REPUBLIC OF TURKEY
ISTANBUL GELISIM UNIVERSITY
INSTITUTE OF GRADUATE STUDIES

Department of Electrical and Electronics Engineering

CNN GOOGLENET AND ALEXNET ARCHITECTURE DEEP LEARNING FOR DIABETIC RETINOPATHY IMAGE PROCESSING AND CLASSIFICATION

Master Thesis

Israa NAEEM OLEIWI AL-MAHDI

Supervisor
Asst. Prof. Dr. Sevcan KAHRAMAN

Istanbul – 2023

THESIS INTRODUCTION FORM

Name and Surname	**:** ISRAA NAEEM OLEIWI AL-MAHDI
Language of the Thesis	**:** English
Name of the Thesis	**:** CNN GOOGLENET And ALEXNET Architecture Deep Learning for Diabetic Retinopathy Image Processing and Classification
Institute	**:** Istanbul Gelisim University Institute of Graduate Studies
Department	**:** Electrical and Electronics Engineering
Thesis Type	**:** Master
Date of the Thesis	**:** 19-07-2023
Page Number	**:** 77
Thesis Supervisors	**:** Asst.Prof.Dr.SEVCAN KAHRAMAN
Index Terms	**:**
Turkish Anstract	**:**
Distribution List	**:** 1. To the Institute of Graduate Studies of Istanbul Gelisim University 2. To the National Thesis Center of YÖK (Higher Education Council)

Signature

ISRAA NAEEM OLEIWI AL-MAHDI

REPUBLIC OF TURKEY
ISTANBUL GELISIM UNIVERSITY
INSTITUTE OF GRADUATE STUDIES

Department of Electrical and Electronics Engineering

CNN GOOGLENET AND ALEXNET ARCHITECTURE DEEP LEARNING FOR DIABETIC RETINOPATHY IMAGE PROCESSING AND CLASSIFICATION

Master Thesis

Israa NAEEM OLEIWI AL-MAHDI

Supervisor
Asst. Prof. Dr. Sevcan KAHRAMAN

Istanbul – 2023

DECLARATION

I hereby declare that in the preparation of this thesis, scientific ethical rules have been followed, the works of other persons have been referenced in accordance with the scientific norms if used, there is no falsification in the used data, any part of the thesis has not been submitted to this university or any other university as another thesis.

Israa NAEEM OLEIWI AL-MAHDI

1-6-2023

TO ISTANBUL GELISIM UNIVERSITY

THE DIRECTORATE OF SOCIAL SCIENCES INSTITUTE

The thesis study of Israa Naeem Oleiwi AL-MAHDI titled as CNN Googlenet and Alexnet Architecture Deep Learning For Diabetic Retinopathy İmage Processing and Classification has been accepted as MASTER THESIS in the department of Electrical- Electronic Engineering by out jury.

Director *Signature*

Asst. Prof.Dr.Sevcan KAHRAMAN

Member *Signature*

Asst. Prof. Dr.Yusuf Gurcan SAHIN

Member *Signature*

Asst. Prof. Dr. Kenan BUYUKATAK

APPROVAL

I approve that the signatures above signatures belong to the aforementioned faculty members.

1 / 06 / 2023

Signature

Prof. Dr. Izzet GUMUS

Director of the Institute

SUMMARY

The objective of this study is to analyze the performance of the two widely used convolutional neural network (CNN) architectures, namely AlexNet and GoogLeNet, in order to determine their relative merits and demerits with regard to the classification of photographic images. The purpose of this test is to evaluate how effectively they can recognize and prepare for a variety of various kinds of pictures. Experiments and analyses include a broad variety of subjects, such as the implementation of algorithms, various pre-processing methods, the continuation of algorithms, and various strategies for fine-tuning. Investigation is conducted into a variety of training period settings, and the findings and graphs obtained from these investigations are compared. The findings indicate that both AlexNet and GoogLeNet have the capability of being utilized in the process of photo classification. AlexNet already has a performance advantage over its rivals after only six rounds of training. In the beginning, GoogleNet was not as accurate as Caffe, but it quickly caught up by engaging in enormous training repetitions. In general, the findings of the study imply that selecting the appropriate CNN architecture should be driven more by necessity than by personal taste. While AlexNet offers high accuracy with fewer epochs, GoogLeNet shows potential for higher performance with further training. This research not only helps academics and practitioners make more educated judgments when selecting models for image classification tasks, but it also increases our understanding of CNN architectures.

Key Words: AlexNet, Convolutional neural network, GoogleNet, Retinopathy diabetes.

ÖZET

Bu çalışmanın amacı, yaygın olarak kullanılan iki evrişimli sinir ağı (CNN) mimarisinin, yani AlexNet ve GoogLeNet'in performansını, fotoğrafik görüntülerin sınıflandırılmasına ilişkin göreli avantajlarını ve dezavantajlarını belirlemek için analiz etmektir. Bu testin amacı, çocukların çeşitli türden resimleri ne kadar etkili bir şekilde tanıyabildiklerini ve bunlara hazırlanabildiklerini değerlendirmektir. Deneyler ve analizler, algoritmaların uygulanması, çeşitli ön işleme yöntemleri, algoritmaların devamı ve çeşitli ince ayar stratejileri gibi çok çeşitli konuları içerir. Çeşitli eğitim periyodu ortamlarında inceleme yapılır ve bu incelemelerden elde edilen bulgular ve grafikler karşılaştırılır. Bulgular, hem AlexNet hem de GoogLeNet'in fotoğraf sınıflandırma sürecinde kullanılma yeteneğine sahip olduğunu göstermektedir. AlexNet, yalnızca altı turluk bir eğitimden sonra rakiplerine göre şimdiden bir performans avantajına sahip. Başlangıçta, GoogleNet, Caffe kadar doğru değildi, ancak muazzam eğitim tekrarlarıyla uğraşarak kısa sürede yakalandı. Genel olarak, çalışmanın bulguları, uygun CNN mimarisinin seçilmesinin kişisel zevkten çok zorunluluk tarafından yönlendirilmesi gerektiğini ima etmektedir. AlexNet daha az dönemle yüksek doğruluk sunarken, GoogLeNet daha fazla eğitimle daha yüksek performans potansiyeli gösteriyor. Bu araştırma, akademisyenlerin ve uygulayıcıların görüntü sınıflandırma görevleri için model seçerken daha eğitimli kararlar vermelerine yardımcı olmakla kalmıyor, aynı zamanda CNN mimarileri hakkındaki anlayışımızı da artırıyor.

Anahtar kelimeler: AlexNet, Konvolüsyonel sinir ağı, GoogleNet, Retinopati diyabeti.

TABLE OF CONTENTS

CHAPTER ONE
LAYOUT CHARACTERISTICS OF THE THESIS

CHAPTER TWO
DIABETIC RETINOPATHY

CHAPTER THREE
CONVOLUTIONAL NEURAL NETWORKS

CHAPTER FOUR
METHODOLOGY

CHAPTER FIVE
RESULTS AND ANALYSIS

ABBREDIVATIONS

CNN:	Convolutional Neural Network
Matlab:	Matrix Laboratory
RGB:	Red Green Blue
DR:	Diabetic Retinopathy
ADAM:	Adaptive Moment Estimation
RMSPROP:	Root Mean Square Propagation
Epoch:	Iteration
AI:	Artificial Intelligence
DL:	Deep Learning
GPU:	Graphics Processing Unit
SGD:	Stochastic Gradient Descent
API:	Application Programming Interface
GUI:	Graphical User Interface
CAD:	Computer-Aided Diagnosis
SVM:	Support Vector Machine

LIST OF TABLES

LIST OF FIGURES

PREFACE

In recent years, the exploration of photo analysis has achieved remarkable success, benefiting both academic researchers and medical professionals. This advancement has significantly improved medical diagnosis, treatment, and overall patient care. However, challenges persist, particularly in addressing diabetic retinopathy (DR), a potential consequence of diabetes that threatens vision. DR is characterized by the growth of abnormal blood vessels on the retina, leading to leakage and potentially blindness.

Overcoming these challenges, innovative diagnostic methods have emerged, enabling early detection and effective disease management. Regular check-ups have become crucial for timely intervention. During examinations, alarming findings like retinal microvasculature aneurysms, hemorrhages, and exudates can now be detected. For instance, microvascular aneurysms appear as patches of redness in the retina, while hemorrhages result from damaged blood vessels leaking blood. Exudates, on the other hand, are fluids seeping from abnormal blood arteries or dilated vessels.

This progress in medical imaging has revolutionized diabetes care and highlighted the importance of maintaining eye health. By using cutting-edge techniques, medical professionals can detect and address diabetic retinopathy at early stages, reducing its potential impact on vision and overall well-being.

.

INTRODUCTION

In recent years, researchers in academic settings as well as professionals working in the medical field have had a great deal of success investigating the process of photo analysis. In today's world, medical imaging provides assistance not only in the process of medical diagnosis but also in a wide range of other aspects of health and treatment, such as the processing and fragmentation of essential medical pictures, analyses, and transfers (Rodríguez et al., 2008).

Research into digital pictures is a part of this discipline since digital pictures have such a wide range of applications; for example, in mathematics, imaging, and the detection of diseases of interest (Nayak et al., 2009). As a direct result of recent advancements in this industry, there has been a significant and widely observable rise in the quality of treatment provided to patients.

A frequent and potentially life-threatening consequence of diabetes is called diabetic retinopathy (DR) (Singh & Tripathi, 2010). It is a disorder that affects the eyes, and it may herald the beginning of diabetes, a disease of the retina that has the potential to cause blindness (Sopharak et al., 2008). As people live longer, the risk of developing the condition increases. Diagnostics and investigations performed within the body have enabled medical professionals to vastly increase their capacity to correctly diagnose sickness, administer the proper therapy, monitor patient progress, and create innovative therapies.

Diabetes can lead to a number of serious complications, one of which is diabetic retinopathy, which is damage to the retina, located in the back of the eye. Retinopathy can lead to blindness if it is not properly diagnosed and treated, which presents significant dangers to the cardiovascular system. Retinopathy is characterized by the growth and development of abnormal blood vessels on the outer retina. This growth and development of abnormal blood vessels can induce leakage from veins inside the retina and the buildup of exudates in the retina's posterior lobe, which is a multi-layered sensitive tissue. There are a few different reasons why a person's vision could become blurry.

It is now feasible to diagnose diabetes with one of the numerous cutting-edge methods that are currently accessible. As a consequence of this, maintaining a regular checkup schedule is absolutely necessary for early detection, which in turn assists in the management of the disease's progression (Singh & Tripathi, 2010). Aneurysms of the retinal microvasculature, hemorrhage, and exudates are some of the most alarming findings that can be made during an examination.

A patient suffering from this ailment may see patches of redness in their retina. These spots are really microvascular aneurysms. Hemorrhages in the retina are caused by blood vessels in the retina that are damaged and unable to mend, which allows blood to leak out of the vessels (Sopharak et al., 2008). The term "exudate" refers to fluid that seeps out of abnormal fatty lumps that are located inside aberrant blood arteries or that seeps out of abnormally dilated blood vessels. Fovea The area in the middle of the retina is where people have the clearest eyesight.

CHAPTER ONE

LAYOUT CHARACTERISTICS OF THE THESIS

The diagnostic procedures that are now in use for retinopathy are time-consuming, challenging to get, and expensive. The fast progression of diabetic retinopathy is a major and immediate cause for worry in terms of health care. The astonishing number of people who are now affected with this condition is only expected to increase in the future, according to projections made by the World Health Organization (WHO). The fact that many of them are unaware that they are infected is, on the other hand, what poses the greatest risk. Researchers estimate that there will be 642 million people living with diabetes by the year 2040 (Wong & Sabanayagam, 2020). There is a risk that one in every three of them will develop retinopathy. Figure 1 is a representation of the anatomy of the retina, and Figure 2 is a representation of both normal and abnormal versions of the retina.

Using a combination of several medical imaging modalities, retinopathy may be identified in all of its many stages and stages of progression, as well as graded. Imaging may be performed in a number of different ways to identify these conditions in patients, but the great majority of specialists in the medical field believe that imaging is the method that is both the most reliable and the one that is utilized the most frequently.

Imaging procedures are utilized in order to identify retinopathy and categorize cases of it according to their degree of severity. Imaging is one of the numerous methods that may be used to diagnose such diseases in humans. It is generally agreed upon that imaging is the method that is the most exhaustive and comprehensive. The identification of retinal disease is highly dependent on a correct diagnosis.

The authors of this thesis propose a machine learning model to assist in the early detection of the stages of the disease and the identification of retinopathy disease based on the segmentation and classification of the disease according to symptoms and degrees according to images. This will allow an ophthalmologist to immediately show whether or not this person is infected on the same day. This will allow an ophthalmologist to immediately show whether or not this person is infected. The purpose of this thesis is to offer a model for the use of machine learning that will assist in the early diagnosis of the phases of the disease as well as the identification of its causes. As a consequence of this, it is of the utmost importance to design a method that shortens the time of the examination, lowers the out-of-pocket expenditures for the patient, and lowers the proportion of patients who have anxiety and the condition in their bodies while they are undergoing the testing procedure.

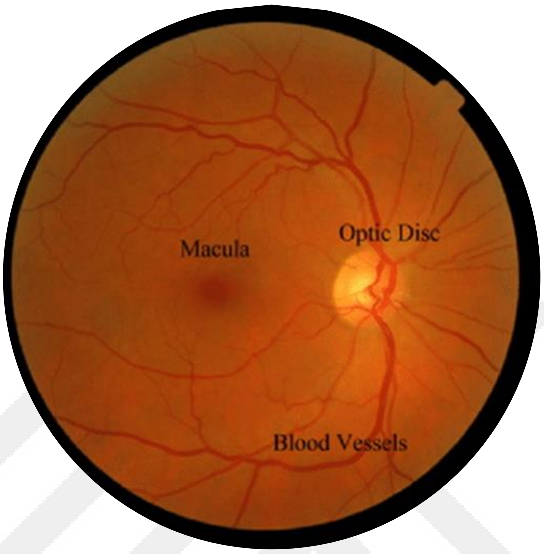

Figure 1. Retina (DIAGNOSIS RETINOPATHY DISEASE USING GLCM AND ANN, 2018)

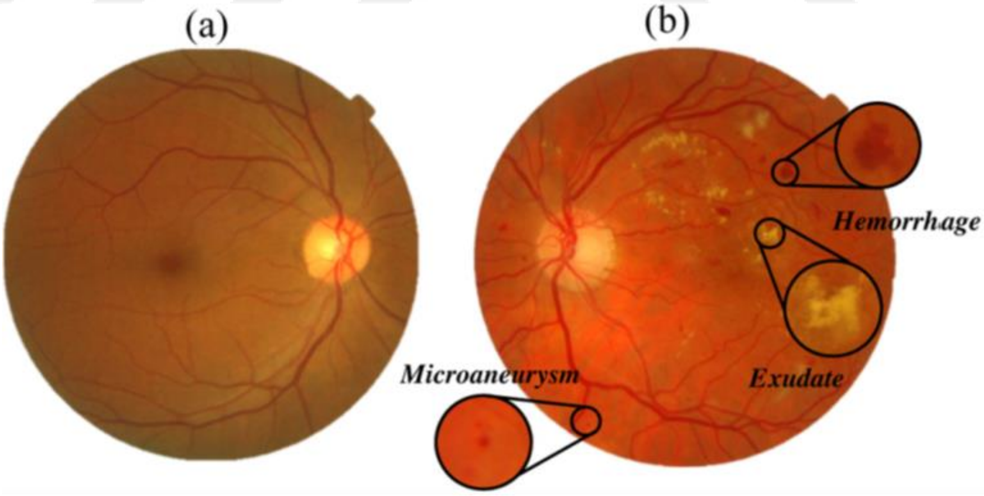

Figure 2. Retina (a) No DR (b) Retina contains DR (DIAGNOSIS RETINOPATHY DISEASE USING GLCM AND ANN, 2018)

1.1. Background

One of the most serious and potentially blinding effects of diabetes is called diabetic retinopathy. Therefore, an early diagnosis is vital if one is interested in both slowing the advancement of the disease and be supportive for the blind people. As a result, a process needs to be carried out once every six months in order to check and understand the progression of the disease (Wong & Sabanayagam, 2020).

In order for the ophthalmologist to efficiently detect and classify fundus images, as well as significantly reduce the risk of blindness caused by DR, an effective algorithm is essential. Therefore, researchers began searching for and developing various methods of image analysis that may assist in the diagnosis of retinopathy.

Within the realm of contemporary image processing, one has access to a dizzying variety of techniques. In addition, artificial intelligence (AI), specialized hardware, and algorithms may all make use of black-and-white and color photographs to extract features that reveal a normal retina and a dysfunctional retina, and then classify the latter into a variety of different illnesses. A concise overview of the pertinent past research will be offered before going into the intricacies of the system, the methodology, and the stages of development, respectively.

By highlighting the presence of secretions in images of the retina that had been processed using image processing software, Franklin et al. (Franklin & Rajan, 2014) were able to identify diabetic retinopathy. The pronounced tone shifts that take place as a direct consequence of these secretions may be observed without any effort. In order to perform this objective, an artificial neural network is trained to generate estimations based on the dimensions, colors, and appearances as well as the textures of reference photographs that were given by medical professionals.

Banerjee and colleagues present a method in the article (S. Banerjee & Chowdhury, 2015) that integrates contextual data with photographs taken from a variety of databases. As a result of this, decision trees that incorporate visuals and contextual information have been developed. The photographs had to be split and preprocessed so that the essential components could be extracted and used to illustrate the various diseases. The hashed photographs from the prototype is directly matched with the filtered hash images since the information contained inside the associated contextual information was quite similar.

A method for identifying and diagnosing DR was presented by Manoj Kumar et al. (Jahiruzzaman & Hossain, 2016) in the form of the k-mean methodology. In which they demonstrated how to segment pictures using the pillar k-means algorithm. After using the pillar approach, the improved k-means algorithm was used to the data in order to further refine the findings. Their recommended technique has the potential to shorten the processing time required for image segmentation using k-means while preserving a high level of accuracy. In each of the several experiments, a unique color space was utilized. Experiments showed that segmenting photos produced sharper images with improved color separation. As a consequence of this, they were successful in achieving even faster computing speeds than k-means while still providing high-quality discoveries.

A Convolutional Neural Network (CNN) technology has been developed by Pratt et al. (Pratt et al., 2016) with the intention of offering an easy automated diagnostic. This technology has been used to classify the five stages of retinopathy. Nevertheless, the findings were not adequate. As a result of the fact that the model was unable to correctly classify the initial stages. In addition to this, they gained an edge by employing a biased data set to their benefit by striking a balance between high specificity and low sensitivity. Additionally, a number of other Ensemble approaches have been suggested.

CNN was used by Gondal et al. (Gondal et al., 2017), who trained their model on the Kaggle data set and tested it on the DiaretDB1 database. CNN was applied by these researchers. A normal stage as well as six more normal stages of non-referred retinopathy were differentiated from the severe phases that were also discovered in this study. In order to classify the aforementioned information, a convolutional neural network was utilized. DiaretDB1 achieved an area under the curve of 95.4.

Multi-Label classification (MLC), an essential idea in the field of computer vision, was utilized by Abdelmaksoud et al. (Abdelmaksoud et al., 2021) for their research. Retinopathy (DR) of varied degrees may be recognized in medical photographs by employing (MLC) features, and the datasets used to do so come from DRIVE, CHASEDB1, STARE, HRF, IDRiD, DIARETDB1, and MESSIDOR respectively. Therefore, ophthalmologists are able to recognize varied degrees of retinopathy, which enables patients to seek out the most effective therapy available and avoid the risk of having their retinal worsen. In their investigation, the researchers made use of a system that is primarily dependent on "deep learning" computer vision. They came to this conclusion after seeing that the progression of DR is linked to a variety of retinal degenerative illnesses. With the use of the ML-CAD technology that is currently being developed, ophthalmologists are able to visualize varied degrees of DR and make an accurate diagnosis of it. The technique is able to differentiate between healthy people and those suffering from DR. Because it makes use of deep learning technology, the system is able to automatically extract four distinct diseases (U-Net). After this, the GLCM approach is applied in order to get rid of any residual symbols. After that, the SVM will classify these characteristics into separate DR stages. The strategy that was recommended achieved a rate of accuracy of 95.1%, sensitivity of 86.1%, and specificity of 86.8%.

1.2. Problem Statement

The correct diagnosis and the planning of the appropriate treatment are dependent on accurately classifying and segmenting the retinopathy. In spite of this, the standard examination may need to be repeated so that the ophthalmologist can establish an accurate diagnosis. This can add additional time to the procedure (up to a few weeks), which is necessary due to the numerous characteristics of retinopathy. Despite this, the standard examination is still recommended (such as color, size, and texture). In addition, a diagnosis must be performed by an ophthalmologist since the symptoms of many eye diseases can be quite similar, which increases the risk of making an incorrect diagnosis. A variety of extrinsic factors can also have an impact on the precision with which a physician makes a diagnosis. A number of factors come into play, including the findings of studies, one's own perceptions, and the characteristics that are unique to the eye. This paradigm is especially beneficial for younger physicians who have less experience in their field.

1.3. Aim

This study aims to improve the quality of computer-aided medical picture segmentation and classification, which is of great benefit to patients who are dealing with long-term conditions. This will be accomplished by developing a tool that uses the convolutional neural network to assist in the diagnosis of retinopathy.

1.4. Importance of the study

The early diagnosis of these diseases, which is made feasible by the use of computer vision to medical imaging, is extremely beneficial. It is absolutely necessary in order to get high-quality photographs and receive correct medical diagnosis.

In addition, these methods assist physicians in obtaining information with a high degree of accuracy, which in turn assists physicians in making a diagnosis and treatment plan in a straightforward manner, without having to rely on possibilities; in this case, the patient receives the best possible care from the physician.

Keep in mind that because to the widespread nature of DR, healthcare institutions will be required to pay more money in order to treat patients. This is due to the fact that conventional systemic treatments take a considerable amount of time to take effect, and due to the fact that the illness is chronic, these treatments require continual monitoring.

If a doctor makes an incorrect assessment of a patient's condition after making a mistake in diagnosing their illness, the patient's life might be in danger. This strategy, which will be described in greater depth in the third chapter, reduces the anxiety that is brought on by the traditional way of diagnosis for patients who fall into the average category.

1.5. Objectives

In cases where a desired outcome was realized,

- The technology mimics the actions of eye doctors.
- The technology is more effective than an amateur ophthalmologist and takes less time to use.
- Gathering and sifting through information about the retina and picking the best of the bunch.
- The creation of a retinopathy segmentation technique for distinguishing between healthy and diseased retinal fundus tissues.
- Construct a system for classifying and grading retinopathies.

- Conducting a number of trials on a predetermined data set to validate the classification and segmentation procedure.
- Conducting a comparison of the proposed patterns with the relevant research in order to learn about and assess the outcomes.

1.6. Outline

This thesis will be organized as following:

- Introduction: This section described the study's overarching goals and provided context for the difficulties, obstacles, and endpoints that were faced along the way.
- Literature review: In this chapter, context for how machine learning techniques may be applied will be provided to accomplish the study's goals. Main achievements from prior research based on both time series and machine learning approaches will be presented and debated.
- Methodology: In this third chapter, deep learning will be used to examine how well the prediction model works. A detailed presentation of the mathematical theory behind the statistical models is planned. This section will detail the data, data processing, feature engineering, and performance metrics. Methodologies for each individual model will be included in the final presentation.
- Results and analysis: In this section, the obtained results are displayed. Results will be analyzed and discussed. There will also be a detailed breakdown of the study's advantages and disadvantages, as well as suggestions on how things may be improved.
- Conclusion: The difficulties of project management and the part that each stakeholder plays in it will be displayed. Findings and suggestions will be compiled when the research is finished.

CHAPTER TWO

DIABETIC RETINOPATHY

The progression of diseases through the different stages of retinopathy is one of the theoretical aspects that will be covered in this chapter. A description will be given of the work's prospective database use. This chapter will go over aspects relating to methodology, such as the application of image processing (which includes picture improvement, numerous optimizations, diagnostics, and classification methods).

2.1. Diabetic Retinopathy

Diabetes is a potentially life-threatening disorder that affects humans. Complications from diabetes can cause damage to the retina of the eye, which can result in vision loss (Klein, 2007).

Vision loss caused by diabetes can be prevented if the disease is detected and treated in its early stages. A leading sign of progressive retinopathy at an early stage, retinal microaneurysms can be seen in the retina. If the aneurysm continues to grow, the patient might not be aware that they are at risk for consequences; nevertheless, if the aneurysm continues to grow. DR Fundus photographs are inspected by ophthalmologists, and the aneurysm is detected and monitored manually. This implies that there is opportunity for substantial error during the inspection, as well as that the process is time intensive and repetitious. When identified at an earlier stage, DR is less likely to progress, which results in a slower loss of vision (Rohan et al., 1989). Early diagnosis of diabetic retinopathy (DR) involves a process that involves monitoring the patient and keeping him under control until the disease is under control. This is done in order to avoid the disease from advancing to a hazardous stage on the retina.

Analyzing photographs of the retinal fundus enables medical professionals to identify a range of conditions, including various stages of DR. A number of computer vision algorithms are utilized in the analysis of these photographs.

However, when blood glucose levels in the retina jump to dangerously high levels, the blood vessels in the retina sustain significant damage. As glucose levels in the blood grow, the blood vessels become more prone to perforation, which in turn impairs the visual system and causes blood to flow into the eye. This condition is known as diabetic retinopathy. This leads to a defect, a major damage, and effects that might potentially be lethal.

It's very amazing that the bodies have the ability to naturally fix themselves. If neighboring brain cells detect a loss of blood, the brain will send messages to those cells, instructing them to take action and organize themselves into a group. By carrying out these steps, you can cause the abnormal creation of new blood vessels (Hajeb Mohammad Alipour et al., 2012).

Because the newly formed blood vessels are fragile and have the potential to cause vision loss, it is essential for diabetic patients to have regular and frequent exams of the retina. This has already been covered in the previous section. Therefore, it is necessary for the ophthalmologist to maintain a careful eye on the retina and to examine it on a regular basis. Fundus fluorescein angiography (FFA), optical coherence tomography (OCT), fundus photography, and slit lamp biomicroscope are some of the several forms of eye examinations that may be performed to diagnose the disease in its initial stages.

2.2. Diabetic Retinopathy Types

This is an essential step because diseases that affect the retina need to be recognized for what they are before therapy can be initiated. It is possible for the terminology to become unclear in situations in which many subtypes are recognized, such as microsystems, blood vessels, hemorrhages, hard exudates, and soft exudates.

The presence of anomalies and the degree to which they manifest are major factors in determining the severity of these illnesses. The first and most important step in treating eye diseases is correctly diagnosing them as abnormalities of the retina. In situations in which a large number of categories are recognized, any and all elements of the diseases in question are discovered and labeled. Retinal photographs are scrutinized by ophthalmologists for the presence of a wide variety of abnormalities. Exudates, hemorrhages, and the constriction of blood vessels and arteries are all examples of this phenomenon (Figueiredo et al., 2015).

2.2.1. Soft Exudates

When retinopathy has progressed to a more advanced level, soft exudates can be detected in the retina. When there is a reduction in the amount of blood that is able to reach nerve fibers, the fibers will enlarge and exude a fluid called exudates (Basha & Prasad, 2008). Cotton-wool spots are an example of a form of exudate known as a soft exudate. As a result, you won't see them until the latter stages of retinopathy. The cotton-wool patches that develop inside this region of the retina induce symptoms that are both rapid and severe (Ashraf et al., 2015). This stage is the most hazardous because of this.

Retinal infarction leads to the occlusion of microscopic blood vessels, which results in the formation of soft exudates (Blankenship & Skyler, 1978). If this happens, the vein that is affected could swell up like a sausage that has been stretched. When searching for exudates, the severity of the sickness is taken into consideration because the size and location of these fluid discharges might vary.

Because of their differing anatomies, hard exudates and soft exudates may be differentiated from one another by checking for telltale differences in size, color, shape, borders, and the human retina. This can be done by comparing the two types of exudates. Figure 3 provides visual representations of soft exudates and may be seen below.

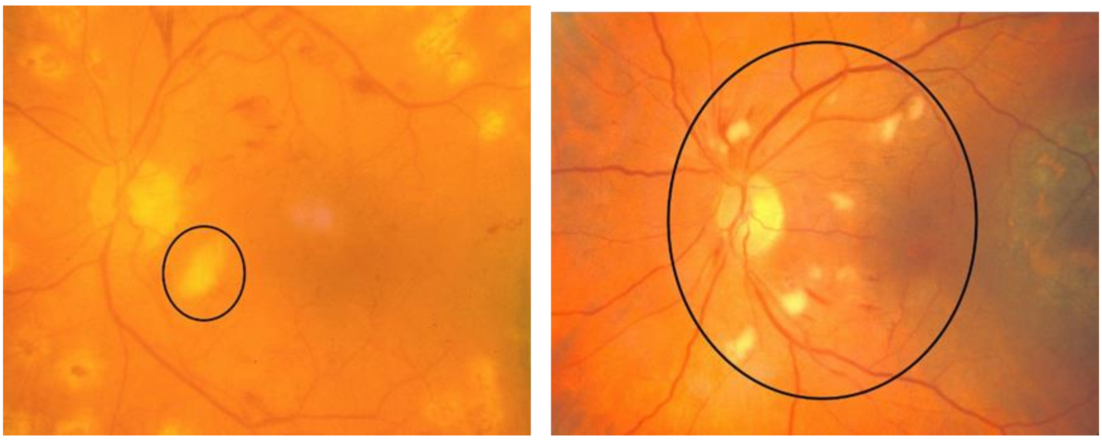

Figure 3. soft exudates (Cotton Wool Spots. COMS Grading Scheme, n.d.)

2.2.2. Hard Exudates

They are fatty deposits in the retina that can be confused for lesions, and they also consist of protein deposits that develop yellow with time. Additionally, they can be mistaken for lesions. They manifest in the fundus, also known as the posterior segment of the eye, and are characterized by an uneven area that has a glossy surface. As a consequence of this, light has a significantly more difficult time reaching the retina, which results in a blurring of vision (Franklin & Rajan, 2014). Leakage occurs as a result of the thin walls of the MAs, which raise the pressure force. A key source for the diagnosis of macular edema is exudate, which also includes fluid and increased vascular permeability. The secretions have a significant role in determining the diagnosis of retinopathy. They may be responsible for a considerable number of cases of blindness that have been documented. The size of the spots that appear on the retina can range from very large to very small, and they may or may not have clear borders, depending on the nature and severity of the condition (Aquino-Brítez et al., 2022). Hard exudates are depicted in the accompanying picture (Figure 4).

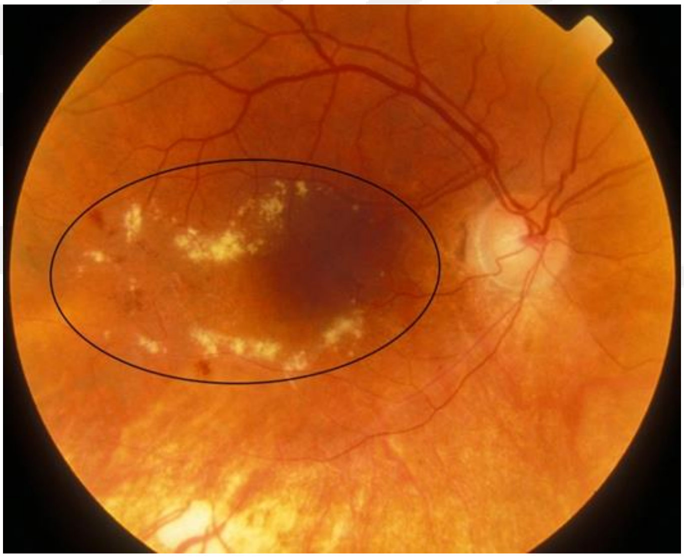

Figure 4. Hard exudates (Hard Exudates. COMS Grading, n.d.)

2.2.3. Hemorrhages

This process takes place inside the retina's innermost layers and is characterized by its creation in a round shape. A blot is the result of blood loss and the outward movement of blood from inside the vasculature's walls. Blood loss and the outward movement of blood from inside the vasculature's walls are what cause a blot. The base of its retina is tinted red, and its skeleton has an asymmetrical shape overall. Hemorrhages in the retina can take place in either the outer or the inner layers, and they often have a circular appearance and the appearance of little blood spots (MAs).

It is the second most prevalent cause of permanent blindness in people who have been diagnosed with retinopathy. Intraretinal hemorrhage is a symptom of serious retinopathy due to this reason. The severity of the illness can be estimated based on the characteristics of the hemorrhages, including their appearance, size, and number. Hemorrhage Control Achieved Through Earlier Automatic Detection (Jitpakdee et al., 2012). A flame-like appearance can be seen on the layer of nerve fibers where the bleeding has manifested (Basha & Prasad, 2008). Even among these professionals, there is still some disagreement as to whether the red flare is the consequence of microaneurysms or spot hemorrhages. This condition is seen in the picture labeled Figure 5.

Figure 5. Hemorrhages (Retinal Hemorrhage. COMS Grading, n.d.)

2.2.4. Microaneurysms

Microaneurysms (MAs) are often the first anomalies to be discovered in the retina during clinical examinations (Aquino-Brítez et al., 2022). The affected, light-sensitive retina may exhibit evidence of microscopic blood vessels or minute, dark red dots, either on their own or in groups (Welikala et al., 2015). Diabetic retinopathy, often known as DR, is the most common form of diabetic retinopathy. It is the leading cause of vision loss and has to be diagnosed as early as possible to prevent total blindness (Sidibé et al., 2015). Microaneurysms are generally spherical in shape, and their diameters range from 10 to 100 microns, which is approximately 1/12 of the average diameter of the optic disc. At this point, there are no indications of a serious illness (Basha & Prasad, 2008). Microaneurysms are depicted as tiny red dots in Figure 6. These dots commonly group together to form bigger, lumpier aggregates.

Figure 6. Microaneurysms (Microaneurysms | Retina Vitreous Resource Center, n.d.)

2.3. Degrees Of Diabetic Retinopathy (Grades)

This article will discuss the many stages (or degrees) of retinopathy that can occur in patients who are affected by this ailment. One can differentiate between proliferative (PDR) and non-proliferative (NPDR) retinopathies (PonniBala & Vijayachitra, 2012). PDR stands for proliferative diabetic retinopathy. The leaking of microscopic blood vessels causes the expansion of the retina that occurs during the non-proliferative stage of the disease. To put it another way, retinal edema is the major cause of blindness in people who have been diagnosed with retinopathy. PDR occurs when the disease has progressed to a point where the retina forms new blood vessels, and these blood vessels grow as a result of bleeding, causing a major defect and blocking the patient's vision (mahmud & Bhattacharjee, 2020). At this point, diabetic retinopathy is considered to be harmful because PDR occurs when the disease has progressed to a point where the retina forms new blood vessels, and these blood vessels grow as a result of bleeding. A comprehensive examination of the stages and severity levels of diabetic retinopathy is presented in Figure 7. DR can be broken down into five distinct phases.

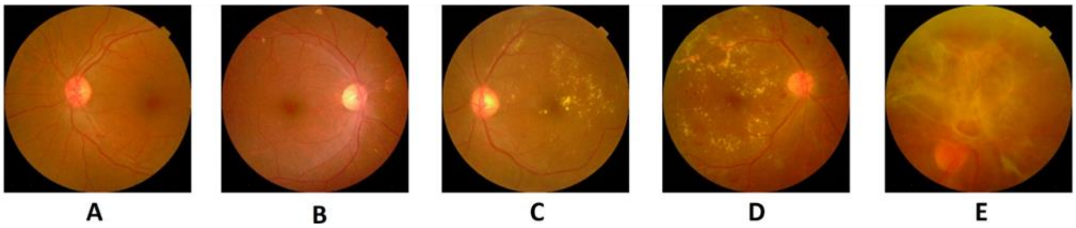

Figure 7. human retinopathy stages (grades) Typical, Mild, Moderate, Severe, and Proliferative (Ramasamy et al., 2021)

2.3.1. Proliferative

The most dangerous stage of the disease is when the blood vessels are unable to saturate with blood, which leads to the proliferation of harmful and dangerous blood vessels on the inner roof of the retina. This, in conjunction with the formation of a somewhat solid, watery liquid (glassy), jelly-like substance in the middle of the eye, marks the most dangerous stage of the disease. Newly formed aberrant blood vessels are notorious for their propensity to leak due to their delicate nature. The severity of vision loss can be directly correlated to the amount of blood that unusually thin blood vessels are able to bleed (Giancardo, 2011). If the surrounding tissue that is associated with the scar recedes and shrinks, there is a risk of retinal detachment occurring. In this scenario, the detachment of the retina results in significant and permanent damage to the patient's vision (Li & Li, 2013). The evolution of vision loss as a result of a ruptured blood artery is seen in Figure 8.

Due to this reason, obtaining an early diagnosis of the condition is crucial to prevent vision loss and initiate timely treatment. Consequently, the ongoing testing activities hold significant importance.

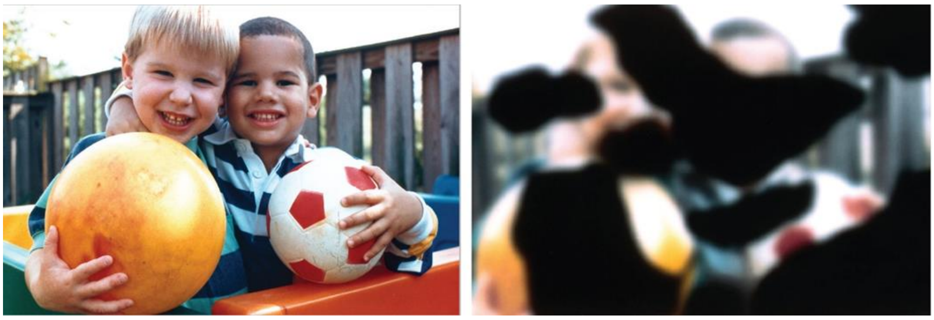

Figure 8. Compare the normal retina on the left to the retina with DR on the right (Right) (Diabetes and Vision Loss | Diabetes | CDC, n.d.)

2.3.2. Normal

As a direct consequence of this, the condition referred to as diabetic retinopathy does not exist any longer. To this day, the understanding of this illness is quite limited. The patient may finally take it easy now that the illness is no longer a danger to them. However, in order to prevent more issues, he has to work with his physician to develop a schedule for regular checkups.

2.3.3. Mild

Ophthalmologists have a tough time diagnosing diabetic retinopathy since the disease causes tiny balloon-like patches within the retina to become inflated and considerably extended. The initial indicator of the problem is a little deformity that originates from the aneurysm. This deformity is the first manifestation of the sickness that may be seen. This disease, known as "capillary leakage," gets its name from the fluids that leak out of the blood vessels and into the retina as the blood vessels widen (Deep Learning for Diabetic Retinopathy Diagnosis & Analysis, n.d.).

2.3.4. Moderate

Capillary leakage of blood and liquid exudates also occur in this instance, and as the disease progresses to a more advanced condition, the blood vessels expand and distort, creating abnormalities in the blood vessels that supply the retina. Capillary leakage of blood and liquid exudates also occur in this situation. There is a possibility that some infected people will lose their ability to give blood (Deep Learning for Diabetic Retinopathy Diagnosis & Analysis, n.d.). The macula has been enlarged, which has resulted in severe changes to the form of the retina.

2.3.5. Severe

At this stage of retinopathy, the severity of the sickness prevents blood from moving through or entering many of the retina's blood vessels, which prevents blood from reaching various sections of the retina. When this occurs, the retina at the back of the eye begins to secrete new blood vessels, which are later determined to be abnormal due to the peculiar look they have. Insufficiency of the capillaries is the name given to this ailment. As a result, a person's vision becomes blurry when the damaged retina, which is brought on by the obstruction of blood vessels, gives rise to the creation and proliferation of dangerous new blood vessels, which all takes place at the back of the retina (Li & Li, 2013). This is what causes the blurring of a person's vision.

2.4. Medical Image

As a result of developments in medical imaging, a clinician is now able to make a correct diagnosis of a significant number of disorders. This is an undeniably significant advancement in a world where the manual diagnosis favored by doctors—which relied on the patient's own words and observations—has been largely rendered obsolete as a result of the advent of medical imaging technology. In other words, this is an advancement that cannot be denied. In the most recent decades, the use of medical imaging technology has been increasingly common among physicians as a result of its numerous advantages, high level of precision, and rapidity in identifying the visible components that need to be identified.

Emerging applications in medical imaging also contribute to the fight against illness by allowing for earlier disease identification, as well as surgery, treatment, and recovery. Medical imaging is an essential component of biological imaging procedures because it utilizes a wide range of imaging modalities, including x-ray imaging, medical ultrasound imaging, magnetic resonance imaging, thermal imaging, endoscopic imaging, and a great number of other imaging modalities as well (Bankman, 2009).

Medical imaging has evolved as a major solution for complex diagnostic difficulties in recent years, particularly when it comes to the non-invasive imaging of the internal structures of the human body. Medical photos are usually fuzzy and of poor resolution because of concerns with color interference and a considerable quantity of noise, in addition to other factors that influence imaging techniques to collect data. These factors all have an impact on the creation of images. For example, due to color interference, increased noise, and blurred edges, directly analyzing images of the retina can be a tough task. Because of this, even if the patient's eyesight improves, it must first be evaluated by an ophthalmologist before it can be deemed to be sufficient. It is reasonable to assume that the significance of medical photographs has increased in recent years due to the fact that computers can now process images in such a way as to make them more usable and clearer.

2.5. Database Description

This database is devoted to pictures of the retinal fundus, and its primary purpose is to facilitate study and comparison. The purpose of this computer-assisted research is to assist in the diagnosis of diabetic retinopathy. Images of the retina make up the great bulk of the 1200 that are contained in this database (Chetoui et al., 2018; Messidor - ADCIS, n.d.). This database is essential because it enables ophthalmologists to practice diagnosing DR of all types and intensities using its extensive collection of retinal photos. This makes the database extremely useful. As a result, it is no longer necessary for patients to have the ability to recognize DR and determine whether or not their state is normal. Using this database, several photographs were examined to determine the various forms and degrees of retinopathy that were present.

In order to acquire retinal pictures for the MESSIDOR (Messidor - ADCIS, n.d.) Imaging Data Project, a 3CCD video camera was utilized. These images were subsequently examined by the Ophthalmology Department to ensure that they were suitable for use in research.

2.6. Flaw Correction In Imaging Pre-Processing

The most recent strategies for image optimization are utilized to sharpen and prepare photographs for processing actions that increase the pictures' overall aesthetic appeal to the observer. In certain contexts, the pictures do not have the required level of clarity, and in order to carry out operations on them, it is necessary to have a high level of precision in the clarity of the image. As a result, the techniques that are currently being used to enhance images are indispensable in certain fields of business.

Examples of this would include things like remote prediction and vision, as well as a variety of tasks that involve the processing of pictures and medical image analysis. Making the required modifications to the image, which may involve scaling it down, scaling it up, or repairing it, may result in an improvement to the image. Picture optimization is the process of improving an image's usefulness so that its qualities may be detected more quickly (Suganya et al., 2013). This is accomplished without adding additional data to the image. Within the realm of optimization, they are working toward a pair of goals (Marques, 2011).

The high resolution of the picture that is being displayed more clearly is related to the first goal, which aims to improve readability. In order to achieve the second goal, the image will be evaluated and edited in such a way as to separate its component parts. One further technique for bettering images is the employment of algorithms, which may be put to work to reduce noise, alter contrast, and accentuate image edges and features. Incorrect application of these techniques to photographs can lead to the loss of gradable information, an increase in background noise, and blurred image boundaries, all of which make it more difficult to locate regions in need of grading and more readily apparent flaws in the original image (Bankman, 2009). Incorrect application of these techniques can also result in the blurring of image boundaries.

Users are obligated to pay great attention and proceed with caution in this area to avoid experiencing any unforeseen repercussions while modifying images. The focus of the present work for this thesis is mostly on classification; therefore, the quality of the enhancement and processing processes used on images must be of an extremely high standard. Identification of retinal problems and grading of illnesses can now be done with more precision, and the quality of the visuals observed can be significantly improved by employing appropriate filters and algorithms. The image processing pipeline encompasses various stages, including segmentation, feature extraction, and classification.

2.6.1. An Equalization of The Histogram That Is Restricted By Contrast (CLAHE)

The CLAHE technique was at first developed for the purpose of improving the clarity and accuracy of medical images; however, it is currently being utilized in a broad range of other situations as well. This technique has the potential to identify minute parts that are in the shape of tiles (Warrier & Viji, 2021), despite the fact that it does not describe all of the regions that are captured in the images.

During the process of picture optimization, the contrast is decreased in specific regions by mixing neighboring tiles using the interpolation method (Suganya et al., 2013). This is done to avoid the image from becoming blurry.

When this occurs, aspects of the picture that were previously obscured become more noticeable, and the contrast between those aspects and the initial version is accentuated. Therefore, the CLAHE technology converts the picture values into other values to obtain a smoother, clearer image that has extra valuable aspects. This makes it possible to deduce a vast amount of information on the illnesses and their severity levels.

2.6.2. Change The Photo to Grayscale

The process of converting color images to grayscale images is an important process because it is easier to detect grayscale images in these areas than it is to identify injury areas in color images due to chromatic affinity. This is because the color images have similar colors, which can sometimes prevent the identification of injury areas in the images of the retina. Images that use a color space that only contains the primary colors red, green, and blue can be processed more slowly than those that use a grayscale.

2.6.3. Filtering Image Content While Taking Photographs

A more polished end result may be achieved by using this technique, which involves removing components that are distracting, as well as lowering noise and distortion. Without warping the image's boundaries, in order to ensure that the finished product seems as authentic as feasible. Research on medical imaging is required in order to perform a method that filters images. This is necessary in order to provide a filtered picture that is free of distortion and is more appealing and clearer to the viewer. Image filtering is used in a variety of important fields, and it is accomplished with specialized software and hardware (He et al., 2013). Because the activity of the filter is reliant on the pixels of the image as well as what is nearby, the picture is filtered and smoothed with a surprising degree of accuracy owing to a matrix. This allows the filter to work more effectively.

CHAPTER THREE

CONVOLUTIONAL NEURAL NETWORKS

3.1. The Technology Of Convolutional Neural Networks

This neural network uses deep learning technology to imitate the same processes that the different layers of the brain do in the actual comprehension that the human mind is capable of. Its configuration has been optimized for detecting and monitoring biological processes, and it has been designed to detect and monitor biological processes (He et al., 2013).

Deep neural networks (DNNs) and convolutional neural networks (CNNs) are not the same in terms of the information that is learned by each type of network, how the networks are formed, or how they are structured. It might be considered a kind of network data due to the fact that it manages information such as picture data. Since data is what it deals with, the network is pixelated and may exist in either two or three dimensions.

Due to the fact that CNN is comprised of computer programs, it is able to classify and identify objects according to the direction in which they are being directed. CNN could make identifications based on certain features found in the data or the graphics. The strong capacity of CNN approaches to simulate key phenomena is the primary source of the benefits that these approaches offer in comparison to other methods and techniques. CNN is able to recognize sickness and classify it into a variety of phases and degrees by making use of the characteristics of medical photos (O'Shea & Nash, 2015; Shanthi & Sabeenian, 2019).

In order to perform its primary function of data processing, a convolutional neural network, also known as CNN, is dependent on the convolution process. A convolutional neural network, often known as CNN, is a type of neural architecture that is specifically designed to extract features from digital images. It consists of hidden layers and an output layer.

CNN views a picture as a three-dimensional matrix that is composed of color pixels (width, height, and depth) that each have a value that ranges from 0 to 1. (0 -

255) If the picture that is being entered has color, then that color will be distributed in both the vertical and horizontal directions. On the other hand, just one-color channel is employed for creating monochromatic photos, which results in grayscale.

CNN is able to successfully categorize information despite the complexity of its underlying layers due to the superior architecture of its system, which is composed of numerous layers. It is made up of one pooling layer, one convolution layer, and one fully connected layer—all of which are hidden layers—and its output may be referred to as either an activation or a feature map, depending on whether or not it is utilized as an input to another layer (Litjens et al., 2017). Additionally, the first layer has a wide variety of filters that may be applied to a picture in order to highlight particular aspects of it, such as blood vessels, contours, or boundaries. Therefore, because to the advanced algorithms that it employs, CNN is able to recognize the same sorts of patterns and edges that the human brain is capable of doing.

Due to the fact that this is the situation, the operational notion of these layers may be described in greater depth by first discussing them individually and then showing how they operate together. Within the CNN, there are layers that are responsible for important activities that are carried out by the neural network. The utilization of this network contributes to the maintenance of lower overall operating expenses. Because the hidden layers are so important to the neurons that make up the CNN, any alterations that are made to them have the potential to have a large effect on the overall performance of the network. A convolutional neural network, often known as a CNN, has minuscule units within each neuron that are solely devoted to doing mathematical calculations and putting those calculations into action. Its major role is to act as a messenger, transmitting information to and from other neurons (Shaban et al., 2020). The illustration of these concealed layers may be found in Figure 9.

The first layer of a convolutional neural network (CNN) is responsible for handling input processes, and the succeeding layers prime themselves for output processes. At the point where the neurons activate and convert the input into output (also in three dimensions), a CNN is structured in three dimensions (length, width, and depth). As a direct consequence of this procedure, Figure 10 illustrates it.

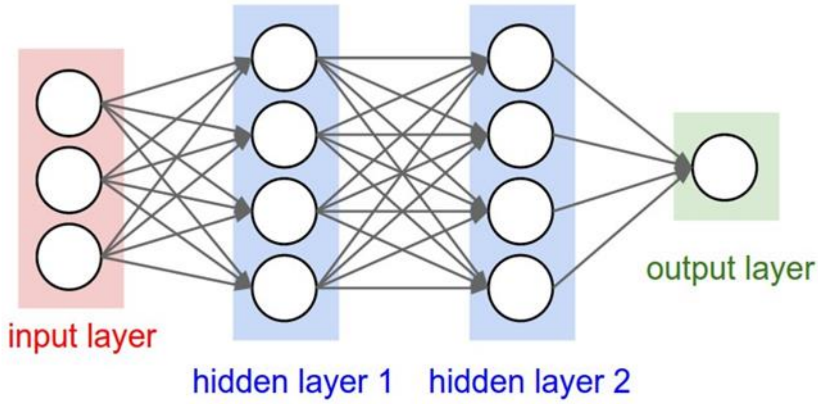

Figure 9. CNNs hidden layer (Bengio et al., 2012)

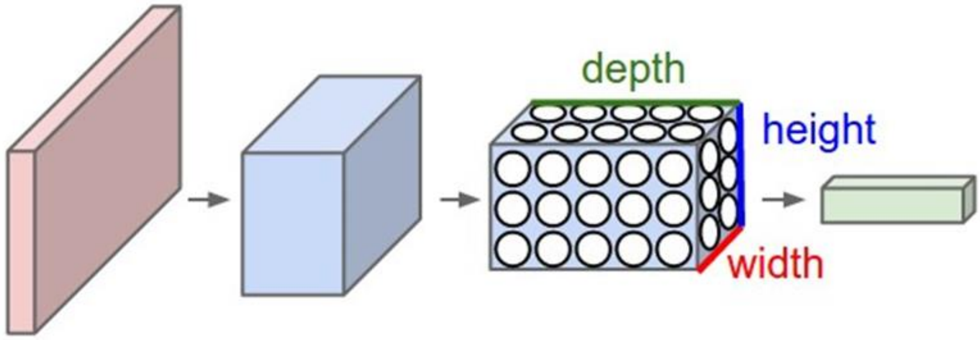

Figure 10. 3D CNN (Bengio et al., 2012)

3.1.1. Convolutional Neural Network Architecture

CNN employs a structure that is made up of many layers in order to perform its analysis on an input image. These layers are connected to one another, and as an example, the first layer starts by inserting an image into it by extracting its properties, the second layer starts assembling (pooling), and the final layer (classifier) starts to shrink the outputs. Figure 11 provides a visual representation of the fundamental components that make up a CNN.

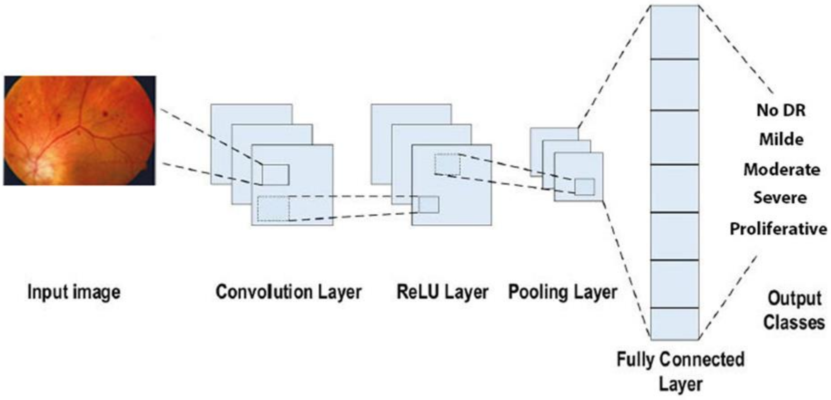

Figure 11. CNN Architecture example

3.1.2. Convolutional Layer

The convolutional layer is regarded as the most important component of a CNN (Flickner et al., 1995), which is a consensus held by the majority of experts. It does this by stacking layers on top of each other and using the findings from these stacked layers as the basis for its hypothesis in order to extract more specific traits. The neural networks' convolutional layer is not connected to the input picture; rather, the whole correlation layer is in continuous contact with the pixel in the receiving field. This layer is responsible for the beginning of the feature processing and mostly provides assistance with calculating and locating features. CNN is able to remove image features effectively thanks to its assessment of the picture using its sliding window convolutional feature. This evaluation begins with image optimization, which is why it is so effective.

The design of CNN, in which neurons in successive convolutional layers are connected to one another, enables it to prioritize low-level characteristics in the first hidden layer before beginning to assemble them into higher-level features in the second hidden layer and so on. This is made possible by the fact that neurons in successive convolutional layers are connected to one another. One of the most important advantages of the convolution layer is that it may reach very deep into the volume input layer because to the numerous (kernel) filters it contains.

These filters, much like the other hidden neurons, have weights associated with them to some degree. With each iteration of the training process known as "squeezing," the filters start the process of looking for new characteristics in the input picture that they have been given (epochs). Following the convolutional processing of the input data by the filter or kernel, the feature map is produced by applying the convolutional filter to the input data, where the matrix is multiplied at each point, and then extracting and collecting the result (Intuitive Guide to Convolution Neural Networks | by Thushan Ganegedara | Towards Data Science, n.d.). The op convolution process involves sliding the green filter (squares) over the blue (squares) input, as illustrated in Figure 12. The outcome of this process is represented in the little square on the right, which indicates the summation and adding of the result into the feature map.

Because the size of the filter will remain at 1 during the entirety of the sliding STRID operation, each sliding step will consist of one pixel. If the filter sliding step is more than a single pixel on the input, then the stride will become more significant. Because the input is always larger than the feature map, the feature map needs to be reduced in size before the convolution operation can be carried out. However, this step can be skipped by making use of padding, which involves the addition of a layer that is surrounded by another layer that contains pixels with zero values.

A feature map is the result of a convolutional layer once it is fully processed. CNN is able to generate non-linear results once the convolution output has been employed and passed through the activation function—in this case, ReLU, where the weighted results and outputs are transferred. This occurs after the convolution output has been employed and passed through the activation function. It is possible to write the ReLU function as (Equation 1 and Figure 13).

$$F\,(z)\ =\ max\,(0, z) \tag{1}$$

The process functions within a three-dimensional matrix consisting of width, height, and depth. However, for better comprehension, an explanation in a simplified two-dimensional format will be provided. Each filter produces its own two-dimensional output feature map, which is then combined and blended with the outputs of other filters to create a unified output volume. When CNN identifies a feature, edge, speck, or color spot, it activates the corresponding filters and starts acquiring knowledge about it. Similarly, when CNN encounters a color spot, it triggers the relevant filters and initiates the learning process. Please consult Figure 14 for a visual depiction of this procedure.

CNN's filters commonly apply 2D convolutions to both the data and the images they process. The figure that is included with this explanation should assist illustrate the process by demonstrating how the filter moves in both the X and Y directions in an iterative manner until the image's dimensions are established (Figure 15).

The following is a mathematical expression that may be used to describe this: D will represent the number of layers that are output, W will stand for the number of layers that are input, P will stand for the total volume of the warp pad, K will define the size of the kernel, and S will stand for the stride size (Abdelmaksoud et al., 2021).

$$D = \frac{[W-2P-K]}{S} + 1 \qquad (2)$$

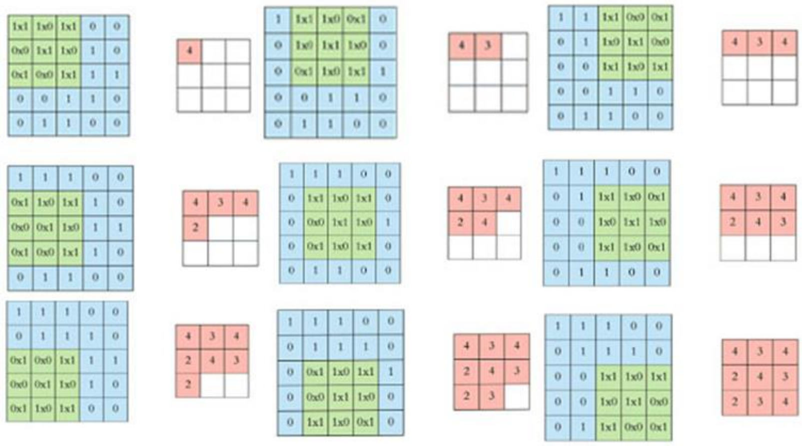

Figure 12. A filter is slid to the left (on the blue squares), and the resulting total is added and sent on to the feature map through the right (red squares)

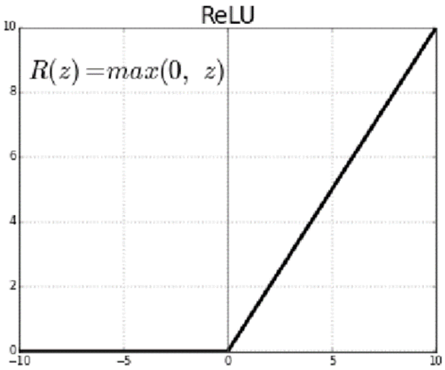

Figure 13. ReLU function

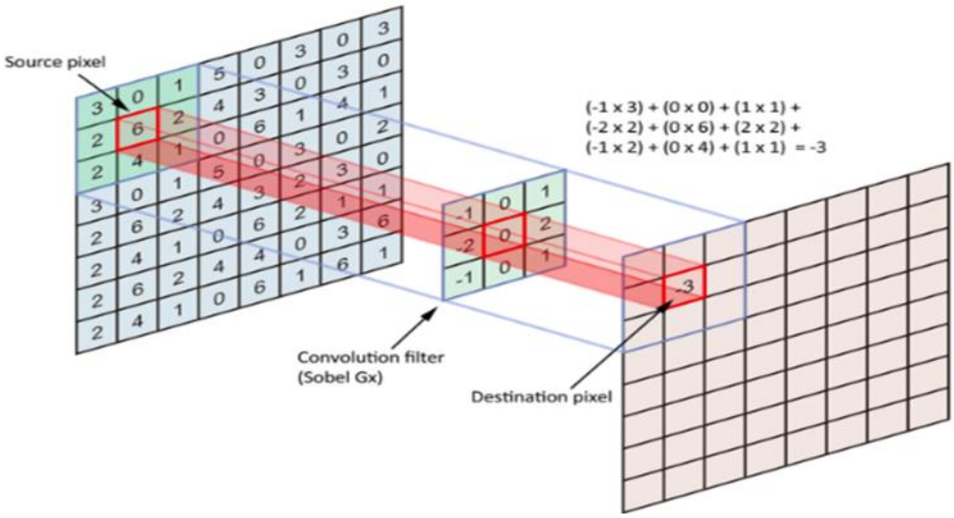

Figure 14. 33 filter different layer CNN example

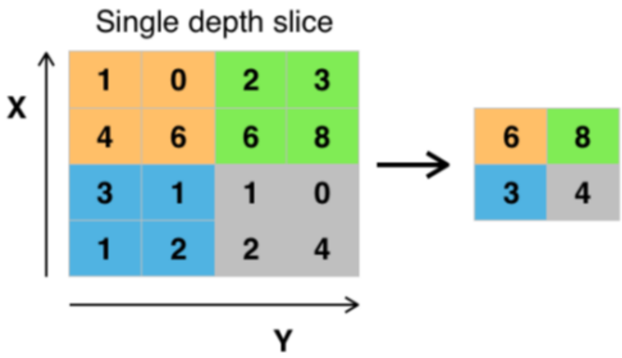

Figure 15. Without depth 2D Convolution (X, Y)

3.1.3. Pooling Layer

The pooling layer is a non-linear model that is an essential component of the CNN theories and is referred to as such. There are several important non-linear pooling layers, but the max pooling layer is by far the most common one. After the picture has been cut up into smaller rectangles by the pooling layer, the maximum output of the layer is determined by randomly sampling from the feature map. Each neuron in this layer is only able to communicate with a small subset of the neurons that came before it (pooling layer). In addition to that, it does its own thing without the necessity of using weights for training. By collecting the inputs from the layer below it and starting to reduce the amount of calculations that take place within the network, the pooling layer aims to minimize the locative volume of the input pictures, the amount of overfitting, and the amount of memory that is used by the network. This is accomplished by beginning to reduce the number of calculations that take place within the network. An illustration of the pooling layer strategy can be seen further down in Figure 16, while the terms MaxPooling and AvgPooling illustrate how aggregating neurons operates.

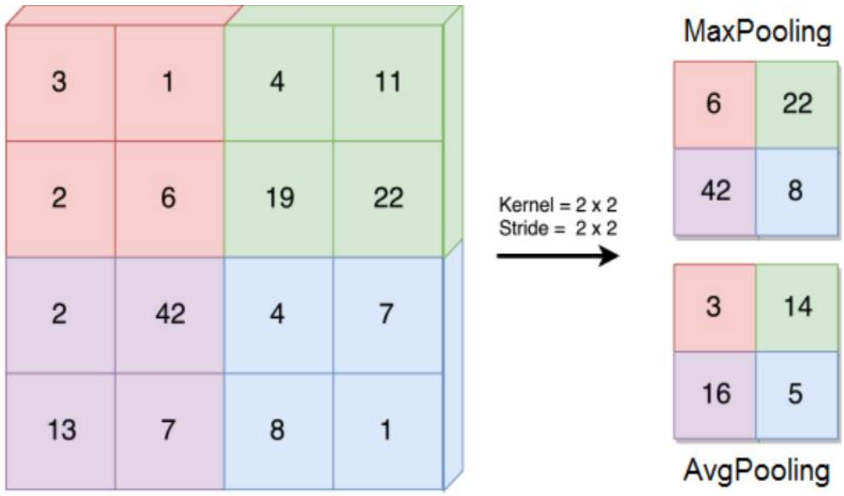

Figure 16. MaxPooling layer, AvgPooling example

3.2. Improvement Algorithms For Trining Neural Networks

In this section, the focus will be on discussing several prominent optimization methods, including stochastic gradient descent, Adam, and Rmsprop. These methodologies play a vital role in the training process of neural networks. To enhance the performance of the model, it is recommended to employ techniques such as dropouts and batch normalization (Kingma & Ba, 2014; Krogh, n.d.; Ruder, 2016). The objective is to minimize the loss function while simultaneously striving to improve the model's accuracy. When optimizing across multiple data instances, the primary aim is to minimize the overall loss to the greatest extent possible.

3.2.1. Root Mean Square Prop (RMSPROP)

The illustrious team of Tieleman and Hinton was responsible for the development of the ADMA optimization method, and they used this variation in their work. The momentum is used to create one-of-a-kind updates, which are then used in the process of rescaling the gradient (Krogh, n.d.). The momentum is used to rescale the gradient, which results in the generation of one-of-a-kind updates. RMSPROP will choose a one-of-a-kind learning rate for the model based on the factors that are currently in play. The speed of the learning process may be altered on its own. The procedure begins with the following equations due to the fact that an individual update must be performed on each parameter.

$$(v_t) = p\, v_{t_1} + (1 - p2) * g_t^2 \tag{3}$$

$$\Delta\omega_t = -\frac{\eta}{\sqrt{v_t + \epsilon}} \tag{4}$$

$$\omega_{t_{+1}} = \omega_t + \Delta\omega_t \tag{5}$$

3.2.2. Stochastic Gradient Descent (SGD)

To be more specific, it makes use of a stochastic gradient descent, abbreviated as SGD, to minimize loss while simultaneously updating the weights of a convolutional neural network, abbreviated as CNN, in order to accurately classify photos. This is accomplished by applying an update to the weights that is derived from a linear combination of the most recent weight update Vt and the negative gradient L. (W). W an is the weight that is calculated by making use of the negative gradient L(W) (Ruder, 2016), where a represents the whole amount of learning. Indicates the pace of change in comparison to the amount of time that has passed since the most recent update (Vt). SGD derives the new value Vt+1 for the weight based on the current weight Wt as well as the most recent update of the weight Vt. This new value represents both the updated weights as well as the new weights at the repetition T+1.

$$V_{t_{+1}} = \mu V_t - a\nabla\mathcal{L}(W_t) \tag{6}$$

$$V_{t_{+1}} = W_t + V_{t_{+1}} \tag{7}$$

3.2.3. ADAM

It is a strategy for optimization that is based on gradients and incorporates an adaptive moment estimation (m t,V t). The update's parameters are included in the following set of equations.

$$(m_t)_i = \left(m_{t_{-1}}\right)_i + (1 - \beta_1)\left(\nabla\mathcal{L}(w_t)\right)_i \tag{8}$$

And

$$(v_t)_i = \beta_2\left(v_{t_{-1}}\right)_i + (1 - \beta_2)\left(\nabla\mathcal{L}(w_t)\right)_i^2 \tag{9}$$

CHAPTER FOUR

METHODOLOGY

4.1. Proposed Architecture

For the detection of DR inserts, it was always intended to make use of a sequence of photographs of the retina. Grayscale images are supplied into a system called CLAHE that optimizes retinopathy diagnosis and detection. Color images are filtered using techniques and applications of optimization, and their features are encoded using GLCM before they are submitted to a convolutional neural network. Figure 17 depicts the overall design of the system, in which color and grayscale images are fed into a CNN network to predict retinopathy subtypes using characteristics that were retrieved during the practical enhancement, transformation, and transfer to GLCM stages. These stages are depicted as having taken place in the system.

To train CNN from scratch, a substantial amount of labeled data is essential. The success of the training process depends on three key factors: the quantity of data available, the quality of the photographs, and the CNN architecture employed (Dreyer & Raymond Geis, 2017; Greenspan et al., 2016; Shen et al., 2017). In situations where the initial data collection is inadequate, transfer learning and augmentation techniques can be employed to address the deficiencies. Transfer learning involves leveraging a pre-trained model to address a different problem. In other words, it is an approach to learning that can be applied across a diverse range of scenarios.

In the field of research on deep learning, this methodology is now regarded as a state-of-the-art approach. Because of this, Deep Neural Networks may be trained with a relatively small amount of data, which is an advantage.

Due to the fact that labeled data sets are often inaccessible for use in practical applications, this is an extremely important function. It is difficult to acquire sufficient data in the medical field because specialized annotations are expensive, certain diseases (eggs, lesions) are uncommon, or necessary data cannot be acquired for reasons such as patient confidentiality. All of these factors contribute to the difficulty of acquiring sufficient data. In this hypothetical situation, CNN models are moved either directly from one medical application to another or from one medical application to another while utilizing data obtained from a natural image data set. The ImageNet database contains more than 1.4 million photos that have been sorted into more than a thousand different categories. Some examples of pre-trained networks that have been trained using ImageNet datasets are as follows: AlexNet (Krizhevsky et al., 2017), VGG-16 (Simonyan & Zisserman, 2014), GoogLeNet (Szegedy et al., 2015), SqueezeNet (Iandola et al., 2022), and ResNet (He et al., 2016). Table 1 presents a few instances of pre-trained networks along with the attributes that are associated with each one. The transfer learning method for image processing may make use of a range of different pre-trained networks, however this is determined by the type of network being used. Previous study (I. Banerjee et al., 2018; Huynh et al., 2016) employing neural networks for medical image processing that were first trained on natural picture datasets showed some good results, which is encouraging because these datasets were used to train the neural networks.

Instead of applying the normal CNN method in the architecture shown in figure 17. GoogLeNet and AlexNet will be used in this study in order to achieve better performance.

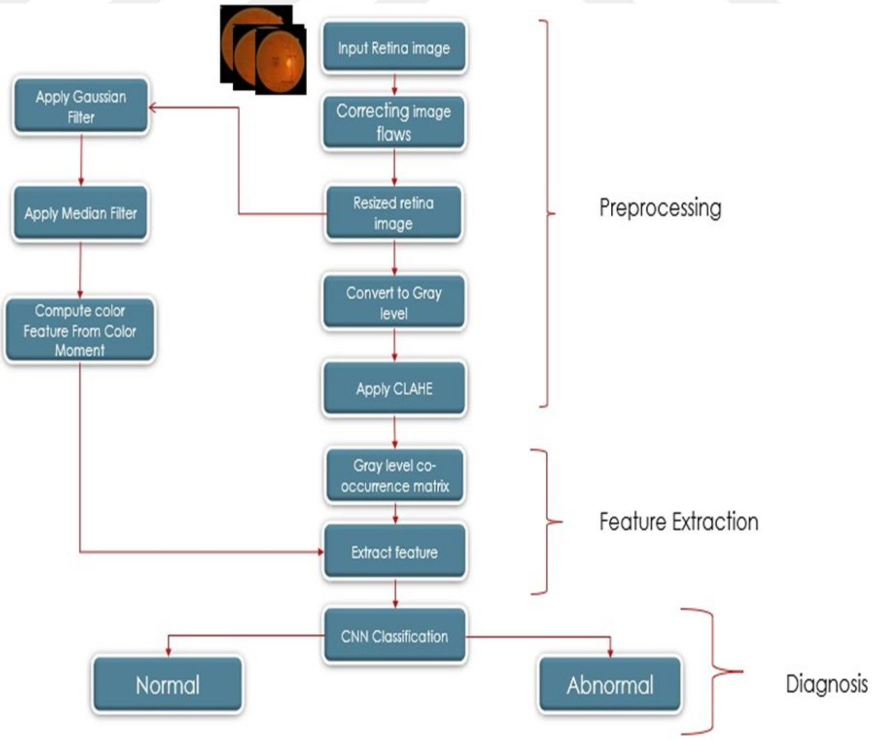

Figure 17. General flowchart CNN- based for proposed method.

Table 1. Some pre-trained networks properties

	Year	Deep	Layers	MB	Architecture
GoogLeNet	2014	22	144	27	Inception module
AlexNet	2012	8	25	227	Traditional sequential network
VGG-16	2014	16	41	515	Traditional sequential network

4.2. Googlenet Model

Google Net was developed with the assistance of a number of colleges and was initially presented in the research publication titled "Going Deeper with Convolutions" which was published in 2014 (Or Inception V1). This arrangement was deemed to be the best overall in the ILSVRC picture categorization contest that took place in 2014. In comparison to the previous two champions, AlexNet and ZF-Net, and even VGG, it has achieved a significant reduction in the error rate (2014 runner up). The design method must therefore involve the utilization of 11 convolutions and the pooling of data from all around the world (Szegedy et al., 2015).

4.2.1. Googlenet Features

Because it makes use of a wide variety of different techniques, such as 11 convolution and global average pooling, it is feasible to construct more complicated systems using it than are now imaginable. during the entirety of the planning and conceptual stages of the building.

The design incorporates eleven convolutions, each of which plays an important role in ensuring the overall success of the conceptual architecture. As a result, the use of these convolutions was necessary throughout the whole process of design in order to cut down on the number of variables. By paring down the number of settings, one may have a richer experience with the architecture. The following are ten instances of phrases that are overly complicated:

When utilizing 48 filters, it is possible to conduct a 5x5 convolution directly, so skipping the 11 convolutions that serve as an intermediate step (Figure 18). The following is a tally of all the procedures that were carried out: Overall, 112.9 MB of data were collected in its entirety (Figure 19).

At the very end of the design process for GoogleNet, global average pooling is included in order to substantially reduce the amount of data that must be stored. This layer takes the average of a 7x7 feature map in order to generate an output that is 1x1. Due to the absence of any additional parameters to learn, the top-1 accuracy has improved by 0.6% following the cessation of new parameter additions.

The final product is constructed by stacking the results of the 11, 33, and 55 convolutions as well as the results of the 33 max pooling that was performed by the Inception module. Additionally, the results of the 11 max pooling that was performed by the Inception module are also included in the final product. Objects of varying sizes can be more effectively managed with the use of convolution filters of varying sizes, according to this line of thinking (Figure 20).

In the middle of the Inception architecture are a few different classifier forks that are solely used for training. All of the layers in the network are connected to one another. There are two layers with 1024 outputs, two layers with 1000 outputs, two average pooling layers with 55 inputs each with a stride of three, and a softmax classification layer. Because of the contribution of this stratum, the overall loss was 0.3 percentage points higher than it would have been otherwise. These layers not only provide regularization, but also help with the problem of gradients fading as they go away from the camera.

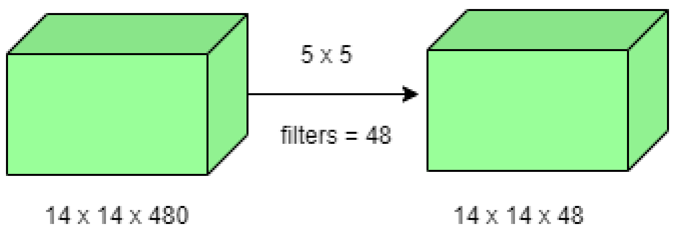

Figure 18. Filter with 5 by 5 size

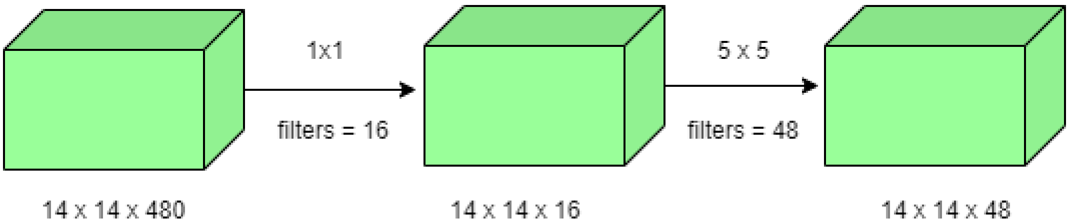

Figure 19. 1by 1 filter use

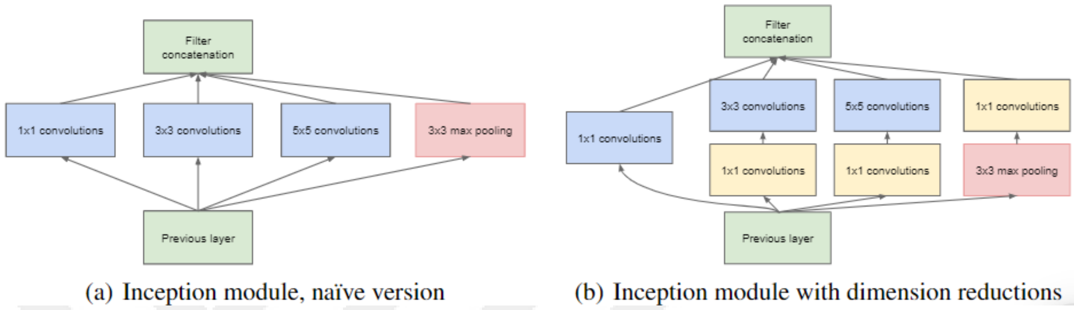

Figure 20. Architecture for inceptual model

4.2.2. Model Architecture

The complete design can be seen in Figure 21, and it comprises of a total of 22 different levels. When conceiving the design, it was essential to take into account how well computers might be utilized in the final product. The core concept is that the design may be implemented on a single device, despite the fact that this device may only have limited computing capabilities. The output of Inception layer 4 is passed into two more classifier levels in the architecture, which are referred known as layers 4a and 4d.

In the sections that follow, the topic of auxiliary classifiers is explained:

- The kit comes with Stride 3, in addition to a conventional pooling layer that has a filter size of 55 microns.
- A convolution involving 128 different filters is used so that dimension reduction and ReLU activation may be achieved.
- The ReLU has been activated in a layer that has a total of 1025 outputs and full connection.

- According to the statistics, the percentage of students who drop out due to Dropout Regularization is 0.70.
- The accuracy of classification provided by a 1000-class softmax classifier is comparable to that provided by the primary softmax classifier.

This layout makes use of a picture with a resolution of 224 by 224 pixels, which enables the RGB color channels to be utilized to their greatest potential. Rectified Linear Units (ReLUs) are the only activation functions that are utilized in these designs (Figure 22).

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Figure 21. General architecture

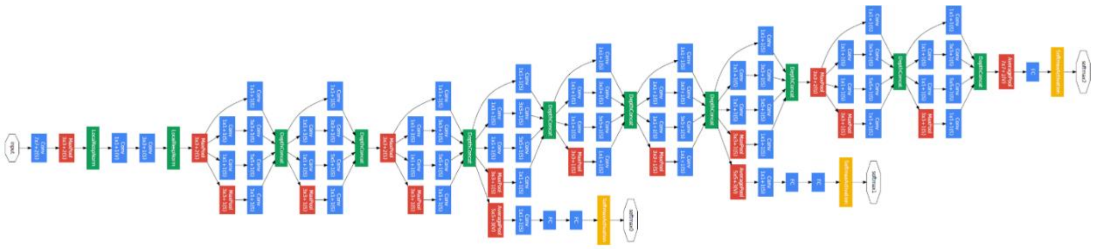

Figure 22. Rectified Linear Units (ReLU) based architecure

4.3. AlexNet Model

AlexNet used their Convolutional Neural Network Architecture to win the 2012 LSVRC.

The LSVRC (Large Scale Visual Recognition Challenge) pits research teams against one another in an effort to increase accuracy across many visual recognition challenges. They achieve this by putting their algorithms through their paces on ImageNet, a vast collection of tagged images. The sides' outlooks on the contest evolved drastically as a result of this.

There are a total of eight weighted layers in AlexNet, with five convolutional layers and three fully connected layers making up the network.

Utilizing ReLu activation at the conclusion of each successive layer, up to the final one, results in the production of a softmax distribution that is applied to all one thousand class labels. The first two linked layers have a dropout operation done on them. In addition to the first and second convolutional layers, as illustrated in the preceding image, the max-pooling operation is also performed after the fifth convolutional layer. Only the maps of kernels from the layer before it, which are kept on the same GPU, are connected to the kernels of the layer that is now being processed in the convolutional layers that are the second, fourth, and fifth respectively. All of the kernel mappings from the second convolutional layer are connected to all of the kernels from the third layer. A layer is said to be completely linked when all of its neurons are connected to the neurons in the layer below it.

ImageNet is an image database that contains about 15 million high-resolution photos that have been tagged with descriptive information. Researchers have been assigned with the goal of lowering the top-1 and top-5 error rates for a subset of the pictures contained in ImageNet as part of this competition. A 256-by-256-pixel RGB picture serves as the input size for AlexNet. The resolution of both the test photos and the training images has to be 256 pixels on each side. This indicates that information must first be transformed into 256x256 format before it can be used for training the network.

One of the distinguishing characteristics of AlexNet is the fact that it employs the nonlinearity of the Rectified Linear Unit (ReLU). In the past, the Tanh and sigmoid activation functions were utilized rather extensively for the process of training a neural network model. Deep convolutional neural networks (CNNs) that are taught utilizing ReLU nonlinearity may be learnt significantly more quickly than those that are trained with saturated activation functions like tanh or sigmoid, as the results of AlexNet's research have shown.

When z is increased to high enough values, the tanh function reaches both its maximum and its lowest. In some parts of the graph, the slope of the function is very close to zero. As a direct consequence of this, the length of time required to get down a slope might potentially rise.

When applied to large positive values of z, the ReLU function has a slope that is not zero. Because of this, the amount of time needed for optimization is cut down. Even though the slope is constant at zero for all z values, the vast majority of neurons that make up a neural network will almost always have positive values.

Because of its enormous size, AlexNet presented a particularly difficult challenge when attempting to eliminate overfitting (60 million parameters). Two strategies that can be utilized in the fight against overfitting are known as data augmentation and dropout.

4.3.1. Data Augmentation

Because the researchers translated and mirrored the photos on the training set horizontally, they were able to add a total of 2048 additional pictures to the collection. Following the application of principal component analysis (PCA) on the RGB pixel data in order to adjust the intensities of the RGB channels, the top-1 error rate was reduced by about 1%.

4.3.2. Dropout

AlexNet relies on a dropout as the second tactic in its arsenal to combat overfitting. The output of each suppressed neuron is individually canceled out with a probability of fifty percent. Because they were "dropped out," these neurons will not take part in the forward pass or the backpropagation that comes after it. As a result, the neural network tries out a different configuration every time it is given a new input. This technique involves inhibiting neurons one at a time in accordance with a predetermined probability. This suggests that neurons that are "turned off" will not engage in the forward pass or backpropagation until they are expressly awakened to take part in either of these processes (Figure 23).

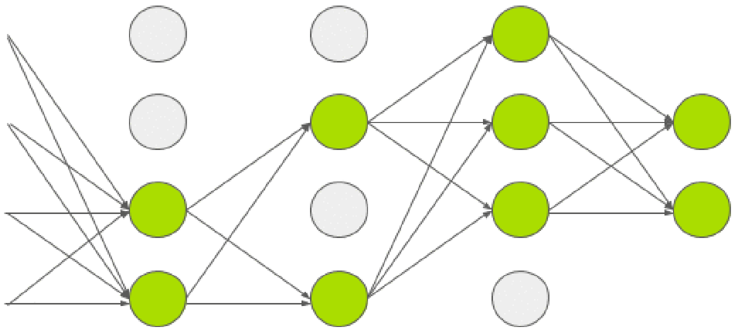

Figure 23. AlexNet Network

4.4. Data Augmentation Techniques

Processing an image that was generated from a huge portion of the CNN (Convolutional Neural Network). The photographs in this collection depict a patient who is from the Asia-Pacific area and who has a diverse variety of traits. To preserve high-resolution photographs, a significant quantity of storage space is required. Figures 23 and 24 contain both of these photographs, respectively. After that, the dataset was reduced in size so that it would be compatible with the input size of 224x224 that GoogleNet and AlexNet uses. There were almost two hundred and fifty datasets taken into consideration. This collection contains 161 DR fundus photos in addition to the normal 39 fundus photographs. Each picture saved in the.jpg format has a resolution of 3152 by 3000 pixels, which adds up to a total of 9 million pixels. All of the photographs have had their file sizes decreased so that they are compatible with the GoogleNet and AlexNet interface. It's possible that RGB photographs, which are more difficult to process and take more time to do so, disclose more details than other image formats.

Figure 24. Normal Fundus Image

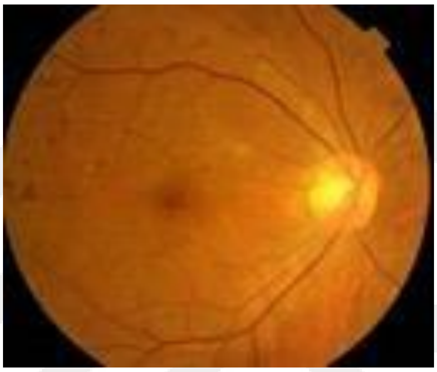

Figure 25. DR Fundus Image

4.5. Image Preprocessing

It is generally accepted that accuracy, sensitivity, specificity, and area under the curve (AUC) are the metrics that are utilized in the process of evaluating classification algorithms (Area Under the Curve). It is generally accepted that a decent metric of the effectiveness of a model is the proportion of cases that have been correctly categorized in relation to the total number of examples.

4.6. Monitor Deep Learning Training Progress: GoogLeNet and AlexNet – Matlab

Developing the practice of keeping a journal of your workout outcomes is a good habit to cultivate. You may monitor the effectiveness of your training over time by charting a number of different indicators. It is necessary to conduct experiments in order to determine whether or not the accuracy of the network is increasing and whether or not it is beginning to overfit the training data.

trainNetwork creates a new figure after each training iteration for the network and displays the training metrics in that new figure when the training option "training-progress" plots is selected in trainingOptions. This occurs after each training iteration for the network. The parameters of the network are adjusted after each iteration of the gradient computation so that they take into account any newly acquired data. If the validation data in trainingOptions is provided for this reason, then the validation metrics will also be exhibited in the image whenever trainNetwork checks the network. This occurs anytime the validation data is provided. The following, which is mirrored in the graph, is what the data seems to suggest:

- The level of accuracy achieved in training may be evaluated based on how successfully fresh subsets of training data are categorized.
- Improved accuracy in training that has been smoothed out: Before the smoothing approach can be utilized, the initial training accuracy needs to be smoothed out first. Because it contains less background noise than the un-smoothed accuracy, the smoothed accuracy makes pattern detection much simpler.
- Accuracy in the validation process: When it comes to testing, precision refers to the degree to which every member of the validation sample falls inside the categories that have been set for that sample (specified using trainingOptions).
- Training loss, smoothed training loss, and validation loss are all distinct concepts that may be distinguished from one another. There are three distinct varieties of training losses: mini-batch, smoothed mini-batch, and validation-set losses. Mini-batch losses are the most common. The classificationLayer is the final layer in your network, and the loss function for this layer is defined in terms of the cross entropy. The principles of loss functions, as they relate to classification and regression, are broken down into easier-to-understand chunks by a group of experienced authors in this article.

CHAPTER FİVE

RESULTS AND ANALYSIS

5.1. Algorithm Implemented

MATLAB, a powerful programming language that is well-suited to scientific work and numerical computations, was utilized in the process of developing the approach and the code for use in this investigation. It would not have been possible to complete this operation without MATLAB's help libraries and other resources, which were readily available to users. Both GoogleNet and AlexNet are being considered for use as CNN architectures for this particular project.

5.2. Pre-processing

During the pre-processing step of image editing, the first layer is always selected manually as the starting point for the work. This layer determines the maximum allowed dimensions for the input picture, which for GoogleNet and AlexNet is 224 by 224 by 3. Before the images are used for training on the network, they are scaled down to meet this size. This process for scaling is applied to both the training set and the testing set. First, the RGB values of each picture are averaged, and then the values of each pixel are subtracted from the average of the color channel to which they are assigned. The development of this pre-processing method, which outperformed previous techniques such as histogram equalization and min-max normalization, was made possible by conducting experiments with a wide range of retinal images. This strategy emerged victorious from a Kaggle competition that was held to determine how best to categorize DR.

5.3. Algorithm Continued

The training data makes up about 70 percent of the whole dataset, whereas the test data only constitutes 30 percent of it. The results of these tests demonstrated that the model frequently overfit the data, which indicates that it performed admirably on the training data but struggled to generalize its findings to new data. In order to fight this issue, a more refined version of the model is currently being developed. This version will reduce the risk of the model being overfit and will enhance the likelihood of the model being able to generalize.

5.4. Fine-tuning

Because of the limits imposed by time, the learning pace of the network is restricted. It was discovered that increasing the epoch size and employing a range of strategies combined resulted in an improvement in the validation procedure's level of precision. It was also found that the model could only classify three DR phases from the input data, despite the fact that the design had included more stages for the model to account for in its construction. The potential of the model to learn and evolve is going to be improved upon as a direct result of this technique for making little adjustments.

5.5. Results and Graph

The entire architecture of GoogleNet, in all of its layered, may be seen in Figures 26 and 27. The entire architecture of AlexNet, in all of its layered, may be seen in Figures 28 and 29. There are offered graphical representations of the outcomes as well as the effectiveness of the various training methodologies.

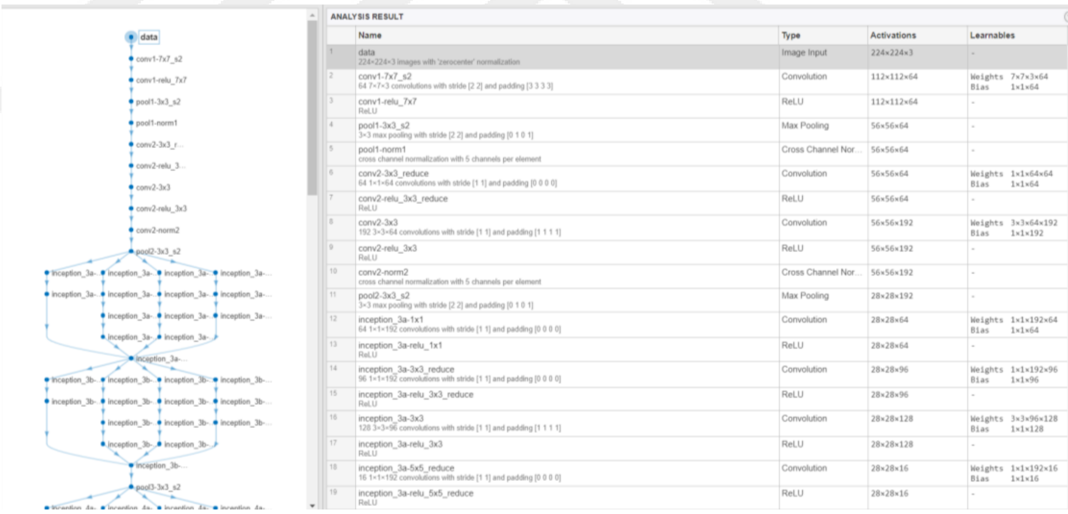

Figure 26. GoogleNet CNN output architecture

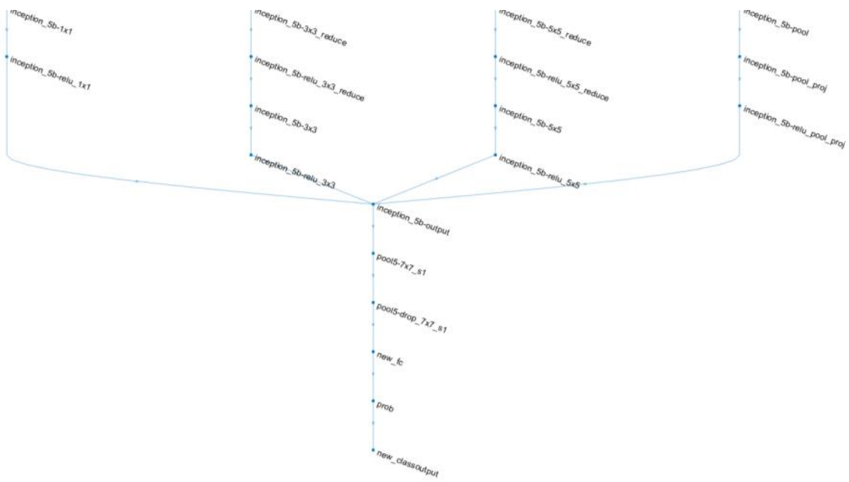

Figure 27. GoogleNet CNN output architecture

ANALYSIS RESULT

	Name	Type	Activations	Learnables
1	data 227×227×3 images with 'zerocenter' normalization	Image Input	227×227×3	-
2	conv1 96 11×11×3 convolutions with stride [4 4] and padding [0 0 0 0]	Convolution	55×55×96	Weights 11×11×3×96 Bias 1×1×96
3	relu1 ReLU	ReLU	55×55×96	-
4	norm1 cross channel normalization with 5 channels per element	Cross Channel Nor…	55×55×96	-
5	pool1 3×3 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	27×27×96	-
6	conv2 2 groups of 128 5×5×48 convolutions with stride [1 1] and padding [2 2 2 2]	Grouped Convolution	27×27×256	Weigh… 5×5×48×128… Bias 1×1×128×2
7	relu2 ReLU	ReLU	27×27×256	-
8	norm2 cross channel normalization with 5 channels per element	Cross Channel Nor…	27×27×256	-
9	pool2 3×3 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	13×13×256	-
10	conv3 384 3×3×256 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	13×13×384	Weights 3×3×256×384 Bias 1×1×384
11	relu3 ReLU	ReLU	13×13×384	-
12	conv4 2 groups of 192 3×3×192 convolutions with stride [1 1] and padding [1 1 1 1]	Grouped Convolution	13×13×384	Weigh… 3×3×192×192… Bias 1×1×192×2
13	relu4 ReLU	ReLU	13×13×384	-
14	conv5 2 groups of 128 3×3×192 convolutions with stride [1 1] and padding [1 1 1 1]	Grouped Convolution	13×13×256	Weigh… 3×3×192×128… Bias 1×1×128×2
15	relu5 ReLU	ReLU	13×13×256	-
16	pool5 3×3 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	6×6×256	-
17	fc6 4096 fully connected layer	Fully Connected	1×1×4096	Weights 4096×9216 Bias 4096×1
18	relu6 ReLU	ReLU	1×1×4096	-
19	drop6	Dropout	1×1×4096	-

Figure 28. AlexNet CNN output architecture

Figure 29. AlexNet CNN output architecture

5.5.1. GoogleNet results

GoogLeNet was trained and tested with both 6 and 20 epochs in order to compare the outcomes of the two different training and testing scenarios.

Following a total of 6 cycles of training, the validation set accuracy ended up being 75% (figure 30). This indicates that the model was effective in classifying 75% of the images in the validation set after just 6 iterations of training. The loss function decreased during the course of training, which is an indication that the model became more adept at generating predictions as it was refined.

The validation set accuracy reached an all-time high of 87.5% after being trained for a total of 20 epochs (figure 31). This suggests that the model's understanding of the data as well as its ability to predict outcomes improved with each consecutive training session. The loss function decreased at a more precipitous rate, which is indicative of a more discernible improvement in the usefulness of the model.

The results of this study demonstrate that GoogLeNet is better than other approaches for the classification of pictures. The accuracy of the model was already rather excellent after only 6 epochs of training, and it continued to improve significantly after 20 epochs of training. The model's ability to learn and adjust to new data was demonstrated by the fact that its performance increased as further training rounds were completed.

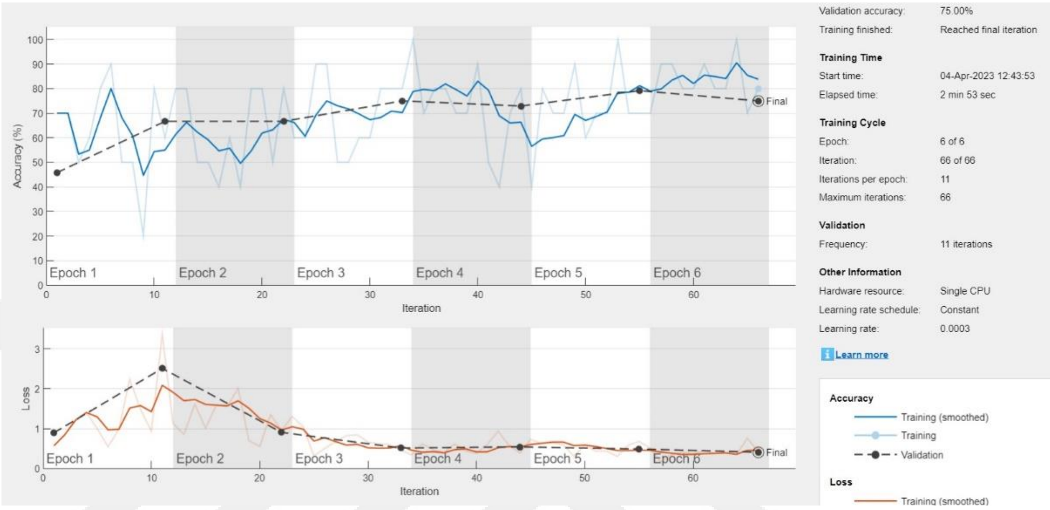

Figure 30. GoogleNet Accuracy for 6 epochs

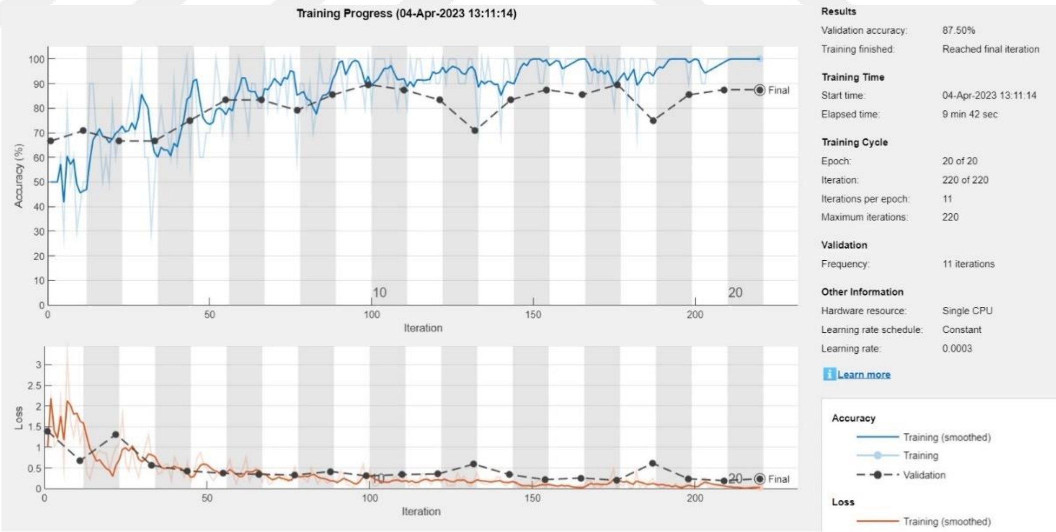

Figure 31. GoogleNet Accuracy for 20 epochs

5.5.2. AlexNet results

In order to evaluate the differences in the results obtained from training and testing AlexNet with 6 and 20 epochs, respectively.

After undergoing 6 cycles of training, AlexNet was able to successfully complete the validation set with an accuracy of 85.42% (figure 32). This indicates that the model was able to correctly classify 85.42 percent of the images included in the validation set after just undergoing training for a total of 6 epochs. The decrease in the loss function that occurred throughout the course of training indicated that the model was growing better at predicting future outcomes.

After training for an additional 20 epochs, the model achieved an accuracy of 100 percent on the validation set (figure 33). This demonstrates that the model's comprehension of the data improved as the number of training iterations rose, which led to a rise in the accuracy of the model's predictions. The loss function fell at a more significant rate, which is indicative of the model's performance improving to a more noticeable degree.

This data provides evidence of the usefulness of AlexNet for the classification of pictures. After only six iterations of training, an acceptable degree of accuracy was achieved, and after twenty iterations, considerable gains were proven to have been made. The steadily diminishing loss function illustrates how the model is able to learn from the dataset and adjust its behavior accordingly. These findings provide evidence that AlexNet is reliable for picture classification tasks in general.

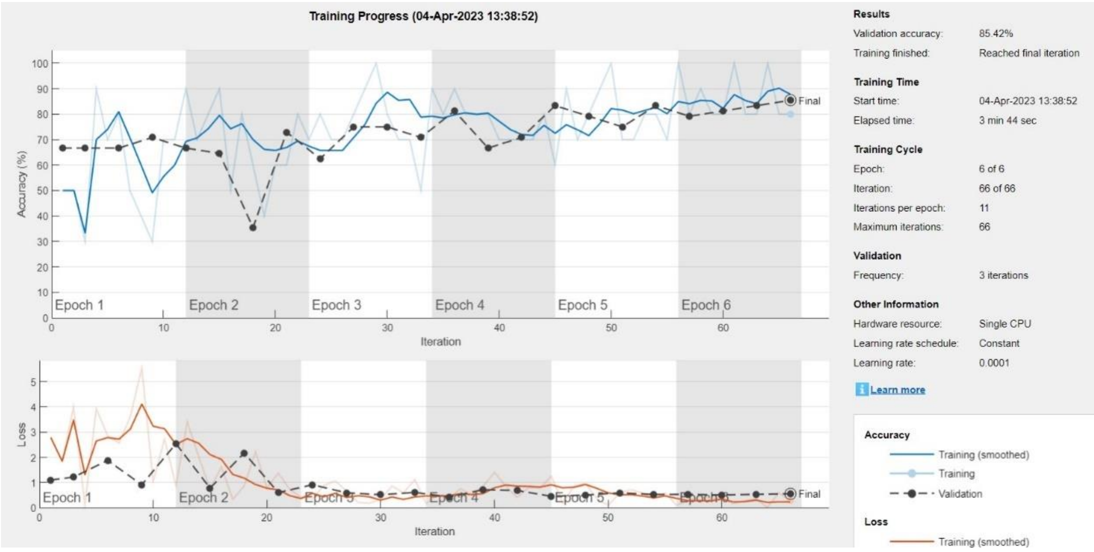

Figure 32. AlexNet Accuracy for 6 epochs

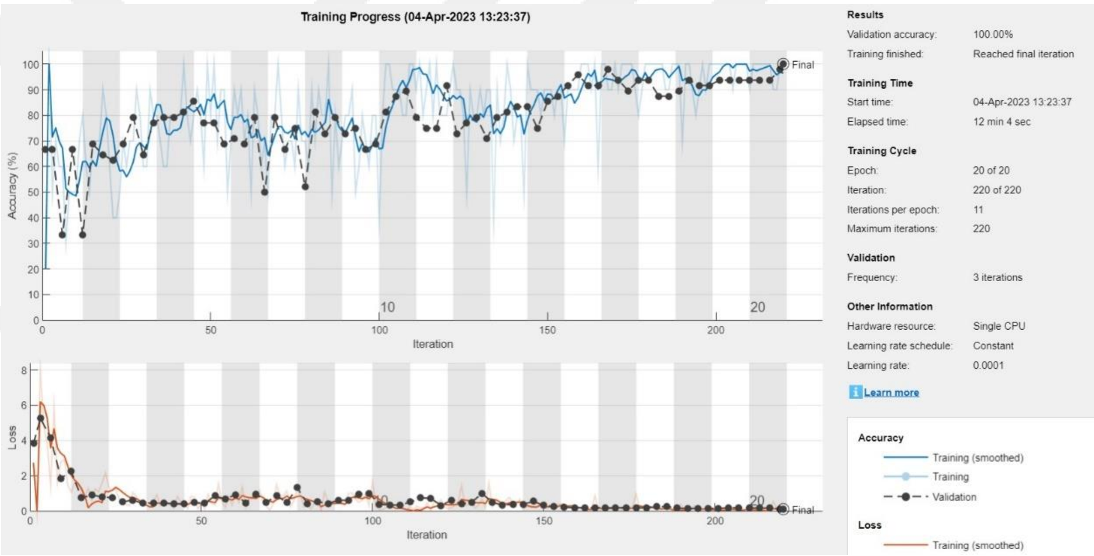

Figure 33. AlexNet Accuracy for 20 epochs

5.6. Indicator to Display Training Progress Information

If the indicator is set to 1, which denotes that something is true, or 0 which denotes that something is untrue, the command window will not disclose specifics about the user's progress while they are being trained.

Table 2 of the verbose output provides further information and specifics on the training procedure.

Table 2. Terminology

Field	Description
Epoch	The date of the beginning of the era. In computing, an epoch represents one complete round of data.
Iteration	the nth iteration in the context of software development, an iteration is a small batch.
Time Elapsed	Hours, minutes, and seconds were measured.
Mini-batch Accuracy	The mini-batch classification accuracy.
Validation Accuracy	The accuracy of categorization is based on the validation data. If you do not give any validation data, this field will not be shown by the function.
Mini-batch Loss	The mini batch was unable to be recovered. When the output layer is a ClassificationOutputLayer object, the cross-entropy loss for multi-class classification problems with mutually exclusive classes is represented by the variable loss.
Validation Loss	The validation data was a failure. Loss is the cross-entropy loss for multi-class classification problems with mutually exclusive classes when the output layer is a ClassificationOutputLayer object. The function does not display this field if you do not provide validation data.
Base Learning Rate	The rate at which a person learns a new skill is measured. When this value is multiplied by the learn rate factors of the layers, the final result is obtained.

5.7. Discussion

In this study, the outcomes of multiple training iterations for AlexNet and GoogleNet were evaluated together with the outcomes of certain epochs. The average accuracy of each model was evaluated after 10 iterations at both 6 and 20 epochs. 79% was the average accuracy achieved by GoogleNet throughout the course of 10 iterations using 6 epochs. This indicates that after 6 epochs of training, GoogleNet achieved an average accuracy of 79% when classifying the validation images. Indicative of the model's reliability and consistency is the fact that its accuracy has improved when compared to the individual result of 75%. GoogLeNet's accuracy experienced a significant boost after 20 iterations, reaching an aggregate average of 91% across all tests after reaching this milestone. As a result of this, the average accuracy is higher than the individual result of 87.5%, which demonstrates that the model has the ability to learn and improve its predictions across several training sessions. After being trained for 6 iterations, AlexNet was able to attain an accuracy of 92% on average across 10 tests. This indicates that after 6 epochs of training, AlexNet was able to correctly recognize the validation photos 92% of the time on average. The excellent starting performance of the model as well as its consistency all the way through the repetitions is reflected in the high average accuracy. After 20 epochs, AlexNet's average accuracy achieved an astounding 99.1 percent, which is an all-time high. This indicates that, if it receives sufficient training repetitions, the model has the potential to learn complex patterns and generate classification results that are virtually faultless. These new findings provide more evidence that demonstrates the repeatability and dependability of the functioning of AlexNet and GoogleNet. The greater average accuracies that AlexNet maintained throughout both the 6 and 20 epochs are indicative of the network's superior beginning performance as well as its superior convergence compared to that of GoogleNet. The accuracy gap between the two models, however, shrunk as GoogLeNet also showed improvement through repeated training. The findings indicate that both AlexNet and GoogLeNet are successful and dependable in picture classification tasks, with AlexNet demonstrating higher average accuracies across a number of repeats and epochs. GoogLeNet, on the other hand, did not do as well in this area. These findings give light on how to choose the most appropriate model for a particular application or dataset based on characteristics like accuracy and convergence speed. Table 4 presents a comparison with previous studies (Gangwar & Ravi, 2021; Harangi et al., 2019; Lam et al., 2018; Mobeen-Ur-Rehman et al., 2019; Qomariah et al., 2019; Zhang et al., 2019). The GoogleNet and AlexNet architectures were utilized, together with the Indian diabetic retinopathy photo collection, in order to carry out this research. This is an important point to make. When the results of this study are compared to those of other studies, you will find that it has the highest level of accuracy. For instance, the GoogleNet model was able to improve upon earlier work by a factor of 75% in the study "Automated Detection of Diabetic Retinopathy using Deep Learning" by Carson Lam et al., and by 72% and 82%, respectively, in the study "Diabetic Retinopathy Detection Using Transfer Learning and Deep Learning" by Akhilesh Kumar Gangwar and Vadlamani Ravi. In addition, after 20 epochs, the AlexNet model had reached 99.1 percent accuracy, which was far higher than the reported accuracy of any previous research by a large margin. This is evidence that your approach appropriately classifies retinal images linked with diabetic retinopathy. It is essential to keep in mind that the outputs might vary widely based on the dataset, preprocessing methods, and model architectures that are utilized. This is something that must be kept in mind at all times. Although other research has utilized alternate datasets such as Messidor and IDRiD and employed different CNN designs such as Inception-ResNet-v2, custom CNNs, and ResNet50, this study used a specific dataset and a CNN architecture. Other studies have also used other CNN designs. In conclusion, the work demonstrates that GoogleNet and AlexNet designs are effective on the dataset that was chosen, which is a contribution to the field of automated diabetic retinopathy detection. The outstanding findings obtained by these models shed light on the potential that deep learning strategies hold for giving accurate diagnoses of diabetic retinopathy. However, further research is necessary to verify the performance of the models on datasets that are larger and more diverse, as well as to examine alternative optimization and preprocessing procedures that can improve the models' performance and generalizability.

Table 3. Results comparison

Training option	epoch	Accuracy	Time elapsed	Average Accuracy after 10 iterations
GoogleNet	6	75%	2 min and 53 seconds	79%
	20	87.5%	9 min and 42 seconds	91%
AlexNet	6	85.42%	3 min and 44 seconds	92%
	20	100%	12 min and 4 seconds	99.1%

Table 4. Comparison with previous studies

Paper Title	Dataset	Models	Accuracy
Automated Detection of Diabetic Retinopathy using Deep Learning	Messidor dataset was supplemented with a Kaggle partition (MildDR) consisting of 550 images	GoogLeNet	74.5%
		AlexNet	68.8%
		ImageNet	57.2%
Classification of Diabetic Retinopathy and Normal Retinal Images using CNN and SVM	77 retinal images from Messidor database	CNN and SVM	95.83%
Diabetic Retinopathy Detection Using Transfer Learning and Deep Learning	Messidor-1 and APTOS dataset	Inception-ResNet-v2	72.33% and 82.18%
Classification of diabetic retinopathy images based on customised CNN architecture	MESSIDOR (1200)	CNN (AlexNet, VggNet16, custom CNN)	98.15%
Automatic screening of fundus images using a combination of convolutional neural network and hand-crafted features	Kaggle (22,700) and IDRiD (516)	CNN (AlexNet)	90.07%
Automated identification and grading system of diabetic retinopathy using deep neural networks	Their Own dataset (13767)	CNN	96.5%
GoogleNet CNN Classifications for Diabetics Retinopathy	Selected Kaggle dataset	GoogleNet	96%
Our method	Own dataset selected from the Indian diabetic retinopathy image dataset of Kaggle	GoogleNet	91%
		Alexnet	99.1%

CONCLUSION AND RECOMMENDATIONS

This study was conducted with the intention of comparing and contrasting the capabilities of two frequently utilized CNN architectures, namely AlexNet and GoogLeNet, when applied to the issue of photo classification. A wide range of tests and investigations have been conducted in order to evaluate the capabilities and effectiveness of various systems for learning and making predictions regarding picture categories.

In Chapter 5, a complete review of the methodologies was presented that were utilized in this research. The various pre-processing approaches were discussed, how to keep the algorithm running, how to fine-tune it, and what type of outcomes and graphs you can anticipate from the various training procedures. By contrasting how well both AlexNet and GoogleNet performed under varying epoch conditions, the study was also able to give a comprehensive analysis of the accuracy and convergence rates of both networks.

According to the findings of the research, both AlexNet and GoogLeNet performed admirably in the image classification tasks given. On the other hand, the distinctions between the two models grew more obvious. After only 6 epochs of training, AlexNet originally demonstrated a higher initial average accuracy (92% against 79% for GoogleNet) than did GoogLeNet. This proved that AlexNet has superior starting performance and learning potential, as indicated by its early-on in training accuracy improvements. Specifically, this was demonstrated by the fact that AlexNet has superior learning potential.

Following the completion of 20 training epochs, the accuracy of both models had greatly improved. AlexNet's average accuracy of 99.1% is rather impressive when compared to GoogleNet's average accuracy of 91%. Based on these findings, it was hypothesized that the models may learn complicated patterns and increase their ability to forecast with more training rounds. The fact that AlexNet has a greater average accuracy after 20 epochs is further evidence of the network's strong convergence and its capacity to achieve almost perfect classification.

In summary, the findings of the research demonstrated that both AlexNet and GoogLeNet are capable of performing admirably when asked to classify images.

AlexNet had a stronger starting performance and average accuracy, whereas GoogleNet exhibited significant gains with additional training rounds. AlexNet was able to learn faster than GoogleNet. It is important to keep in mind that the performance of the model may shift dramatically depending on the data and the activity at hand.

In conclusion, the CNN architecture that should be employed is decided upon according on the specific requirements and objectives of the job. When it comes to performing image classification tasks with high accuracy and in a relatively short amount of time spent on training, AlexNet is a compelling choice to consider. On the other hand, GoogleNet has the potential to improve with more training cycles, which makes it a viable solution for software programs that require extensive learning and convergence.

This research contributes to the improved grasp of CNN architectures, particularly insofar as it relates to the subject of the classification of pictures. Additional research might be done to evaluate the generalizability of the models' domains and datasets in terms of concerns such as the efficiency of computation, the complexity of the models, and their interpretability. Academics and practitioners will have a better sense of which CNN architecture will perform best for their specific image classification jobs if they compare AlexNet with GoogLeNet.

REFERENCES

Abdelmaksoud, E., El-Sappagh, S., Barakat, S., Abuhmed, T., & Elmogy, M. (2021). Automatic Diabetic Retinopathy Grading System Based on Detecting Multiple Retinal Lesions. IEEE Access, 9, 15939–15960. https://doi.org/10.1109/ACCESS.2021.3052870

Aquino-Brítez, D., Gómez, J. A., Noguera, J. L. V., García-Torres, M., Román, J. C. M., Gardel-Sotomayor, P. E., Benitez, V. E. C., Matto, I. C., Pinto-Roa, D. P., Facon, J., & Grillo, S. A. (2022). Automatic Diagnosis of Diabetic Retinopathy from Fundus Images Using Neuro-Evolutionary Algorithms. Studies in Health Technology and Informatics, 290, 689–693. https://doi.org/10.3233/SHTI220166

Ashraf, A., Akram, M. U., Sheikh, S. A., & Abbas, S. (2015). Detection of macular whitening and retinal hemorrhages for diagnosis of malarial retinopathy. IST 2015 - 2015 IEEE International Conference on Imaging Systems and Techniques, Proceedings. https://doi.org/10.1109/IST.2015.7294515

Banerjee, I., Crawley, A., Bhethanabotla, M., Daldrup-Link, H. E., & Rubin, D. L. (2018). Transfer learning on fused multiparametric MR images for classifying histopathological subtypes of rhabdomyosarcoma. Computerized Medical Imaging and Graphics : The Official Journal of the Computerized Medical Imaging Society, 65, 167–175. https://doi.org/10.1016/J.COMPMEDIMAG.2017.05.002

Banerjee, S., & Chowdhury, A. R. (2015). Case Based Reasoning in the Detection of Retinal Abnormalities Using Decision Trees. Procedia Computer Science, 46, 402–408. https://doi.org/10.1016/J.PROCS.2015.02.037

Bankman, I. N. (2009). Handbook of Medical Image Processing and Analysis. Handbook of Medical Image Processing and Analysis. https://doi.org/10.1016/B978-0-12-373904-9.X0001-4

Basha, S., & Prasad, K. (2008). Automatic Detection of Hard Exudates in Diabetic Retinopathy Using Morphological Segmentation and Fuzzy Logic.

Bengio, Y., Courville, A., & Vincent, P. (2012). Representation Learning: A Review and New Perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1798–1828. https://doi.org/10.48550/arxiv.1206.5538

Blankenship, G. W., & Skyler, J. S. (1978). Diabetic retinopathy: a general survey. Diabetes Care, 1(2), 127–137. https://doi.org/10.2337/DIACARE.1.2.127

Chetoui, M., Akhloufi, M. A., & Kardouchi, M. (2018). Diabetic Retinopathy Detection Using Machine Learning and Texture Features. Canadian Conference on Electrical and Computer Engineering, 2018-May. https://doi.org/10.1109/CCECE.2018.8447809

Cotton wool spots. COMS Grading Scheme. (n.d.). Retrieved December 15, 2022, from http://webeye.ophth.uiowa.edu/dept/coms/grading/cotton-wool-spots.htm

Deep Learning for Diabetic Retinopathy Diagnosis & Analysis. (n.d.).

Diabetes and Vision Loss | Diabetes | CDC. (n.d.). Retrieved December 15, 2022, from https://www.cdc.gov/diabetes/managing/diabetes-vision-loss.html

DIAGNOSIS RETINOPATHY DISEASE USING GLCM AND ANN. (2018).

Figueiredo, I. N., Kumar, S., Oliveira, C. M., Ramos, J. D., & Engquist, B. (2015). Automated lesion detectors in retinal fundus images. Computers in Biology and Medicine, 66, 47–65. https://doi.org/10.1016/J.COMPBIOMED.2015.08.008

Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., & Yanker, P. (1995). Query by Image and Video Content: The QBIC System. Computer, 28(9), 23–32. https://doi.org/10.1109/2.410146

Franklin, S. W., & Rajan, S. E. (2014). Diagnosis of diabetic retinopathy by employing image processing technique to detect exudates in retinal images. IET Image Processing, 8(10), 601–609. https://doi.org/10.1049/IET-IPR.2013.0565

Gangwar, A. K., & Ravi, V. (2021). Diabetic Retinopathy Detection Using Transfer Learning and Deep Learning. Advances in Intelligent Systems and Computing, 1176, 679–689. https://doi.org/10.1007/978-981-15-5788-0_64/COVER

Giancardo, L. (2011). Automated fundus images analysis techniques to screen retinal diseases in diabetic patients. (Analyse de.

Gondal, W. M., Kohler, J. M., Grzeszick, R., Fink, G. A., & Hirsch, M. (2017). Weakly-supervised localization of diabetic retinopathy lesions in retinal fundus images. Proceedings - International Conference on Image Processing, ICIP, 2017-September, 2069–2073. https://doi.org/10.48550/arxiv.1706.09634

Hajeb Mohammad Alipour, S., Rabbani, H., & Akhlaghi, M. R. (2012). Diabetic retinopathy grading by digital curvelet transform. Computational and Mathematical Methods in Medicine, 2012. https://doi.org/10.1155/2012/761901

Harangi, B., Toth, J., Baran, A., & Hajdu, A. (2019). Automatic screening of fundus images using a combination of convolutional neural network and hand-crafted features. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2699–2702. https://doi.org/10.1109/EMBC.2019.8857073

Hard exudates. COMS Grading. (n.d.). Retrieved December 15, 2022, from https://webeye.ophth.uiowa.edu/dept/coms/grading/hard-exudates.htm

He, K., Sun, J., & Tang, X. (2013). Guided image filtering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(6), 1397–1409. https://doi.org/10.1109/TPAMI.2012.213

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition (pp. 770–778). http://image-net.org/challenges/LSVRC/2015/

Huynh, B. Q., Li, H., & Giger, M. L. (2016). Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. Journal of Medical Imaging (Bellingham, Wash.), 3(3), 034501. https://doi.org/10.1117/1.JMI.3.3.034501

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2022). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. https://github.com/DeepScale/SqueezeNet

Intuitive Guide to Convolution Neural Networks | by Thushan Ganegedara | Towards Data Science. (n.d.). Retrieved December 16, 2022, from https://towardsdatascience.com/light-on-math-machine-learning-intuitive-guide-to-convolution-neural-networks-e3f054dd5daa

Jahiruzzaman, M., & Hossain, A. B. M. A. (2016). Detection and classification of diabetic retinopathy using K-means clustering and fuzzy logic. 2015 18th International Conference on Computer and Information Technology, ICCIT 2015, 534–538. https://doi.org/10.1109/ICCITECHN.2015.7488129

Jitpakdee, P., Aimmanee, P., & Uyyanonvara, B. (2012). A survey on hemorrhage detection in diabetic retinopathy retinal images. 2012 9th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON 2012. https://doi.org/10.1109/ECTICON.2012.6254356

Klein, B. E. K. (2007). Overview of epidemiologic studies of diabetic retinopathy. Ophthalmic Epidemiology, 14(4), 179–183. https://doi.org/10.1080/09286580701396720

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84–90. https://doi.org/10.1145/3065386

Krogh, A. (n.d.). Neural Network Ensembles, Cross Validation, and Active Learning.

Lam, C., Yi, D., Guo, M., & Lindsey, T. (2018). Automated Detection of Diabetic Retinopathy using Deep Learning. AMIA Summits on Translational Science Proceedings, 2018, 147. /pmc/articles/PMC5961805/

Li, B., & Li, H. K. (2013). Automated analysis of diabetic retinopathy images: principles, recent developments, and emerging trends. Current Diabetes Reports, 13(4), 453–459. https://doi.org/10.1007/S11892-013-0393-9

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88. https://doi.org/10.1016/J.MEDIA.2017.07.005

mahmud, tareq, & Bhattacharjee, I. (2020). (THESIS) Diabetic Retinopathy Classification from Retinal Images using Machine Learning Approaches. Thesis for Bachelor of Science in Computer Science & Engineering. https://doi.org/10.13140/RG.2.2.31248.43524/1

Marques, O. (2011). Practical Image and Video Processing Using MATLAB®. Practical Image and Video Processing Using MATLAB®. https://doi.org/10.1002/9781118093467

Messidor - ADCIS. (n.d.). Retrieved December 16, 2022, from https://www.adcis.net/en/third-party/messidor/

Microaneurysms | Retina Vitreous Resource Center. (n.d.). Retrieved December 15, 2022, from https://louisvillediabeticeyedoctor.com/diseases-conditions/hypertensive-retinopathy/dc_hypertensive-ret_fig2/

Mobeen-Ur-Rehman, Khan, S. H., Abbas, Z., & Danish Rizvi, S. M. (2019). Classification of Diabetic Retinopathy Images Based on Customised CNN Architecture. Proceedings - 2019 Amity International Conference on Artificial Intelligence, AICAI 2019, 244–248. https://doi.org/10.1109/AICAI.2019.8701231

Nayak, J., Bhat, P. S., & Acharya, U. R. (2009). Automatic identification of diabetic maculopathy stages using fundus images. Journal of Medical Engineering & Technology, 33(2), 119–129. https://doi.org/10.1080/03091900701349602

O’Shea, K., & Nash, R. (2015). An Introduction to Convolutional Neural Networks. https://doi.org/10.48550/arxiv.1511.08458

PonniBala, M., & Vijayachitra, S. (2012). Computerised Retinal Image Analysis to Detect and Quantify Exudates Associated with Diabetic Retinopathy. International Journal of Computer Applications, 54(2), 7–12. https://doi.org/10.5120/8536-2077

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. https://doi.org/10.1016/J.PROCS.2016.07.014

Qomariah, D. U. N., Tjandrasa, H., & Fatichah, C. (2019). Classification of diabetic retinopathy and normal retinal images using CNN and SVM. Proceedings of 2019 International Conference on Information and Communication Technology and Systems, ICTS 2019, 152–157. https://doi.org/10.1109/ICTS.2019.8850940

Ramasamy, L. K., Padinjappurathu, S. G., Kadry, S., & Damaševičius, R. (2021). Detection of diabetic retinopathy using a fusion of textural and ridgelet features of retinal images and sequential minimal optimization classifier. PeerJ. Computer Science, 7, 1–21. https://doi.org/10.7717/PEERJ-CS.456

Retinal Hemorrhage. COMS Grading. (n.d.). Retrieved December 15, 2022, from https://webeye.ophth.uiowa.edu/dept/COMS/grading/retinal-hemorrhage.htm

Rodríguez, R., Castillo, P. J., Guerra, V., Azuela, J. H. S., Suáreza, A. G., & Izquierdo, E. (2008). A comparison between two robust techniques for segmentation of blood vessels. Computers in Biology and Medicine, 38(8), 931–940. https://doi.org/10.1016/J.COMPBIOMED.2008.06.002

Rohan, T. E., Frost, C. D., Wald, N. J., lecturer Christopher Frost, senior D., Nicholas Wald, lecturer J., & Wald, P. (1989). Prevention of blindness by screening for diabetic retinopathy: a quantitative assessment. BMJ : British Medical Journal, 299(6709), 1198. https://doi.org/10.1136/BMJ.299.6709.1198

Ruder, S. (2016). An overview of gradient descent optimization algorithms. https://doi.org/10.48550/arxiv.1609.04747

Shaban, M., Mahmoud, A. H., Shalaby, A., Ghazal, M., Sandhu, H., & El-Baz, A. (2020). Low-complexity computer-aided diagnosis for diabetic retinopathy. Diabetes and Retinopathy, 133–149. https://doi.org/10.1016/B978-0-12-817438-8.00007-9

Shanthi, T., & Sabeenian, R. S. (2019). Modified Alexnet architecture for classification of diabetic retinopathy images. Computers & Electrical Engineering, 76, 56–64. https://doi.org/10.1016/J.COMPELECENG.2019.03.004

Sidibé, D., Sadek, I., & Mériaudeau, F. (2015). Discrimination of retinal images containing bright lesions using sparse coded features and SVM. Computers in Biology and Medicine, 62, 175–184. https://doi.org/10.1016/J.COMPBIOMED.2015.04.026

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. https://doi.org/10.48550/arxiv.1409.1556

Singh, N., & Tripathi, R. C. (2010). Automated Early Detection of Diabetic Retinopathy Using Image Analysis Techniques. International Journal of Computer Applications, 8(2), 18–23. https://doi.org/10.5120/1186-1648

Sopharak, A., Uyyanonvara, B., Barman, S., & Williamson, T. H. (2008). Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods. Computerized Medical Imaging and Graphics : The Official Journal of the Computerized Medical Imaging Society, 32(8), 720–727. https://doi.org/10.1016/J.COMPMEDIMAG.2008.08.009

Suganya, P., Gayathri, S., & Mohanapriya, N. (2013). Survey on Image Enhancement Techniques. International Journal of Computer Applications Technology and Research, 623–627. https://doi.org/10.7753/IJCATR0205.1019

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going Deeper With Convolutions (pp. 1–9).

Warrier, G. S., & Viji, Dr. K. S. A. (2021). A Method for Enhancing The Images for Early Detection of Scoliosis. International Journal of Engineering Research & Technology, 9(7). https://doi.org/10.17577/IJERTCONV9IS07023

Welikala, R. A., Fraz, M. M., Dehmeshki, J., Hoppe, A., Tah, V., Mann, S., Williamson, T. H., & Barman, S. A. (2015). Genetic algorithm based feature selection combined with dual classification for the automated detection of proliferative diabetic retinopathy. Computerized Medical Imaging and Graphics, 43, 64–77. https://doi.org/10.1016/J.COMPMEDIMAG.2015.03.003

Wong, T. Y., & Sabanayagam, C. (2020). Strategies to Tackle the Global Burden of Diabetic Retinopathy: From Epidemiology to Artificial Intelligence. Ophthalmologica. Journal International d'ophtalmologie. International Journal of Ophthalmology. Zeitschrift Fur Augenheilkunde, 243(1), 9–20. https://doi.org/10.1159/000502387

Zhang, W., Zhong, J., Yang, S., Gao, Z., Hu, J., Chen, Y., & Yi, Z. (2019). Automated identification and grading system of diabetic retinopathy using deep neural networks. Knowledge-Based Systems, 175, 12–25. https://doi.org/10.1016/J.KNOSYS.2019.03.016

APPENDIXES

RESUME

Personal Information

Surname, name	:	
Nationality	:	
Birth date and place	:	
Telephone	:	
Fax	:	
e-mail	:	

Education

Degree	Education Unit	Graduation Date
Master		
Bachelor		
High School		

Work Experience

Year	Place	Title

Foreing Language

Publications

Hobbies