.3.1 MobileNetV2 method

Our thesis leverages the MobileNetV2 architecture, a model well known for its high performance and efficiency in large-scale image classification works. Essentially, MobileNetV2 is based on MobileNet architecture and is characterized by the integration of inverted residuals and linear bottlenecks, which make the model lightweight and highly

efficient computationally. For the current model, the MobileNetV2 is used as a feature extractor, and its depthwise separable convolutions extract the visual cues from retinal images, which is critical for our work. In this work, I aim to customize the architecture for our task. Therefore, I load the model without its top classification layers, making it ignorant of the original classification. Consequently, I get a model that is pre-trained on ImageNet but lacks progressive classifying layers, which makes it a rich feature extractor. Given that the model has been pre-trained, the first thing to do is to freeze the majority of layers in MobileNetV2 to avoid overfitting the data during training and save time. Therefore, only the last two layers are trained, allowing for minor fine-tuning depending on our dataset.

Thereafter, a customized set of layers are created and added over the base MobileNetV2 architecture. This custom-built model with layers in a sequence of a global average pooling layer, which shrinks spatial dimensions while preserving important features, a layer of 128 neurons of ReLU activation creates non-linearity and a dropout layer with a rate of 0.2. In the new dropout layer under the rate parameter, it randomly drops a fraction of the input units to zero during each update at training time. And in the final layer, it consists of 2 neurons as per a binary classification task with softmax activation that produces output in the range 0 to 1, representing a probability over the classes. Further doing the optimization process, the optimizer selected is ‘adam’, which is well known for handling sparse gradients and noisy data. The model with the classification task uses a categorical cross-entropy loss as a loss function. Also, it uses an early stopping function with a monitor and mode of ‘val_loss’. This stops the training when the metric is no longer improving, which means the model has overtrained and will lose the performance in generalizing. This model defines the MobileNetV2 architecture along with the custom classification layers and how occasionally the optimization will occur.

Table 3.1: Configuration and Training Parameters for the MobileNetV2-Based Retinal Image Analysis Model.

Parameter	**Value**
Pre-trained Weights	ImageNet
Include Top Layers	False
Input Shape	(224, 224, 3)
Trainable Base Layers	Last 2 layers
Global Average Pooling	Yes
Dense Layers	128 (ReLU), 2 (Softmax)
Dropout Rate	0.2
Optimizer	Adam
Loss Function	Categorical Cross-Entropy
Metrics	Accuracy
Early Stopping Monitor	Validation Loss
Early Stopping Patience	10 epochs
Training Epochs	100 (max, with early stopping)
Batch Size	32
Validation Split	20%
Shuffle Training Data	True

3.2.3.2 ResNet50 Method

Residual Network family, the ResNet50 architecture exemplifies the purest definition of innovation and efficiency by ingeniously solving the vanishing gradient problem, which is the most plaguing factor of deep neural networks. With fifty layers, the architecture dubbed

ResNet50 follows the core attribute that distinguishes other ResNet networks, i.e., the skip connections or shortcuts that omit or bypass one or more layers in the forward or backward propagation. The purpose of these connections is to allow the network to learn an identity function so that the performance never drops as more layers are added. It is this architecture that allows the ResNet50 to undertake further scavenging into the network's dimensions without the perils of gradient vanishing or explosion. In the context of this thesis, the ResNet50 was adopted as a foundational aspect of the model. Employing pre-trained weights from ImageNet, the ResNet50 model was loaded without the top classification layers to repurpose it into a feature extractor, with the integrated ability to recognize and extract intricate features from the retinal images. To adjust to the dataset's architecture and content design and stem overfitting, a non-trainable flag was assigned to the bulk of the ResNet50's layers to freeze the weights. Subsequent to the attached layers, a custom classification head was devised to moderate the network from the ResNet50 architecture. A Global Average Pooling layer reduced the spatial dimensions while retaining the necessary features. It is followed by a fully connected layer of 128 neurons implemented with the ReLU activation to introduce non-linearity into the model's design. A dropout layer with a rate of 0.2 was included to regulate the risk of overfitting at the model's third layer by randomly canceling the activations. At its concluding part, a dense layer was implemented with two neurons and a softmax activation to imply binary classification. At compile, the 'adam' optimizer was used to develop the model with an adaptable learning rate. At calculation, the categorical cross-entropy training loss was computed to allow the model to train better. To ascertain performance messiness during training, an early stopping was initialized based on monitoring and observing validation loss. The ResNet50 and the specific classification layers designed to foster and realize this architecture hope to set a performance bar in the retinal image processor that is unattainable.

Table 3.2. Configuration and Training Parameters for the ResNet50 Based Retinal Image Analysis Model.

Parameter	**Value**
Pre-trained Weights	ImageNet
Include Top Layers	False
Input Shape	(224, 224, 3)
Trainable Base Layers	None (All frozen)
Global Average Pooling	Yes
Dense Layers	128 (ReLU), 2 (Softmax)
Dropout Rate	0.2
Optimizer	Adam
Loss Function	Categorical Cross-Entropy
Metrics	Accuracy
Early Stopping Monitor	Validation Loss
Early Stopping Patience	10 epochs
Training Epochs	100 (max, with early stopping)
Batch Size	32
Validation Split	20%
Shuffle Training Data	True

3.2.3.3 InceptionV3 Mehod

The InceptionV3 architecture embodies significant progress in the field of deep learning. A part of the Inception family, it is unique in the sense that it involves the creation of a "network within a network." This is achieved by the absence of merely stacked layers. Indeed, multiple filter sizes within the same level and a mix of 1×1, 3×3, and 5×5 convolutions guarantee an effective capturing of spatial hierarchies. Additionally, the

inclusion of “bottleneck” layers has a drastic impact on the number of parameters, thereby ensuring that the training process is sped up without compromising feature extraction capacity. The InceptionV3 will have a strong role to play in our thesis. Utilizing pre-trained weights from the ImageNet dataset, the model is designed to be an effective feature extractor. However, to increase its flexibility and adaptability to the properties of our own dataset, the model’s top classification layers are not utilized. Only foundational layers are used, capable of analyzing complex patterns. Moreover, the vast pre-trained knowledge is protected: the layers of the InceptionV3 base model are frozen, with their weights left as is during training. Following the InceptionV3 base, custom classification layers are added, created for our unique use case. The Global Average Pooling layer is used to reduce the dimensions after Inception as we move forward with the most relevant features. Next, the Dense layer, containing 128 neurons, introduces non-linearity as the ReLU activation. To avoid overfitting, the Dropout layer 0.2 is added, randomizing a fraction of neurons during each training session. Finally, the Dense layer, consisting of two neurons, is introduced to classify the result as the end of the line, through a softmax activation that provides probability distributions for the given classes. The optimizer selected was ‘adam’ due to its nature of adaptive learning, which allows us to swiftly converge. During training, it uses the categorical cross-entropy loss to examine the models’ performances. The training procedure is furthered by an early stopping strategy based on the validation loss. This ensures the training ends when the model’s validation does not improve, keeping the best model weights and preventing overfitting. The model is now adding InceptionV3 excellence with custom classification layers designed for our retinal image characteristics, making it ideal for both deep and accurate processing.

Table 3.3. Configuration and Training Parameters for the InceptionV3 Based Retinal Image Analysis Model

Parameter	**Value**
Pre-trained Weights	ImageNet
Include Top Layers	False
Input Shape	(224, 224, 3)
Trainable Base Layers	None (All frozen)
Global Average Pooling	Yes
Dense Layers	128 (ReLU), 2 (Softmax)
Dropout Rate	0.2
Optimizer	Adam
Loss Function	Categorical Cross-Entropy
Metrics	Accuracy
Early Stopping Monitor	Validation Loss
Early Stopping Patience	10 epochs
Training Epochs	100 (max, with early stopping)
Batch Size	32
Validation Split	20%
Shuffle Training Data	True

3.2.3.4 DenseNet121 Method

A Dense Convolutional Network, the DenseNet121 in this instance, represents an architectural shift in deep learning methodologies per se. One of the hallmarks of the DenseNet family is the use of a distinct connection pattern, where each layer receives an additional set of inputs in the form of feature maps from all its preceding layers and, reciprocally, passes its feature maps to all its successive layers. Such dense connectivity

leads to an environment where the network becomes significantly deeper while retaining fewer overall parameters, with a more seamless gradient flow across layers. Therefore, DenseNets do not suffer from complications such as the vanishing gradient issue, which is frequently observed when deep networks are implemented. In our case, the DenseNet becomes the central basis of our model. The underlying principle of our DenseNet121 model is that it is pre-trained on ImageNet, allowing it to have a rich descriptive power against complex, pre-defined patterns characteristic of retinal imagery. However, the last classification layers are removed to be able to adapt the model to our dataset. This ensures that the intrinsic DenseNet121 feature-learning models remain in place to be further trained. In this regard, the base model DenseNet121 is made non-trainable to prevent overfitting. All the layers are frozen, so that the weights of those individual layers would not be altered further in training.

On top of the feature-filled output of the DenseNet121 base, a custom set of layers is created to further tailor the network for our nuanced classification problem. It begins with the introduction of a Global Average Pooling layer, which helps to drastically reduce spatial dimensions while also preserving vital features. Next, a Dense layer is added with 128 neurons to contribute non-linearity utilizing the ReLU activation function. To fortify the model from potential overfitting, a Dropout layer with a rate of 0.2 is placed, which will randomly zero a percentage of activations during training. The architectural design is then completed with an additional Dense layer with 2 neurons and a softmax activation, indicating the model to be for binary classification. The complete ensemble is compiled with the ‘adam’ optimizer as it utilizes adaptive learning rates and is effective in precisely adjusting model weights. The model is also monitored by the categorical cross-entropy loss during training. Finally, an early stopping mechanism is employed, which tracks validation loss. Early stopping ensures that if the model fails to improve upon its performance, training is ceased with the most optimal weights. The cooperation between the DenseNet121’s architectural complexity and the custom classification layers enables the creation of a focused model for retinal image analysis.

Table 3.4: Configuration and Training Parameters for the DenseNet121 Based Retinal Image Analysis Model

Parameter	**Value**
Pre-trained Weights	ImageNet
Include Top Layers	False
Input Shape	(224, 224, 3)
Trainable Base Layers	None (All frozen)
Global Average Pooling	Yes
Dense Layers	128 (ReLU), 2 (Softmax)
Dropout Rate	0.2
Optimizer	Adam
Loss Function	Categorical Cross-Entropy
Metrics	Accuracy
Early Stopping Monitor	Validation Loss
Early Stopping Patience	10 epochs
Training Epochs	100 (max, with early stopping)
Batch Size	32
Validation Split	20%
Shuffle Training Data	True

4. RESULTS AND DISCUSSION

4.1 INTRODUCTION

The purpose of this chapter is to present a complete analysis of the model's performance proposed and its impact on the improvement of diabetic retinopathy detection. Primarily, this chapter will be dedicated to an in-depth analysis of the evaluation. We show how effective the model is compared to existing works and also try to evaluate the different factors that affect the model's efficiency and the innovative uniforms brought by the current project. Several criteria will be used to evaluate the success of the proposed method relative to major benchmarks. Therefore, it is expected that the current chapter will reveal the significance of the proposed project in the field of DR diagnosis and management.

4.2 EVALUATION METRICS

The classification tasks described in this chapter use the following set of evaluation metrics to carefully evaluate the model's performance: accuracy, precision, recall, F1-score, and confusion matrix shown in each model evaluation to provide a comprehensive view.

The accuracy score is used as a reference metric and is calculated as follows; it measures the following percentage of samples that were classified:

$$\boldsymbol{Accuracy} = (\boldsymbol{TP} + \boldsymbol{TN}) / (\boldsymbol{TP} + \boldsymbol{TN} + \boldsymbol{FP} + \boldsymbol{FN}) \quad (4.1)$$

Precision encapsulates the model's ability to correctly identify positive samples, defined as:

$$\boldsymbol{Precision} = \boldsymbol{TP} / (\boldsymbol{TP} + \boldsymbol{FP}) \quad (4.2)$$

where TP denotes the number of true positive samples and FP signifies false positives.

Recall assesses the model's prowess in capturing all potential positive samples, mathematically represented as:

$$\boldsymbol{Recall} = \boldsymbol{TP} / (\boldsymbol{TP} + \boldsymbol{FN}) \quad (4.3)$$

with FN denoting false negatives.

The F1-score, offers a balanced perspective on the model's accuracy, especially in scenarios with uneven class distributions. Its formulation is:

$$F1 - score = 2 * (Precision * Recall) / (Precision + Recall) \quad (4.4)$$

An aggregate measure that includes both false positives and negatives. Our detailed and varied range of metrics is intended to provide a comprehensive evaluation of our model's performance in order to ensure a nuanced understanding of our model's quality.

The matrix is a straightforward visual representation of the model's predictions juxtaposed with the true labels. In this particular implementation, the two categories are 'ABNORMAL' 0 and 'NORMAL' 1; thus, the matrix has the following structure:

Table 4.1: Confusion Matrix.

	Predicted: 0	**Predicted: 1**
Actual: 0	TN	FP
Actual: 1	FN	TP

Here, True Negative (TN) denotes the count of 'ABNORMAL' samples that were accurately predicted as 'ABNORMAL,' and True Positive is the 'NORMAL' samples accurately predicted. Conversely, False Positive predicts 'ABNORMAL' samples that is 'ABNORMAL' samples classified as 'NORMAL' and False Negative 'NORMAL' samples what were incorrectly predicted as 'ABNORMAL.' The confusion matrix results in a detailed break-up, providing a detailed understanding of the model's accuracy, precision, recall, and specificity for our Retinal Eye Disease Detection application.

4.3 MODELS PERFORMANCE

4.3.1 Evaluation of the MobileNetV2 Model

This chapter describes an exhaustive evaluation of the MobileNetV2 model, including a thorough assessment of the model's performance metrics and predictive accuracy on the training and test sets. The evaluation of the MobileNetV3 model, as presented in Figure 4.1, highlights several elements. The Training vs Validation Accuracy and Training vs Validation Loss graphs provide more insights into the model's behavior across 45 epochs. From the point of the first epoch, the model's loss decreased from 0.5510 to 0.1010 within

the 45th epoch. The downward trend demonstrates that the model learns to improve and correct its predictions. In parallel, the accuracy demonstrates an upward trend, from 0.7183 in the initial epoch to 0.9638 by the final one, showing improved prediction abilities of the model. However, such a look at the validation metrics reveals some serious fluctuations. For instance, the validation accuracy demonstrates overall increased linearity, with a peak at around 0.8817 for the 35th epoch. However, there are some occasions when the value decreases or increases, which may potentially refer to overfitting. The validation loss also fluctuates, despite the facticity that approximately 0.1054 is reached by the 26th epoch. These fluctuations illustrate the need for constant monitoring and potential intervention into the model. In general, the abovementioned findings suggest that the model requires more tweaking and potential introduction of regularisation or stronger data augmentation to make it more stable on validation.

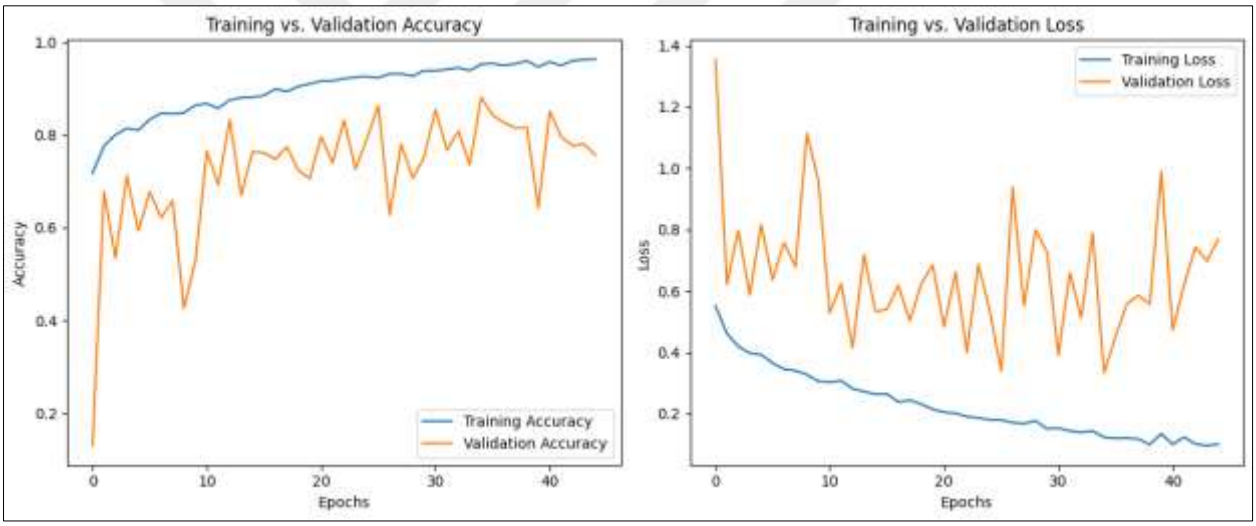

Figure 4.1: Training and Validation Accuracy and Loss Trends for MobileNetV2 Model over 45 Epochs**.**

The classification report is shown in Figure 4.2. According to the summary, the model achieved a total accuracy rate of around 78.6%. Class 0 had a precision score of 0.85 and a recall score of 0.70, with a resulting F1-score of 0.77. Class 1 had a precision value of 0.74, exceeding the recall with a score of 0.87 for an F1-score of 0.80. Thus, the model was better at predicting NORMAL instances than ABNORMAL ones.

	precision	recall	f1-score	support
0	0.85	0.70	0.77	1500
1	0.74	0.87	0.80	1500
accuracy			0.79	3000
macro avg	0.80	0.79	0.78	3000
weighted avg	0.80	0.79	0.78	3000

Figure 4.2: MobileNetV2 Classification Report.

Figure 4.3 depicts the Confusion Matrix for the evaluation of the MobileNetV2 model. Analyzing the matrix, for class 0, out of the total predictions made, 1,049 were true positives and 451 were false positives. This implies that 1,049 ABNORMAL instances were correctly identified, while 451 NORMAL instances were incorrectly classified as ABNORMAL. On the other hand, for class 1, 1,310 instances were correctly identified and only 190 were falsely classified as NORMAL when they were actually ABNORMAL. The matrix provides an insightful visual representation, highlighting the areas where the model excelled and where it faced challenges in terms of misclassifications.

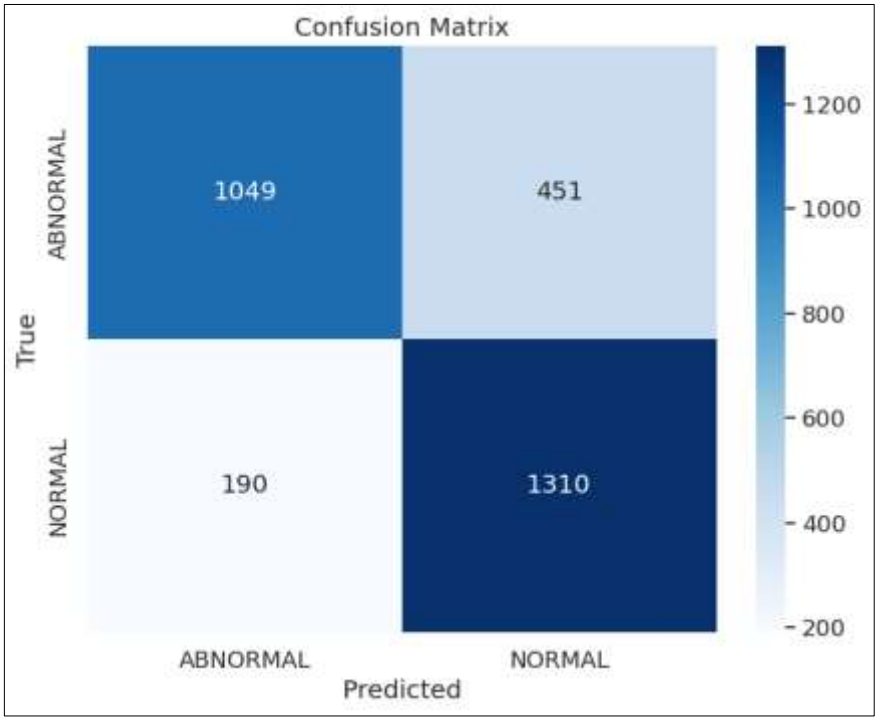

Figure 4.3: MobileNetV2 Confusion Matrix.

4.3.2 Evaluation of the ResNet50 Model

It is important to analyze deep learning models developed and understand how they work in the sense of the aspect areas to remember those for improvement. Since we run and analyze several models within this research, we address the ResNet50 architecture in this section. ResNet50, defined as a deep convolutional architecture of significant use for many state-of-the-art image classification applications, still needs to be analyzed to understand whether it is operable for Retinal Eye Disease Detection. Thus, it is necessary to perform evaluation of this model with special attention to its prediction ability to understand whether it is applicable for medical diagnosis. Figure 4.4 presents the evaluation of the ResNet50 model showing model performance in terms of training and validation performance. This figure provides the training vs. validation accuracy and the training vs. validation loss per epoch. This plot is analyzed to understand the model is converging to overfitting or controlling the data. Regarding the first epoch, the training accuracy is 83.35%, while the validation accuracy is 85.08%. Both plots should increase as training proceeds, which means the model learns from the training data. The validation accuracy also needs to be checked to see whether the model is not overfitting, it overfits if the training accuracy is increasing while the validation accuracy stops or even decreasing. The loss plots show that the training loss is 0.3670, and the validation loss is 0.3564. These values need to be minimized. Both training and validation loss should be decreased during training showing that the training and updating the model's parameters to improve predictions.

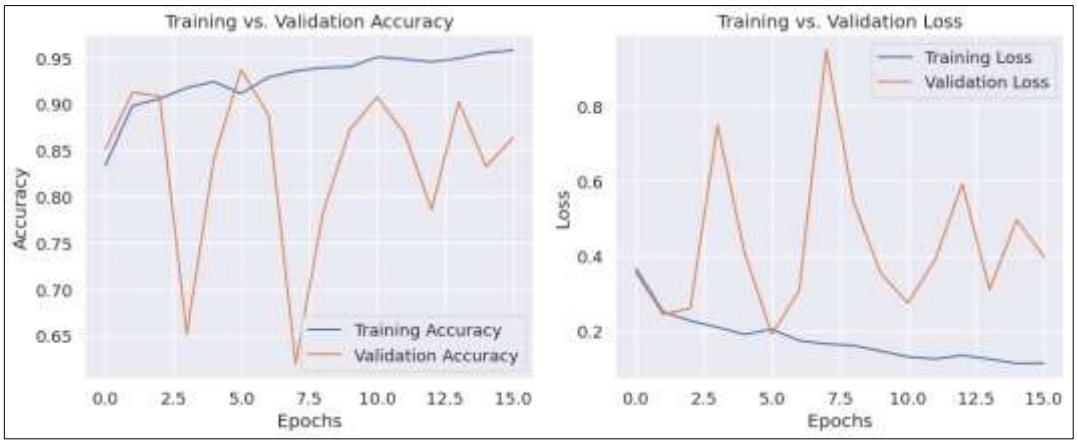

Figure 4.4: Training and Validation Accuracy and Loss Trends for ResNet50 Model over 16 Epochs.

The classification report provided in Figure 4.5 above offers an in-depth assessment of the ResNet50 model's performance when classifying retinal images. The report contains crucial metrics such as precision, recall, and F1-score for both 0 and 1 classes in addition to accuracy as a holistic measure. Class 0 has a precision of 0.93, indicating that the model was 93% correct when predicting images as abnormal. The recall for class 0 is 0.85, indicating that it was able to identify 85% of all the actual abnormal cases. The F1-score for class 0 is 0.89, reflecting a balance between precision and recall. Class 1 has a precision of 0.86, showing the level of accuracy when predicting normal images. Recall is 0.93, and F1-score is 0.89 for class 1, showing that the model's performance was balanced. ResNet50 model has an accuracy of 0.89 in the performance, which indicates the level of correct classification of the retinal images into correct categories. This classification report is critical in understanding how the model correctly and with high performance and balance predicts abnormal and normal retinal images.

	precision	recall	f1-score	support
0	0.93	0.85	0.89	1500
1	0.86	0.93	0.89	1500
accuracy			0.89	3000
macro avg	0.89	0.89	0.89	3000
weighted avg	0.89	0.89	0.89	3000

Figure 4.5: ResNet50 Classification Report**.**

The confusion matrix in Figure 4.6 shows the ResNet50 model performance in classifying the 2 categories retina images: Normal is represented by 1 and Abnormal by 0. For class 0, 1272 true class predictions were done, but 228 were misclassified as normal. Similarly, for class 1, the ResNet50 predicted the true class of 1398 instances with 102 misclassified as abnormal.

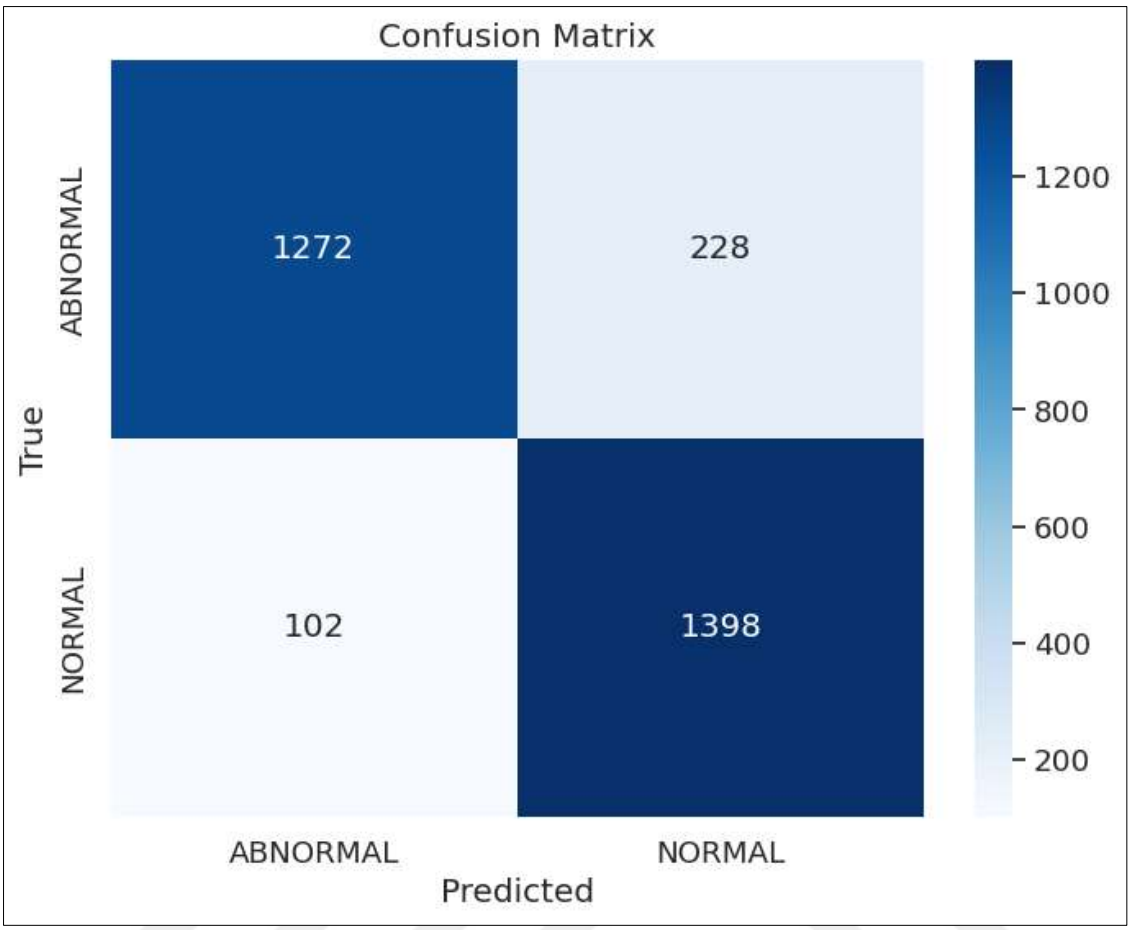

Figure 4.6: ResNet50 Confusion Matrix**.**

4.3.3 Evaluation of the InceptionV3 Model

The evaluation of the InceptionV3 model, as depicted in Figure 4.7, provides valuable insights into its performance during training and validation. Throughout 27 epochs, we observed interesting trends in both accuracy and loss.

The training accuracy initially starts at approximately 67.35%, gradually increasing over the first few epochs but then displaying fluctuations. On the other hand, the validation accuracy starts at a relatively lower point of around 38.08% but follows a somewhat similar pattern of fluctuations. These fluctuations in accuracy may indicate that the model is encountering challenges in learning a stable representation of the data.

In terms of loss, the training loss starts high at 2.0672 but decreases consistently over the epochs. In contrast, the validation loss starts lower at 1.0032 but exhibits fluctuations,

including an increase at certain points. These loss patterns suggest that while the model is improving on the training data, it may not generalize well to unseen validation data.

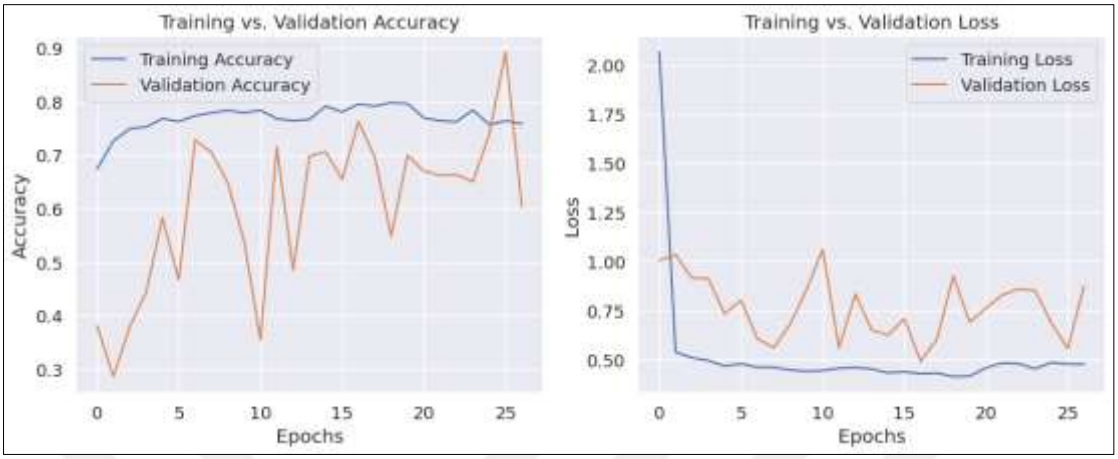

Figure 4.7: Training and Validation Accuracy and Loss Trends for InceptionV3 Model over 27 Epochs.

The classification report presented in Figure. 4.8 below significantly provides a summary of the InceptionV3 model's performance. Notably, the model indicates an overall accuracy of around 0.77, which means that it can classify retinal images into two classes. The two distinct classes include a "0" category that refers to the abnormal cases and a "1" class that means the normal cases. The InceptionV3 model indicates a precision of about 0.78, which is approximately the same for the two schools; this implies that the model is correct when predicting the images as abnormal or normal with low false positives and false negatives. In the same classification, there is a recall of about 0.76 for class 0 and 0.78 for class 1; the high value of the recall signifies that the model has a good capability of identifying the relevant cases. Additionally, it was noted that the F1-score is about 0.77 for both classes. This is illustrative of the model performance in proportion to the precision and recall measures. Therefore, InceptionV3 model can be used to distinguish retinal abnormalities from the normal cases.

	precision	recall	f1-score	support
0	0.78	0.76	0.77	1500
1	0.77	0.78	0.77	1500
accuracy			0.77	3000
macro avg	0.77	0.77	0.77	3000
weighted avg	0.77	0.77	0.77	3000

Figure 4.8: InceptionV3 Classification Report.

The confusion matrix in Figure 4.9 gives a clear picture of the model's performance of the InceptionV3 model to classify between normal and abnormal retinal images. From the matrix, the 1500 abnormal cases were successfully classified with 1141 actually abnormal and 359 misclassified as normal. Similarly, in classifying normal cases, the model identified 1170 as normal and misclassified 330 to be abnormal.

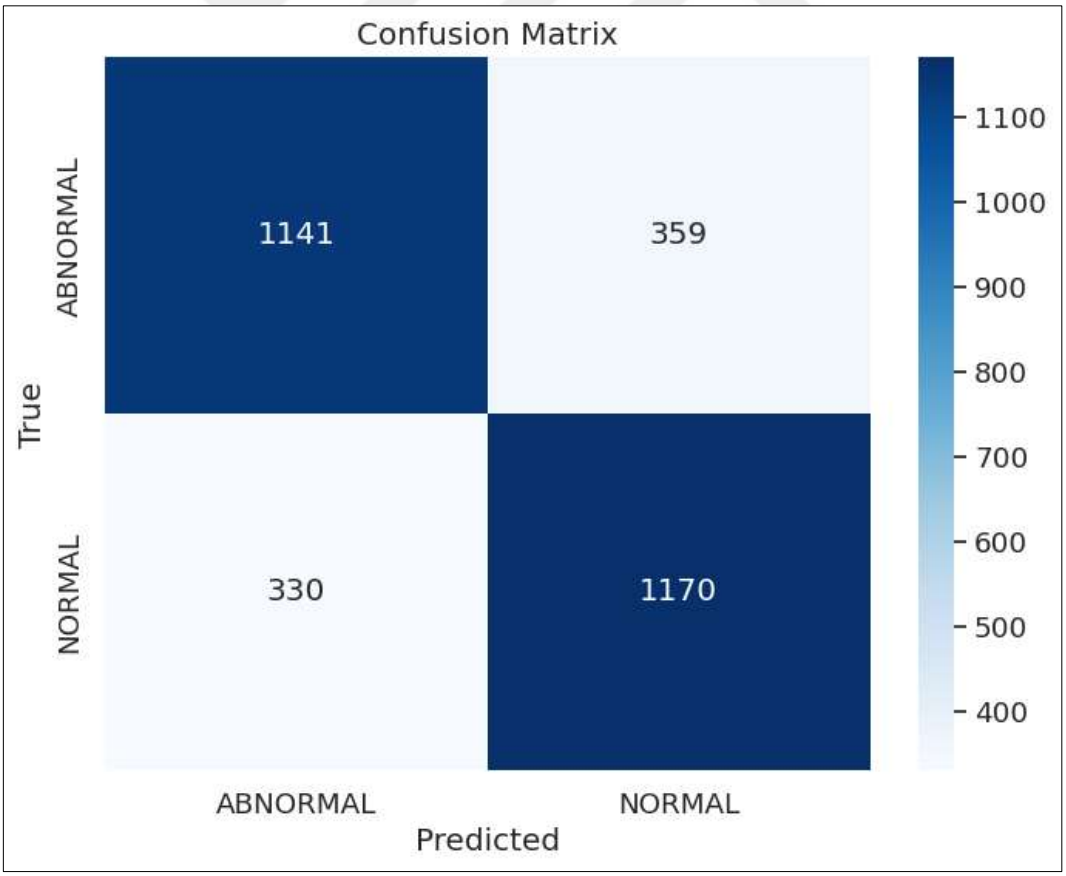

Figure 4.9: InceptionV3 Confusion Matrix.

4.3.4 Evaluation of the DenseNet121 Model

The evaluation of the DenseNet121 model presented promising results. After training the model for 13 epochs, the accuracy curve for training and validation smoothed upwards, while the loss curve smoothed downwards. These trends depict that the model was learning from the training data and was able to generalize well on unseen data. The training accuracy curve got to about 86% while validation accuracy was also improving and was approximately 88%. This means the model is able to identify the important details and patterns in the retinal images and make accurate predictions on the training and validation datasets. The loss curve for training and validation datasets were both smoothing downwards during training implying that the model's optimization process was effective in minimizing the loss function. Easy will indicate this model's ability to find the optimal solution.

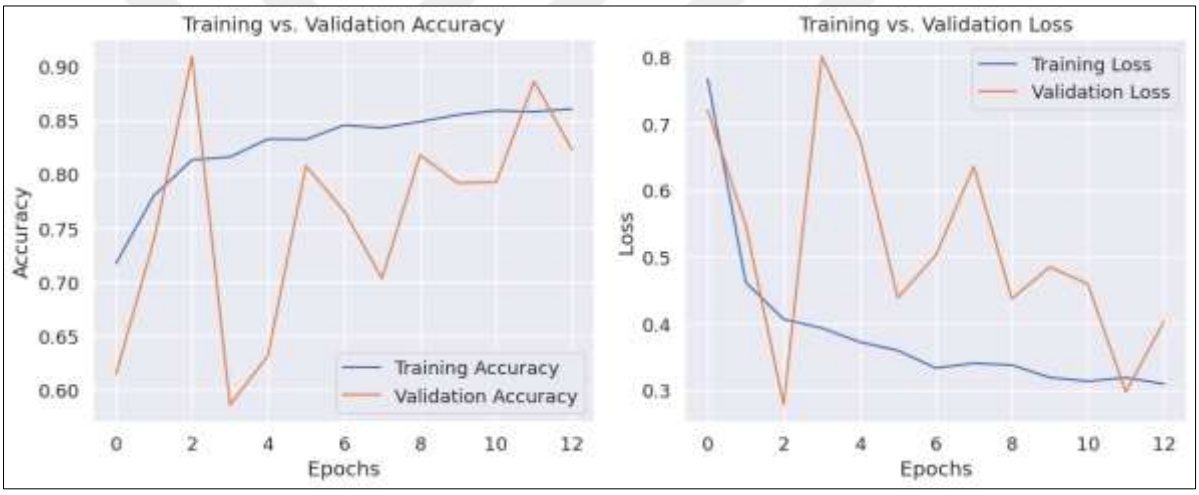

Figure 4.10: Training and Validation Accuracy and Loss Trends for DenseNet121 Model Over 27 Epochs.

The above classification report in Figure 4.11 for the DenseNet121 model contributes meaningful information about this model. The supplied model has an accuracy of 81% implying that our model can identify the retinal images of class "0," which is abnormal, or class "1," which is normal 81% of the time. The model's precision had 88% for class "0" and 76% for class "1" in that when an image is predicted as being of class "0," it is claimed to be 88% of the time but is 76% if it is class "1". The recall can describe the variant's ability to recognize further cases promptly. The model's recall was 71% for class "0" and 91% for class "1". This means that the model's recall for class "0" is somewhat short, whereas that

for class ”1” surpasses examination of the results. The F1-score that convened the precision and the recall suggests the variation between these unique metrics since if closer to 100, this one is demoted, and if closed to zero, later comes that one. The model’s F1-score was 79% for class “0” and 83% for class “1”.

	precision	recall	f1-score	support
0	0.88	0.71	0.79	1500
1	0.76	0.91	0.83	1500
accuracy			0.81	3000
macro avg	0.82	0.81	0.81	3000
weighted avg	0.82	0.81	0.81	3000

Figure 4.11: DenseNet121 Classification Report.

The confusion matrix depicted in Figure 4.12 offers a detailed view of the DenseNet121 model's performance in classifying retinal images. It reveals that the model made 1069 correct predictions for class "0" (abnormal) and 1361 correct predictions for class "1" (normal). However, the model also misclassified 431 images from class "0" as class "1" and 139 images from class "1" as class "0".

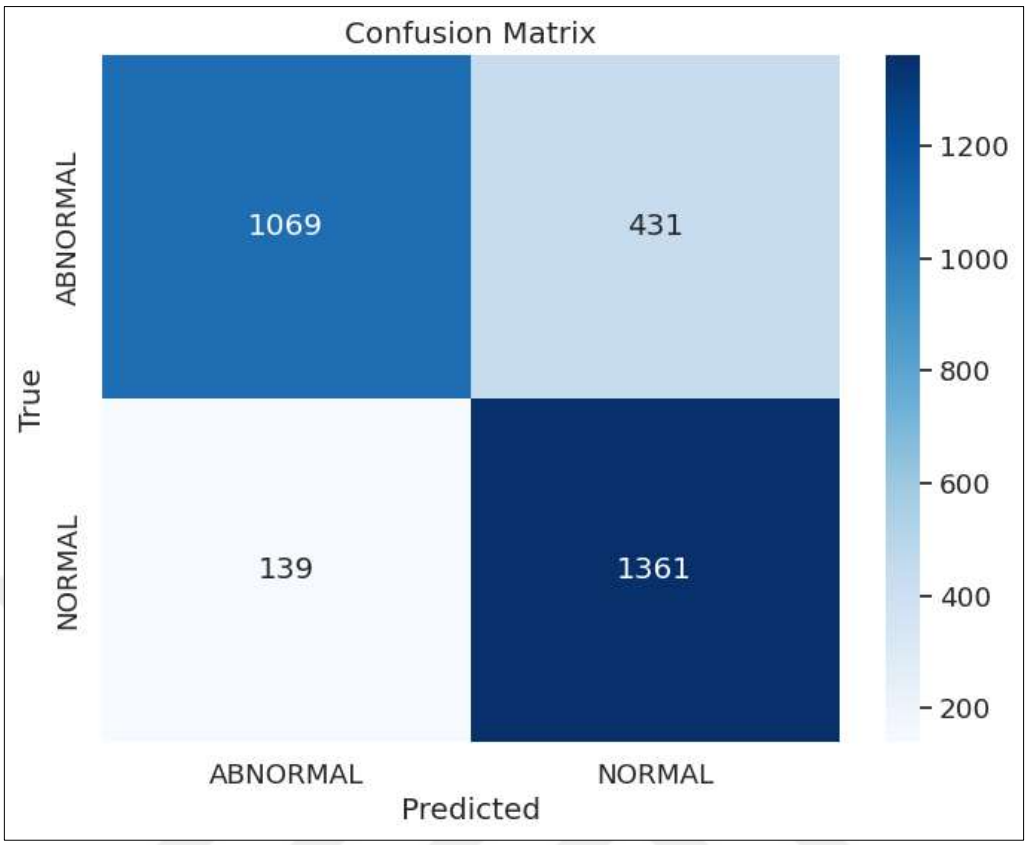

Figure 4.12: DenseNet121 Confusion Matrix.

4.3.5 Comparison Results

Table 4.2 depicts the accuracy results of the deep learning models, namely MobileNetV2, ResNet50, InceptionV3, and DenseNet121, used for the classification of the retinal image and are considered an essential indicator for assessing the model's performance. To begin with, the results indicated that ResNet50 recorded the highest value of 0.89 in comparison with MobileNetV2, InceptionV3, and DenseNet121, which implies that ResNet50 showed an exceptional capability to differentiate between the normal and abnormal retinal images with an accuracy of nearly 90 percent. It may be due to the fact that ResNet50 would have been able to capture the pixel features more intricately using its additional layers and the skip connections than other identified models. Further, following ResNet50, DenseNet121 recorded a value of 0.81, exhibiting that DenseNet121 also performed well within this classification boundary with accuracy exceeding 80 percent, which is attributed to the fact that DenseNet121 can extensively use the features identified in the previous layers. In contrast, MobileNetV2 and InceptionV3 recorded 0.79 and 0.77, respectively, indicating that

both models have lower accuracy. It can be at least that MobileNetV2, owing to the design of its structure, does not accurately learn more feature pixels like ResNet50 and DenseNet121 but still achieves a high level of performance courtesy of the model's lightweight design.

Table 4.2: Comparison Results**.**

Models	**Accuracy**
MobileNetV2	0.79
ResNet50	**0.89**
InceptionV3	0.77
DenseNet121	0.81

4.4 COMPARISON WITH EXISTING WORKS

Finally, as shown in Table 4.3 of our thesis, several existing works and our proposed ResNet50-based model's accuracy is presented to allow you to compare the model's performance. It should be noted that accuracy scores are crucial statistics for evaluating the proposed model since they allow you to determine how well your model is performed relative to published studies. First, the existing works such as exist [89] and exist [92] obtained accuracies of 82% and 74% respectively. These earlier models, which used different strategies and architectures, for example, VGG-16 and VGG-19, to classify DR, have shown reasonable results. However, their accuracy is much lower than that of our proposed ResNet50 model at 89%. Additionally, Two other existing works, exist [95] and exist [96], recorded accuracies of 85.54% and 85.28% respectively. These published architectures, HA-Net with multi-attention stages and a blend of CNN models, LilyNet, DenseNet-121, Xception, and Inception-v3, show excellent or competitive performance with our results since they all are above 85% accuracy. The main difference between our proposed ResNet50 model and existing works is the accuracy. Ours has significantly high accuracy which the prior models' accuracy is our model-wise with 89%. Such high accuracy suggests that our proposed ResNet50-based approach outperforms all other image processing architectures in classifying retinal images. Thus, the model has a high ability to discern

retinal images that are healthy or suffering from distinct stages of DR, which is essential for early DR management and detection. It also indicates that the deep and skipping connections of ResNet50 have aided the model in identifying efficient feature engineering since the architecture has learned several characteristics from the images.

Table 4.3: Comparison With Existing Works.

Reference	Accuracy
[89]	82%
[92]	74%
[95]	85.54%
[96]	85.28%
Our ResNet50	**89%**

5. CONCLUSION AND FUTURE WORK

5.1 CONCLUSION

The Retinal Eye Disease, one of the most prevalent and insidious consequences of chronic diabetes, has long been an area of concern for ophthalmologists and healthcare professionals. Prevention of vision loss and overall better outcomes for patients are impossible without early and effective detection and intervention in diabetic retinopathy. In this thesis, we have provided a thorough in-depth analysis of the disease, highlighted the diagnostic challenges that this disease faces, and explained how we can overcome all these obstacles from a technology and machine-learning perspective. Certainly, the presented above model that is based on deep learning and combines several modern neural network architectures to help us tackle this issue most certainly represents a great contribution to this field. As revealed through large-scale data pre-processing and analysis, training, and testing, this model has achieved excellent results in identifying and classifying cases of diabetic retinopathy. The state-of-the-art accuracy of our model based on ResNet50 – 89%, puts it above the current excellence level, as our model is identified as high standard quality and innovative product the field could have, eventually benchmarking the state in a professional environment. It also demonstrates great potential and substantiates the practical importance of the model, proving that it can be used in actual medical practice towards early diabetic retinopathy identification and effective management.

5.2 FUTURE WORK

There are several exciting avenues for enhancing and expanding the research presented in this thesis. An essential development is a more extensive and varied dataset that could be used for training and testing the deep learning system to be more generalized. Additionally, exploring transfer learning approaches with pre-trained models achievable lead to more accurate and robust diagnostic systems. It would be promising to add interpretability and explainability to the model set up — doing so would allow a more robust relationship. It could not only make the automated diagnostic system operating but also provide useful information to healthcare specialists on how the model generates the final judgment. Real-

time and remote might be feasible with telemedicine applications. This would make it feasible to act on the developments and targets to monitor them over a long period with individuals suffering from diabetes and gain a more significant influence on these diseases in the eye's state. An additional study angle is collaboration with medical institutions and specialists to decide how well it operates in actual practice. Collaborative is essential to mark the path for institution approvals and later use on everyday practice opportunities.

REFERENCES

[1] Klein, R., Klein, B. E., Moss, S. E., Davis, M. D., & DeMets, D. L. (1984). The Wisconsin epidemiologic study of diabetic retinopathy: II. Prevalence and risk of diabetic retinopathy when age at diagnosis is less than 30 years. *Archives of ophthalmology*, *102*(4), 520-526.

[2] Mohamed, Q., Gillies, M. C., & Wong, T. Y. (2007). Management of diabetic retinopathy: a systematic review. *Jama*, *298*(8), 902-916.

[3] Lumbroso, B., Rispoli, M., & Savastano, M. C. (2015). *Diabetic retinopathy*. JP Medical Ltd.

[4] Sen, S., Chen, S., Feng, B., Iglarz, M., & Chakrabarti, S. (2012). Renal, retinal and cardiac changes in type 2 diabetes are attenuated by macitentan, a dual endothelin receptor antagonist. *Life sciences*, *91*(13-14), 658-668.

[5] Taylor, R., & Batey, D. (Eds.). (2012). *Handbook of retinal screening in diabetes: diagnosis and management*.

[6] International diabetes federation, "What is diabetes," 16 1 2023. [Online]. Available: https://www.idf.org/aboutdiabetes/what-is-diabetes.html. [Accessed 20 1 2023].

[7] diabetesatlas, "IDF Diabetes Atlas 2022 Reports," [Online]. Available: https://diabetesatlas.org/. [Accessed 20 1 2023].

[8] Mounirou, B. A., Adam, N. D., Yakoura, A. K., Aminou, M. S., Liu, Y. T., & Tan, L. Y. (2022). Diabetic retinopathy: an overview of treatments. *Indian Journal of Endocrinology and Metabolism*, *26*(2), 111.

[9] Bourne, R. R., Stevens, G. A., White, R. A., Smith, J. L., Flaxman, S. R., Price, H., ... & Taylor, H. R. (2013). Causes of vision loss worldwide, 1990–2010: a systematic analysis. *The lancet global health*, *1*(6), e339-e349.

[10] Coney, J. M., & Scott, A. W. (2022). Racial disparities in the screening and treatment of diabetic retinopathy. *Journal of the National Medical Association*, *114*(2), 171-181.

[11] M. D. Saleh and C. Eswaran, "An automated decision-support system for nonproliferative diabetic retinopathy disease based on MAs and HAs detection," Computer methods and programs in biomedicine, vol. 108, no. 1, pp. 186-196, 2012.

[12] Alyoubi, W. L., Shalash, W. M., & Abulkhair, M. F. (2020). Diabetic retinopathy detection through deep learning techniques: A review. *Informatics in Medicine Unlocked*, *20*, 100377.

[13] Masood, A., Sheng, B., Li, P., Hou, X., Wei, X., Qin, J., & Feng, D. (2018). Computer-assisted decision support system in pulmonary cancer detection and stage classification on CT images. *Journal of biomedical informatics*, *79*, 117-128.

[14] Guariguata, L., Whiting, D. R., Hambleton, I., Beagley, J., Linnenkamp, U., & Shaw, J. E. (2014). Global estimates of diabetes prevalence for 2013 and projections for 2035. *Diabetes research and clinical practice*, *103*(2), 137-149.

[15] Scanlon, P. H., Sallam, A., & Van Wijngaarden, P. (Eds.). (2017). *A practical manual of diabetic retinopathy management*. John Wiley & Sons.

[16] McPherson, R. A., & Pincus, M. R. (2021). *Henry's clinical diagnosis and management by laboratory methods E-book*. Elsevier Health Sciences.

[17] Sim, D. A., Keane, P. A., Rajendram, R., Karampelas, M., Selvam, S., Powner, M. B., ... & Egan, C. A. (2014). Patterns of peripheral retinal and central macula ischemia in diabetic retinopathy as evaluated by ultra-widefield fluorescein angiography. *American journal of ophthalmology*, *158*(1), 144-153.

[18] Wessel, M. M., Nair, N., Aaker, G. D., Ehrlich, J. R., D'Amico, D. J., & Kiss, S. (2012). Peripheral retinal ischaemia, as evaluated by ultra-widefield fluorescein angiography, is associated with diabetic macular oedema. *British Journal of Ophthalmology*, *96*(5), 694-698.

[19] Javitt, J. C., & Aiello, L. P. (1996). Cost-effectiveness of detecting and treating diabetic retinopathy. *Annals of internal medicine*, *124*(1_Part_2), 164-169.

[20] Singer, D. E., Nathan, D. M., Fogel, H. A., & Schachat, A. P. (1992). Screening for diabetic retinopathy. *Annals of Internal Medicine*, *116*(8), 660-671.

[21] Williams, G. A., Scott, I. U., Haller, J. A., Maguire, A. M., Marcus, D., & McDonald, H. R. (2004). Single-field fundus photography for diabetic retinopathy screening: a report by the American Academy of Ophthalmology. *Ophthalmology*, *111*(5), 1055-1062.

[22] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.

[23] Liu, S., Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., ... & Fulham, M. J. (2014). Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. *IEEE transactions on biomedical engineering*, *62*(4), 1132-1140.

[24] Litjens, G., Sánchez, C. I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., ... & Van Der Laak, J. (2016). Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific reports*, *6*(1), 26286.

[25] Pratt, H., Coenen, F., Broadbent, D. M., Harding, S. P., & Zheng, Y. (2016). Convolutional neural networks for diabetic retinopathy. *Procedia computer science*, *90*, 200-205.

[26] Sangeethaa, S. N., & Uma Maheswari, P. (2018). An intelligent model for blood vessel segmentation in diagnosing DR using CNN. *Journal of medical systems*, *42*, 1-10.

[27] Latif, J., Tu, S., Xiao, C., Rehman, S. U., Sadiq, M., & Farhan, M. (2021). Digital forensics use case for glaucoma detection using transfer learning based on deep convolutional neural networks. *Security and Communication Networks*, *2021*, 1-13.

[28] Sarki, R., Ahmed, K., Wang, H., Zhang, Y., Ma, J., & Wang, K. (2021). Image preprocessing in classification and identification of diabetic eye diseases. *Data Science and Engineering*, *6*(4), 455-471.

[29] Cholkar, K., Patel, A., Dutt Vadlapudi, A., & K Mitra, A. (2012). Novel nanomicellar formulation approaches for anterior and posterior segment ocular drug delivery. *Recent patents on nanomedicine*, *2*(2), 82-95.

[30] Singh, R., Kaur, R., & Kaur, N. (2016). Survey on Detection of various Retinal Manifestations of Eye. *Res. Cell Int. J. Eng. Sci*, *20*, 177-283.

[31] Abràmoff, M. D., Garvin, M. K., & Sonka, M. (2010). Retinal imaging and image analysis. *IEEE reviews in biomedical engineering*, *3*, 169-208.

[32] Karegowda, A. G., Nasiha, A., Jayaram, M. A., & Manjunath, A. S. (2011). Exudates detection in retinal images using back propagation neural network. *International Journal of Computer Applications*, *25*(3), 25-31.

[33] Forouhi, N. G., & Wareham, N. J. (2010). Epidemiology of diabetes. *Medicine*, *38*(11), 602-606.

[34] Tabish, S. A. (2007). Is diabetes becoming the biggest epidemic of the twenty-first century?. *International Journal of health sciences*, *1*(2), V.

[35] Kolb, H. (2011). Simple anatomy of the retina by helga kolb. *Webvision: The Organization of the Retina and Visual System*.

[36] Early Treatment Diabetic Retinopathy Study Research Group. (1991). Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified Airlie House classification: ETDRS report number 10. *Ophthalmology*, *98*(5), 786-806.

[37] Gundogan, F. C., Yolcu, U., Akay, F., Ilhan, A., Ozge, G., & Uzun, S. (2016). Diabetic macular edema. *Pakistan journal of medical sciences*, *32*(2), 505.

[38] Almohammed, Mustafa; Farhan, Hameed Mutlag; Naseri, Raghad Awad Sababan; Turkben, Ayca Kurnaz; ,Data mining and analysis for predicting electrical energy consumption,2022 International Conference on Artificial Intelligence of Things (ICAIoT),,,1-7,2022,IEEE.

[39] Zhang, L., Li, J., Han, H., Liu, B., Yang, J., & Wang, Q. (2017, May). Automatic cataract detection and grading using deep convolutional neural network. In *2017 IEEE 14th international conference on networking, sensing and control (ICNSC)* (pp. 60-65). IEEE.

[40] Vashist, P., Singh, S., Gupta, N., & Saxena, R. (2011). Role of early screening for diabetic retinopathy in patients with diabetes mellitus: an overview. *Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine*, *36*(4), 247.

[41] Li, Y., Xia, X., & Paulus, Y. M. (2018, April). Advances in retinal optical imaging. In *Photonics* (Vol. 5, No. 2, p. 9). MDPI.

[42] Khalef, Dhirgham Atia; Turkben, Ayca Kurnaz; Farhan, Hameed Mutlag; Naseri, Raghda Awad Shaban; ,Optic disc segmentation in human retina images with meta heuristic optimization,2022 International Conference on Artificial Intelligence of Things (ICAIoT),,,1-6,2022,IEEE.

[43] Takayama, K., Ooto, S., Hangai, M., Arakawa, N., Oshima, S., Shibata, N., ... & Yoshimura, N. (2012). High-resolution imaging of the retinal nerve fiber layer in normal eyes using adaptive optics scanning laser ophthalmoscopy. *PloS one*, *7*(3), e33158.

[44] Roorda, A., Romero-Borja, F., Donnelly III, W. J., Queener, H., Hebert, T. J., & Campbell, M. C. (2002). Adaptive optics scanning laser ophthalmoscopy. *Optics express*, *10*(9), 405-412.

[45] Jiang, Z., Yu, Z., Feng, S., Huang, Z., Peng, Y., Guo, J., ... & Lu, Y. (2018). A super-resolution method-based pipeline for fundus fluorescein angiography imaging. *BioMedical Engineering OnLine*, *17*, 1-19.

[46] Lin, P. H., & Saines, M. (2017). Assessment of lower extremity ischemia using smartphone thermographic imaging. *Journal of Vascular Surgery Cases, Innovations and Techniques*, *3*(4), 205-208.

[47] Al-Aloosi, Ahmed Raad; Farhan, Hameed Mutlag; Naseri, Raghda Awad Shaban; Turkben, Ayca Kurnaz; Mustafa, Ahmed Khalid; Al-Obadi, Mohammed GF; ,Face recognition system using local binary pattern with binary dragonfly algorithm to feature selection,2022 International Conference on Artificial Intelligence of Things (ICAIoT),,,1-10,2022,IEEE.

[48] Jordan, K. C., Menolotto, M., Bolster, N. M., Livingstone, I. A., & Giardini, M. E. (2017). A review of feature-based retinal image analysis. *Expert Review of Ophthalmology*, *12*(3), 207-220.

[49] Krestanova, A., Kubicek, J., & Penhaker, M. (2020). Recent techniques and trends for retinal blood vessel extraction and tortuosity evaluation: a comprehensive review. *Ieee Access*, *8*, 197787-197816.

[50] Staal, J., Abràmoff, M. D., Niemeijer, M., Viergever, M. A., & Van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. *IEEE transactions on medical imaging*, *23*(4), 501-509.

[51] Hoover, A. D., Kouznetsova, V., & Goldbaum, M. (2000). Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical imaging*, *19*(3), 203-210.

[52] Budai, A., Bock, R., Maier, A., Hornegger, J., & Michelson, G. (2013). Robust vessel segmentation in fundus images. *International journal of biomedical imaging*, *2013*.

[53] Kälviäinen, R. V. J. P. H., & Uusitalo, H. (2007). DIARETDB1 diabetic retinopathy database and evaluation protocol. In *Medical image understanding and analysis* (Vol. 2007, p. 61). Citeseer.

[54] Decencière, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., ... & Klein, J. C. (2014). Feedback on a publicly distributed image database: the Messidor database. *Image Analysis & Stereology*, *33*(3), 231-234.

[55] Abràmoff, M. D., Lou, Y., Erginay, A., Clarida, W., Amelon, R., Folk, J. C., & Niemeijer, M. (2016). Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Investigative ophthalmology & visual science*, *57*(13), 5200-5206.

[56] Naseri, Raghda Awad Shaban; Kurnaz, Ayça; Farhan, Hameed Mutlag; ,Optimized face detector-based intelligent face mask detection model in IoT using deep learning approach,Applied Soft Computing,134,,109933,2023,Elsevier

[57] Kauppi, T., Kalesnykiene, V., Kamarainen, J. K., Lensu, L., Sorri, I., Raninen, A., ... & Pietilä, J. (2007, September). The diaretdb1 diabetic retinopathy database and evaluation protocol. In *BMVC* (Vol. 1, No. 1, p. 10).

[58] Alyoubi, W. L., Abulkhair, M. F., & Shalash, W. M. (2021). Diabetic retinopathy fundus image classification and lesions localization system using deep learning. *Sensors*, *21*(11), 3704.

[59] Feeny, A. K., Tadarati, M., Freund, D. E., Bressler, N. M., & Burlina, P. (2015). Automated segmentation of geographic atrophy of the retinal epithelium via random forests in AREDS color fundus images. *Computers in biology and medicine*, *65*, 124-136.

[60] Al-Bander, B., Williams, B. M., Al-Nuaimy, W., Al-Taee, M. A., Pratt, H., & Zheng, Y. (2018). Dense fully convolutional segmentation of the optic disc and cup in colour fundus for glaucoma diagnosis. *Symmetry*, *10*(4), 87.

[61] Al-Obadi, Mohamed Ghanim; Farhan, Hameed Mutlag; Naseri, Raghda Awad Shaban; Turkben, Ayca Kurnaz; Mustafa, Ahmed Khalid; Al-Aloosi, Ahmed Raad;

,Data mining techniques for extraction and analysis of covid-19 data,2022 International Conference on Artificial Intelligence of Things (ICAIoT),,,1-7,2022,IEEE.

[62] Diaz-Pinto, A., Morales, S., Naranjo, V., Köhler, T., Mossi, J. M., & Navea, A. (2019). CNNs for automatic glaucoma assessment using fundus images: an extensive validation. *Biomedical engineering online*, *18*, 1-19.

[63] Batista, F. J. F., Diaz-Aleman, T., Sigut, J., Alayon, S., Arnay, R., & Angel-Pereira, D. (2020). Rim-one dl: A unified retinal image database for assessing glaucoma using deep learning. *Image Analysis & Stereology*, *39*(3), 161-167.

[64] Li, L., Xu, M., Liu, H., Li, Y., Wang, X., Jiang, L., ... & Wang, N. (2019). A large-scale database and a CNN model for attention-based glaucoma detection. *IEEE transactions on medical imaging*, *39*(2), 413-424.

[65] Kamran, S. A., Saha, S., Sabbir, A. S., & Tavakkoli, A. (2019, December). Optic-net: A novel convolutional neural network for diagnosis of retinal diseases from optical tomography images. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)* (pp. 964-971). IEEE.

[66] Alsaih, K., Lemaitre, G., Rastgoo, M., Massich, J., Sidibé, D., & Meriaudeau, F. (2017). Machine learning techniques for diabetic macular edema (DME) classification on SD-OCT images. *Biomedical engineering online*, *16*, 1-12.

[67] Luo, X., Li, J., Chen, M., Yang, X., & Li, X. (2021). Ophthalmic disease detection via deep learning with a novel mixture loss function. *IEEE Journal of Biomedical and Health Informatics*, *25*(9), 3332-3339.

[68] Li, N., Li, T., Hu, C., Wang, K., & Kang, H. (2021). A benchmark of ocular disease intelligent recognition: One shot for multi-disease detection. In *Benchmarking, Measuring, and Optimizing: Third BenchCouncil International Symposium, Bench 2020, Virtual Event, November 15–16, 2020, Revised Selected Papers 3* (pp. 177-193). Springer International Publishing.

[69] Niemeijer, M., Van Ginneken, B., Cree, M. J., Mizutani, A., Quellec, G., Sánchez, C. I., ... & Abràmoff, M. D. (2009). Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. *IEEE transactions on medical*

imaging, *29*(1), 185-195.

[70] Zhang, J., Dashtbozorg, B., Bekkers, E., Pluim, J. P., Duits, R., & ter Haar Romeny, B. M. (2016). Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores. *IEEE trans*

[71] Holm, S., Russell, G., Nourrit, V., & McLoughlin, N. (2017). DR HAGIS—a fundus image database for the automatic extraction of retinal surface vessels from diabetic patients. *Journal of Medical Imaging*, *4*(1), 014503-014503.

[72] Mateen, M., Wen, J., Nasrullah, Song, S., & Huang, Z. (2018). Fundus image classification using VGG-19 architecture with PCA and SVD. *Symmetry*, *11*(1), 1.

[73] Grisan, E., Foracchia, M., & Ruggeri, A. (2008). A novel method for the automatic grading of retinal vessel tortuosity. *IEEE transactions on medical imaging*, *27*(3), 310-319.

[74] Kubicek, J., Timkovic, J., Krestanova, A., Augustynek, M., Penhaker, M., & Bryjova, I. (2018). Morphological segmentation of retinal blood vessels and consequent tortuosity extraction. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, *10*(1-4), 73-77.

[75] Tham, Y. C., Lim, S. H., Shi, Y., Chee, M. L., Zheng, Y. F., Chua, J., ... & Cheng, C. Y. (2018). Trends of visual impairment and blindness in the Singapore Chinese population over a decade. *Scientific reports*, *8*(1), 12224.

[76] Oloumi, F., Rangayyan, R. M., & Ells, A. L. (2015, August). Computer-aided diagnosis of plus disease in retinal fundus images of preterm infants via measurement of vessel tortuosity. In *2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 4338-4342). IEEE.

[77] Kashkool, Hawraa Jaafar Murad; Farhan, Hameed Mutlag; Naseri, Raghda Awad Shaban; Kurnaz, Sefer; ,"Enhancing Facial Recognition Accuracy and Efficiency Through Integrated CNN, PCA, and SVM Techniques",International Conference on Forthcoming Networks and Sustainability in the AIoT Era,,,57-70,2024,Springer Nature Switzerland Cham.

[78] Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of*

pharmaceutical and biomedical analysis, *22*(5), 717-727.

[79] Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., & Demir, I. (2020). A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology*, *82*(12), 2635-2670.

[80] Podareanu, D., Codreanu, V., van Leeuwen, G. C., & Weinberg, V. (2019). Best practice guide-deep learning. *Partnership for Advanced Computing in Europe (PRACE), Tech. Rep*, *2*.

[81] Riyam Laythkhaleel, Abdullahi Abdu Ibrahim, Raghda Awad Shaban Naseri, Hameed Mutlag Farhan, An efficient faults and attacks categorization model in IoT-based cyber physical systems using Dilated CNN and BiLSTM with multi-scale dense Attention module,Biomedical Signal Processing and Control, Volume 96, Part B, 2024,106637,ISSN 1746-8094,

[82] Yingge, H., Ali, I., & Lee, K. Y. (2020, February). Deep neural networks on chip-a survey. In *2020 IEEE International Conference on Big Data and Smart Computing (BigComp)* (pp. 589-592). IEEE.

[83] Conza, K. A. L. Deep learning for decision support in dermatology.

[84] Glorot, X., & Bengio, Y. (2010, March). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249-256). JMLR Workshop and Conference Proceedings.

[85] Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)* (pp. 807-814).

[86] Glorot, X., Bordes, A., & Bengio, Y. (2011, June). Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics* (pp. 315-323). JMLR Workshop and Conference Proceedings.

[87] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

[88] Zheng, Y., He, M., & Congdon, N. (2012). The worldwide epidemic of diabetic retinopathy. *Indian journal of ophthalmology*, *60*(5), 428.

[89] Nguyen, Q. H., Muthuraman, R., Singh, L., Sen, G., Tran, A. C., Nguyen, B. P., & Chua, M. (2020, January). Diabetic retinopathy detection using deep learning. In *Proceedings of the 4th international conference on machine learning and soft computing* (pp. 103-107).

[90] Tymchenko, B., Marchenko, P., & Spodarets, D. (2020). Deep learning approach to diabetic retinopathy detection. *arXiv preprint arXiv:2003.02261*.

[91] Pour, A. M., Seyedarabi, H., Jahromi, S. H. A., & Javadzadeh, A. (2020). Automatic detection and monitoring of diabetic retinopathy using efficient convolutional neural networks and contrast limited adaptive histogram equalization. *IEEE Access*, *8*, 136668-136673.

[92] Thota, N. B., & Reddy, D. U. (2020, August). Improving the accuracy of diabetic retinopathy severity classification with transfer learning. In *2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS)* (pp. 1003-1006). IEEE.

[93] Mushtaq, G., & Siddiqui, F. (2021, February). Detection of diabetic retinopathy using deep learning methodology. In *IOP conference series: materials science and engineering* (Vol. 1070, No. 1, p. 012049). IOP Publishing.

[94] Karki, S. S., & Kulkarni, P. (2021, March). Diabetic retinopathy classification using a combination of efficientnets. In *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)* (pp. 68-72). IEEE.

[95] Shaik, N. S., & Cherukuri, T. K. (2022). Hinge attention network: A joint model for diabetic retinopathy severity grading. *Applied Intelligence*, *52*(13), 15105-15121.

[96] Oulhadj, M., Riffi, J., Chaimae, K., Mahraz, A. M., Ahmed, B., Yahyaouy, A., ... & Tairi, H. (2022). Diabetic retinopathy prediction based on deep learning and deformable registration. *Multimedia Tools and Applications*, *81*(20), 28709-28727.

[97] Lahmar, C., & Idri, A. (2023). Deep hybrid architectures for diabetic retinopathy classification. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *11*(2), 166-184.

[98] Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., ... & Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *cell*, *172*(5), 1122-1131.