A MODIFIED RESNET-50 CNN MODEL FOR CLASSIFICATION OF EYE DISEASES

2023
MASTER THESIS
COMPUTER ENGINEERING

Sajad Abdlkadhim Abdlhusein ALKHAYKANE

Thesis Advisor
Assist. Prof. Dr. Sait DEMİR
Assist. Prof. Dr. Ashwan A. ABDULMUNEM

A MODIFIED RESNET-50 CNN MODEL FOR CLASSIFICATION OF EYE DISEASES

Sajad Abdlkadhim Abdlhusein ALKHAYKANE

Thesis Advisor

Assist. Prof. Dr. Sait DEMİR

Assist. Prof. Dr. Ashwan A. ABDULMUNEM

T.C.
Karabuk University
Institute of Graduate Programs
Department of Computer Engineering

Prepared as
Master Thesis

KARABUK
August 2023

I certify that in my opinion the thesis submitted by Sajad Abdlkadhim Abdlhusein ALKHAYKANE titled “A MODIFIED RESNET-50 CNN MODEL FOR CLASSIFICATION OF EYE DISEASES” is fully adequate in scope and in quality as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Sait DEMİR ....

Thesis Advisor, Department of Computer Engineering

This thesis is accepted by the examining committee with a unanimous vote in the Department of Computer Engineering as a Master of Science thesis. Aug 25, 2023

Examining Committee Members (Institutions)		Signature
Chairman	: Assist. Prof. Dr. Yusuf Yargı BAYDİLLİ (HU)	Online
Member	: Assist. Prof. Dr. Kasım ÖZACAR (KBU)	....
Member	: Assist. Prof. Dr. Sait DEMİR (KBU)	....

The degree of Master of Science by the thesis submitted is approved by the Administrative Board of the Institute of Graduate Programs, Karabuk University.

Assoc. Prof. Dr. Zeynep ÖZCAN ....

Director of the Institute of Graduate Programs

All the information incorporated in this thesis has been collected and presented under academic regulations and ethical principles. Moreover, I have conscientiously adhered to the demands specified by these regulations and principles, duly acknowledging all sources referenced in this work that are not original to it."

Sajad Abdlkadhim Abdlhusein ALKHAYKANE

ABSTRACT

M. Sc. Thesis

A MODIFIED RESNET-50 CNN MODEL FOR CLASSIFICATION OF EYE DISEASES

Sajad Abdlkadhim Abdlhusein ALKHAYKANE

Karabuk University
Institute of Graduate Programs
The Department of Computer Engineering

Thesis Advisor:
Assist. Prof. Dr. Sait DEMİR
Assist. Prof. Dr. Ashwan A. ABDULMUNEM

June 2023, 58 pages

Delay in diagnosing and treating eye illnesses causes blindness in millions worldwide. In response, there has been a push towards an efficient automated detection technique in medical imaging, such as retinal fundus images. In this research, the system has been proposed to automatically distinguish between healthy and unhealthy retinal fundus images given a set of training images based on pre-trained deep learning architectures. A preprocessing based on adaptive histogram equalization followed by morphological operations is demonstrated to provide an improved class distinction between normal and diseased images compared to using the raw images alone.

Specific effective deep convolutional neural network (CNN) based architectures are developed using the pre-trained weights gained through transfer learning. Comparing the results of several deep CNN architectures - ResNet50, Inceptionv3, GoogleNet, and MobileNet, A new modified model has been created by adding new layers on top of the last output layer to the model with the highest accuracy (ResNet50). Dense layers are used to downsize the data and improve performance. The final layer is used for predicting image classes using softmax activation to enhance the classification performance significantly. Extensive testing on a big ophthalmology dataset reveals encouraging results. The various diseases represented in the fundus images and the situations in which they were collected strong evidence for the method's viability in clinical settings, Where the proposed models (ResNet50, InceptionV3, GoogLeNet, and MobileNet) achieved an accuracy of 99.1%, 95.3%, 94.7 and 93.1, respectively. In comparison, the modified model achieved an accuracy of 99.25%.

Keywords **:** Deep learning, CNN, Pre-Trained Deep Learning, Ophthalmology.
Science Code : 92432

ÖZET

Yüksek Lisans Tezi

GÖZ HASTALIKLARININ SINIFLANDIRILMASI İÇİN MODİFYE EDİLMİŞ RESNET-50 CNN MODELİ

Sajad Abdlkadhim Abdlhusein ALKHAYKANE

Karabük Üniversitesi
Lisansüstü Eğitim Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı:
Dr. Öğr. Üyesi Sait DEMİR
Dr. Öğr. Üyesi Ashwan A. ABDULMUNEM
Haziran 2023, 58 sayfa

Göz hastalıklarının teşhis ve tedavisindeki gecikmeler dünya çapında milyonlarca insanın kör olmasına neden olmaktadır. Bu nedenle, retina fundus görüntüleri gibi tıbbi görüntüler kullanan etkili bir otomatik algılama tekniğine yönelik ihtiyaç olmuştur. Bu araştırmada, önceden eğitilmiş derin öğrenme mimarilerine dayalı bir dizi eğitim görüntüsü verilen sağlıklı ve sağlıksız retina fundus görüntülerini otomatik olarak ayırt etmek için bir sistem önerilmiştir. Uyarlanabilir histogram eşitleme ve ardından uygulanan morfolojik işlemlere dayanan bir ön işlemenin, yalnızca ham görüntüleri kullanmaya kıyasla normal ve hastalıklı görüntüler arasında gelişmiş bir sınıf ayrımı sağladığı gösterilmiştir.

Transfer öğrenme yoluyla elde edilen önceden eğitilmiş ağırlıklar kullanılarak etkili derin konvolüsyonel sinir ağı (CNN) tabanlı mimariler geliştirilmiştir. ResNet50, Inceptionv3, GoogleNet ve MobileNet gibi çeşitli derin CNN mimarilerinin sonuçları karşılaştırılarak, en yüksek doğruluğa sahip olan modelin (ResNet50), son çıktı katmanının üstüne yeni katmanlar eklenerek yeni bir modifiye model oluşturulmuştur. Verileri küçültmek ve performansı artırmak için yoğun katmanlar kullanılmıştır. Son katman, sınıflandırma performansını önemli ölçüde artırmak için softmax aktivasyonunu kullanarak görüntü sınıflarını tahmin etmek için kullanılmıştır. Büyük bir oftalmoloji veri kümesi üzerinde yapılan kapsamlı testler önemli sonuçlar ortaya koymaktadır. Fundus görüntülerinde temsil edilen çeşitli hastalıklar ve bunların içinde bulunduğu durumlar, yöntemin klinik uygulamalarda uygulanabilirliği için güçlü kanıtlar sunmuştur. Önerilen modeller (ResNet50, InceptionV3, GoogLeNet ve MobileNet) sırasıyla %99,1, %95,3, 94,7 ve 93,1 doğruluk oranlarına ulaşmıştır. Diğer modeller ile karşılaştırıldığında, modifiye edilmiş model ile %99,25'lik doğruluk oranı elde etmiştir.

Anahtar Kelimeler : Derin öğrenme, CNN, Önceden Eğitilmiş Derin Öğrenme, Oftalmoloji.

Bilim Kodu : 92432

ACKNOWLEDGMENT

Finally, I can only pray to God Almighty to grant us guidance and chastity Those who lack a strong will are more likely to spend more time in the stations of exhaustion, failure, and despair than those who do. The first to whom to give thanks and praise during the night and the end of the day is the Highest, the Kahar, the First and the Last, the Apparent and the Inward, Who showers us with his countless blessings, bless us with His endless provision and illuminates our paths. May God's blessings, mercy, and peace be with Muhammad bin Abdullah and his loved ones as He sent His servant and Messenger among us. He dispatched him with his unmistakable Qur'an to fill us in on our ignorance and encourage us to seek out information wherever it might be found. All praise and thanks are due to God for granting us success and inspiring us with patience and eagerness for that which confronts us to accomplish this humble work.

Thanks go to every teacher who helped me with their knowledge from the first stages of this study until now. I also extend a word of thanks to Asst. Prof. Dr. Sait DEMİR and Asst. Prof. Dr. Ashwan A. Abdulmunem. I wish them all the best and much success. I also thank the Head of the Department of Computer Engineering and his staff for the assistance they have given me. Thanks also go to my colleagues, wishing them the best ranks and success This labor of love is dedicated to my lovely mother, sisters, and brother, who gave me life, hope, and education infused with curiosity and curiosity's offspring, wisdom, patience, and loyalty. As I prepare to return to my two countries, Iraq and Turkiye, and whoever had a role, from near or far, in completing this study, I ask the Lord Almighty to reward everyone with the best reward in this world and the Hereafter. Finally, I can only pray to God almighty to grant us guidance, purity, wealth, health, and safety, and who inspires me with more determination and strength to upgrade my scientific thesis and study for the Ph.D., Inshallah. Thanks a lot.

CONTENTS

Page

APPROVAL ii
ABSTRACT iv
ÖZET vi
ACKNOWLEDGMENT viii
CONTENTS ix
LIST FIGURES xii
LIST OF TABLES xiii
ABBREVITIONS INDEX xiv

PART 1 1
INTRODUCTION 1
1.1. MOTIVATIONS 2
1.2. THESIS PROBLEM 3
1.3. AIMS OF THE STUDY 3
1.4. RELATED WORK 4
1.5. THESIS ORGANIZATION 6

PART 2 8
THEORETICAL BACKGROUND 8
2.1. DIAGNOSIS OF EYE 8
2.2. DEEP LEARNING 9
2.3. CNN CONCEPT 9
2.3.1. Convolutional Layers 12
2.3.2. Non-Linear Layer 13
2.3.2.1. Pooling Layer (PL) 15
2.3.2.2. Fully Connected Layers 16
2.4. HYPER AND MODEL PARAMETERS 17
2.5. LOSS FUNCTION 17
2.6. CHALLENGES AND NETWORK EDUCATION 18

Page

2.6.1. Parameter Initialization ... 20
2.6.2. Optimizer Selection ... 21
2.6.3. Regularization to CNN ... 22
2.6.3.1. Batch Normalization ... 22
2.6.3.2. Dropout ... 23
2.6.4. Early Stopping ... 23
2.6.5. K-Fold Validation ... 23
2.6.6. Pre-trained Convolutional Neural Network (CNN) Models ... 24
2.7. EVALUATION MATRICES ... 24
2.7.1. Accuracy ... 24
2.7.2. F1 Score ... 25
2.7.3. Confusion Matrix ... 25
2.7.4. AUC-ROC ... 25
2.8. SUMMARY ... 26

PART 3 ... 27
METHODOLOGY ... 27
3.1. OVERVIEW ... 27
3.2. THE PROPOSED WORK ... 27
3.3. TECHNICAL REQUIREMENTS ... 28
3.4. DATASET ... 28
3.5. PREPROCESSING ... 29
3.6. TRANSFER LEARNING ... 32
3.7. DATA AUGMENTATION ... 33
3.8. MODELS TRAINING ... 33
3.8.1. Pre-trained (CNN) Models ... 34
3.8.1.1. ResNet-50 ... 35
3.8.1.2. Modified ResNet-50 ... 35
3.8.1.3. Inceptionv3 ... 35
3.8.1.4. GoogLeNet ... 36
3.8.1.5. MobileNet ... 37
3.8.2. Preparing the Structure of the Proposed Models ... 37

Page

PART 4 41
RESULTS 41

PART 5 46
DISCUSSION 46

PART 6 49
CONCLUSION 49
    6.1. UPCOMING WORK 49

REFERENCES 50

RESUME 58

LIST FIGURES

Page

Figure 2.1. An example artificial neural network with a hidden layer 11
Figure 2.2. CNN at their most fundamental level of design 11
Figure 2.3. Rectifier functions 13
Figure 2.4. Activation Function Forms 14
Figure 2.5. Average Pooling and Max Pooling Operation... 16
Figure 2.6. Fully Connected Layers Architecture 16
Figure 2.7. Overfitting and Underfiting 18
Figure 2.8. Gradient Descent 22
Figure 3.1. Training process of a pre-trained model. 28
Figure 3.2. Examples of CFP pictures showing eye diseases. 29
Figure 3.3. pictures of the enhancement effect. 30
Figure 3.4. Data Augmentation. 33
Figure 4.1. Performance comparison of the proposed models without enhancement technique and augmentation. 42
Figure 4.2. Confusion matrix for classification using a) ResNet50 b) Inceptionv3. 43
Figure 4.3. Confusion matrix for classification using a)GoogLeNet b)MobileNet .. 43
Figure 4.3. Performance comparison of the proposed models with enhancement technique and augmentation. 44
Figure 4.4. Confusion matrix for classification using a) ResNet50 b) Inceptionv3. 44
Figure 4.5. Confusion matrix for classification using a)GoogLeNet b)MobileNet .. 45
Figure 4.6. Comparison of the confusion matrix of the (M-ResNet-50) model, a) After adding optimization techniques b) Before adding optimization techniques. 45
Figure 5.1. Class-based performance indicators are shown. 48

LIST OF TABLES

Page

Table 2.1. Hyper Parameters and Parameters 17

Table 3.1. The structure of all proposed deep models 38

Table 4.1. Performance evaluation of models without enhancement technique and augmentation.. 42

Table 4.2. Performance comparison of the proposed models with enhancement technique and augmentation. 44

Table 5.1. Class-based performance indicators are shown. 47

Table 5.2. Analysis of the systems' performance concerning those of similar systems in existing studies.. 48

ABBREVITIONS INDEX

CNN	: Convolutional Neural Network
RELU	: Rectified Linear Units
AI	: Artificial Intelligent
FC	: Fully Connected Layer
OCT	: The optical coherence tomography
CFP	: Color Fundus Photography
OIH	: Ophthalmic Image Handling
IDRiD	: Indian Diabetic Retinopathy Image Collection
HRF	: High-Resolution Fundus
M-ResNet-50	: Modified ResNet-50 Model

PART 1

INTRODUCTION

The retina is a delicate layer that lines the back of the eye. It processes and recognizes objects by transforming incoming light into nerve impulses transmitted to the brain's optical cortex [1]. Failure to diagnose eye problems promptly might result in permanent vision loss due to retinal damage. As a result, getting the right treatment requires a prompt diagnosis [2], [3]. CFP and optical tomography (OCT) are the most common methods for assessing retinal pictures to detect eye disorders [4]. The optical coherence tomography (OCT) technology creates images to quantify retinal thickness for eye disease diagnosis, but it comes at a high cost. Retinal abnormalities can be spotted with the help of CFP images of the inside of the eyes. CFP is more accurate, less expensive, and non-invasive in diagnosing eye disorders than other approaches. Continuous assessments with the CFP approach are highly suggested for all adults [5]. Ophthalmologists rely on fundus images to assess and diagnose a variety of eye conditions, including but not limited to cataracts, diabetic retinopathy, and glaucoma [6]. Aging, immunological abnormalities, genetics, trauma, metabolic problems, radiation, and other causes of lens opacity can bring on cataracts. Because the lens protein has become desaturated, the person has developed a cataract and can no longer see [7]. Diabetic retinopathy is caused by persistently elevated blood sugar. Blindness results from high blood glucose because of vascular leaking in the retina [8]. Early detection of retinal disease allows for treatment of all stages. Optic disc atrophy, optical domain defects, depression, and vision loss are all symptoms of glaucoma, a potentially blinding eye condition. Glaucoma is caused by high blood pressure and inadequate blood flow to the optic nerve [9]. Since their initial symptoms are identical, ophthalmologists face difficulty making a correct and timely diagnosis in the early stages of most eye illnesses. In addition, it takes a lot of time and effort to diagnose many photos created using CFP manually. In addition, there is a lack of ophthalmologists in low-income regions, making manual diagnosis impossible. Therefore, automatic diagnosis is desperately needed to boost diagnostic precision, relieve pressure from ophthalmologists, and back up their decisions. Computer-aided diagnosis is an efficient way for automatic illness identification, which is especially important given the rising number of patients with eye problems and the shortage of ophthalmologists. Both the academic and business sectors have committed significant resources to the development of automatic methods, and in particular CNN models, for interpreting biological pictures. CNN models have excellent accuracy in illness classification for ocular disorders. CNN's goal is to aid medical professionals by easing their workload and speeding up the process of disease classification. Preventing blindness through advanced stages of eye disease requires prompt diagnosis. Fundus image analysis has mainly been studied in the context of early diagnosis of a specific disease. This research aims to create computerized models that can examine photographs of the back of the eye (the fundus) and label different types of eye illnesses based on their appearance. The early stages of several eye illnesses share similar clinical and vital indicators. Hence the Pre-trained CNN models focused on extracting and combining subtle and non-obvious characteristics.

1.1. MOTIVATIONS

Human observers may have difficulty distinguishing between different forms of eye disorders in the early stages because the symptoms and visual manifestations might be extremely similar. This is where instrumental diagnosis with the help of computers comes in.

To aid in diagnosing, categorizing, and monitoring various eye disorders, computer-assisted diagnostic approaches analyze fundus images using cutting-edge image processing, pattern recognition, and machine learning algorithms. These methods can enhance the precision and timeliness of diagnosis while providing objective and quantitative evaluations to help guide clinical decision-making.

1.2. THESIS PROBLEM

Significant obstacles exist to study and solution development in fundus photography-based diagnosis of eye diseases.

One of these difficulties is a lack of data sets, which could prevent reliable diagnostic tools from being developed and tested. This difficulty inspired this research, and it seeks to solve several concerns about collecting data for detecting and diagnosing eye illnesses using (CNN) models. Some examples of such inquiries concern the use of (CNNs) for feature extraction and classification and the collection, organization, and enhancement of data sets.

The study aims to improve our understanding of the difficulties in detecting eye diseases by answering these questions and shedding light on possible remedies. The ultimate goal is to aid in the development of more precise, reliable, and timely diagnostic tools for eye disease, which can boost patient satisfaction and facilitate earlier treatment.

- Should pre-trained (CNN) models be used with fundus pictures to diagnose eye diseases?
- How can we best enhance model performance, and what resources are available?
- Which model has the best chance of outperforming the others and achieving high accuracy?

1.3. AIMS OF THE STUDY

The following are the results of this study:

- Provide a summary of the relevant literature.
- Accurately identifying and diagnosing eye illnesses from fundus images with great performance and accuracy.

- Show that the pre-trained (CNN) model can correctly categorize these images.

1.4. RELATED WORK

The early detection of COD relies on color fundus pictures obtained by fundus cameras. Improving diagnostic performance mainly depends on the changes to the input color fundus images before feeding them to the deep neural models. Color fundus photos can benefit from preprocessing to improve the accuracy of DL-based COD diagnosis by eliminating sources of noise, including inconsistent lighting, low contrast, irrelevant details, etc. Generic digital image preprocessing approaches can be applied to any COD documented in the literature. However, some preprocessing techniques (hence referred to as "specific preprocessing techniques") are limited in their applicability to a subset of ocular illnesses.

To categorize disorders of the eye using many markers, DCNet was proposed by Junjun et al. [10]. Three sequential nodes comprise the network's architecture: a feature extraction node for fundus images, a spatial correlation node, and a classification node. To identify multi-class fundus images for an Ocular Disease Intelligent Recognition (ODIR) dataset containing changes in anatomical components such as the optic disc, macula, and blood vessels, Neha et al. [11] developed four CNN models with different optimizers. VGG16 trained with SGD optimizers performed best when classifying fundus images compared to other models. A deep learning model with a mixture loss function was proposed for illness detection in fundus images by Xiong et al. [12]. The categorization of the eye illness dataset was enhanced by introducing a deep learning model that combines the loss function with focal loss robustness. Retinal vasculature can be segmented from CFP pictures using the method provided by Kai et al. [13], which is based on the CNN model. A probability map is generated using the CNN model's loss function. The feature maps were used to construct a convolutional neural network model that could accurately extract data about retinal blood vessels. Clement et al. [14] suggested using a convolutional structure to train the Messidor dataset and reinforce it with supervised learning. At the same time that lesions are detected, they are split into their red and brilliant components. Slices with red and glowing lesions are generated and checked for accuracy at the pixel level. The area Under Curve (AUC) was 83.9%. Therefore, the system performed admirably. CNN and support vector machine (SVM) were proposed by Rahul et al. [15] to micro-read fundus pictures for cataract identification. After extracting features from the source photos, the dataset grew in size. CNN achieved an accuracy of 87.08%, whereas SVM achieved an accuracy of 87.5%.

This study aims to classify a dataset of eye diseases using feature extraction and preprocessing techniques. Four approaches were developed to classify color fundus photography (CFP) images to aid in diagnosing eye diseases.

A pre-trained Convolutional Neural Network (CNN) algorithm is used; comparing the results of several deep CNN architectures, ResNet50, Inceptionv3, GoogLeNet, and MobileNet, can significantly boost the classification performance.

By employing these hybrid techniques, combining the strengths of pre-trained CNN algorithms and feature fusion and dimensionality reduction methods, this study aimed to achieve more accurate and reliable classification results for diagnosing eye diseases using CFP images.

Although the choice of preprocessing approaches relies on the nature and needs of a specific ocular disease diagnosis, to the best of our knowledge, no extensive examination of the influence of preprocessing on the performance of DL models has yet been conducted. More attention should be paid in this study to qualitatively analyzing the preprocessing methods that have been implemented. The input fundus images must be preprocessed, which uses computing resources. The prediction performance may be enhanced by some preprocessing techniques while being negatively affected by others. Some preprocessing methods may be optimal for treating certain kinds of eye problems. But in real-world settings, especially those with a wide range of eye illnesses, they may fall short of expectations. Therefore, there is much room for improvement in quantitatively assessing the benefits and drawbacks of such methods for automated chronic disease diagnosis.

A growing number of studies have used CNN-based models for COD detection recently. CLAHE preprocessed RGB fundus pictures of size (32 32) were predicted using a shallow CNN proposed by Islam et al. [51]. Overfitting occurs in neural networks when working with images of such low resolution because the most important details are lost. Histogram equalization was used by Wang et al. [52] on (448 x 448) RGB and grayscale photos. Following their training on these photos, two EfficientNet-B3 [53] networks produced average predictions for the final prediction. More processing power is needed during the training and inference stages since more training parameters must be used to accommodate larger input images. Left and right fundus RGB pictures of size (256 256) were joined by Gour and Khanna [54], and COD was categorized with a sigmoid activation function. Eight different classifiers were employed to classify COD after characteristics from the left and right eyes' CNN networks were fused by Li et al. Left and right fundus RGB pictures (448 448) were used to generate ResNet features, which were then improved using a spatial correlation module (He et al., [25, 56]). The authors cross-validated their approach on 1166 fundus pictures by randomly splitting the training ODIR dataset. By using a teacher network trained on fused features from both eye pictures and the 102 diagnostic keywords, He et al. [57] were able to improve upon previous attempts [25, 56]. When it is impractical to gather fundus images of both eyes during the same visit, the resulting CDSS should be adaptable enough to handle the situation. Due to the integration of several input fundus photos, the information gained by the CNN models for predicting the output COD is not easily shown. Ophthalmologists have a hard time trusting automated systems, so CDSS should offer a clear, defensible verdict (even if it's wrong) rather than an extremely accurate but opaque one. To solve these issues, we present an automated RoI segmentation and ensemble technique that trains CDSSs to identify even the tiniest lesions, leading to more accurate early identification of COD and the ability to examine the precise features of the input images that led to this detection.

1.5. THESIS ORGANIZATION

This thesis is contained six sections. PART 1 contains the introduction, motivation letter, problem, and objective of the thesis, in addition to related works. PART 2, Which contains the theoretical background that talks about the science of artificial intelligence, passing through deep learning, then convolutional neural networks and their contents; the third part also talks about the methodology involved in the proposed technology and pre-treatment and then ends with a detailed structure of the proposed models, The fourth part dealt with the results, the fifth part discussed the results. It ended with the conclusion in PART 6. Furthermore, the references and the material used to construct this thesis report are listed.

PART 2

THEORETICAL BACKGROUND

This chapter will look at the criteria used to classify eye diseases. In addition to covering issues related to network training, it provides an overview of deep learning technology as described in two convolutional neural networks (CNN) models. Next, we'll investigate software testing and metrics and introduce a design model approach.

2.1. DIAGNOSIS OF EYE

The retina, a tiny layer at the back of the eye, is very important to vision because it receives light and converts it into electrical impulses sent to the brain. Early detection and treatment can prevent permanent vision loss from eye illnesses. Color fundus photography (CFP) and optical coherence tomography (OCT) are typical methods for analyzing retinal images to diagnose and treat eye diseases.

CFP is the most popular option because it is non-invasive, inexpensive, and simple. In particular, it aids in diagnosing retinal abnormalities by providing clear pictures of the eye's internal structures. Optometric coherence tomography (OCT), on the other hand, generates images that assess the retinal thickness, which aids in identifying eye diseases. However, compared to CFP, OCT is more costly [87].

Many people suffer from ocular illnesses such as glaucoma, retinopathy, and cataracts. There is a growing demand for automatic diagnosis utilizing computer-aided approaches to improve diagnostic accuracy and relieve the workload of ophthalmologists. Models based on Convolutional Neural Networks (CNNs) have performed well in classifying ocular disorders. These models aim to speed up the process of disease classification by assisting doctors and specialists in making decisions.

2.2. DEEP LEARNING

To model neurons in the human brain is the focus of deep learning, a branch of machine learning [33]. The algorithms used in deep learning serve two purposes: principles of data mining and the simulation of technical processes [16]. Machine learning focuses on data learning and pattern recognition [17]. Analytical pondering and information acquisition are at the heart of the machine learning subfield [18]. Deep learning can extract features automatically, and accurate results can be returned. However, more data types must be preprocessed because features must be obtained by hand in machine learning. Deep learning can extract features from a wide range of dimensions, regardless of whether or not they are immediately apparent. Clinical decision-making based on deep learning could one day mimic human experts. Medical X-ray detection is just one use of deep learning architectures employed in fields like computer vision and image processing [19]. The medical business has utilized deep learning to understand better optimal outcomes, disease progression probabilities, and the rapid creation of reliable medical images [20].

2.3. CNN CONCEPT

In computer vision, CNNs (short for Convolutional Neural Networks) are a popular deep learning method for tasks including image categorization, object recognition, and picture segmentation. Images, which have a gridlike layout, are ideal data for CNNs to analyze [21] CNNs rely heavily on a concept known as the convolution process. In convolution, the input data is processed through a series of filters (sometimes called kernels or feature detectors). These filters are implemented as tiny matrices that undergo element-wise multiplication and summing operations as they are convolved (slid) over the input data. This convolutional procedure yields a feature map that draws attention to distinct characteristics within the original data [22]. CNNs are constructed from "convolutional layers," often of several filters. Each filter is designed to identify a different type of feature or pattern in the input data at a different place in space, such as edges, corners, or textures. The filters convolve the input data, producing feature maps sensitive to regional variations. A CNN's ability to learn a hierarchy of progressively complicated features is made possible using many filters [23]. CNNs typically contain pooling layers after convolutional layers because they assist in minimizing the spatial dimensions of the feature maps while maintaining relevant information. When doing a pooling process, it is common practice to save only the maximum or average value inside a certain region of the feature map and to delete all other values. This down-sampling procedure decreases the network's computational cost and improves translation invariance (i.e., pattern recognition regardless of input position) [24]. One or more fully connected layers, like those used in traditional neural networks, are typically used after the convolutional and pooling layers in CNNs. Predictions can be made using the information in these fully connected layers, which were trained on the more abstract attributes. To get the appropriate output, such as class probabilities in image classification, CNNs frequently use activation functions like softmax in the final layer [25]. To train a CNN, the network optimizes its parameters (weights and biases) by minimizing the gap between its predictions and the true labels in the training data. This process is known as back-propagation. The network can learn and generalize beyond the specific instances used during training because the parameters are updated iteratively to minimize the loss function [26].

Additionally, inputs of varying sizes can be processed by convolution. In typical neural network layers, each output unit communicates with each input unit [27]. See Figure 2.1. In general, CNNs have made tremendous strides in computer vision by allowing state-of-the-art performance on various tasks and automatically extracting important features from images [28]. CNNs can effectively collect and recognize complicated patterns in visual input because of their hierarchical architecture, which uses convolution and pooling operations [29].

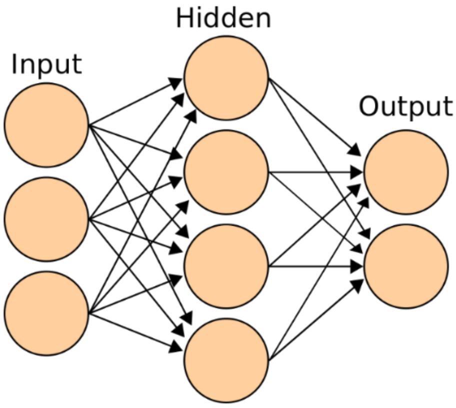

Figure 2.1. An example artificial neural network with a hidden layer [30].

Subtle changes are made to an input picture or feature bus (with one input node for each input) by a neural networks hidden layers and non-linear activation functions. Each sublayer has a group of neurons in constant contact with all the cells below them. The output layer is the final stage of a fully connected neural network and displays the network's newly acquired categorization abilities [31]. Scaling a neural network correctly is challenging because of factors such as image resolution, the number of inputs, the number of weights, several hidden layers, and the total number of nodes [32]. Convolutional neural networks (CNNs) with a forward propagation architecture have lately shown promise in various computer vision applications, as seen in Figure 2.2. [41],[52].

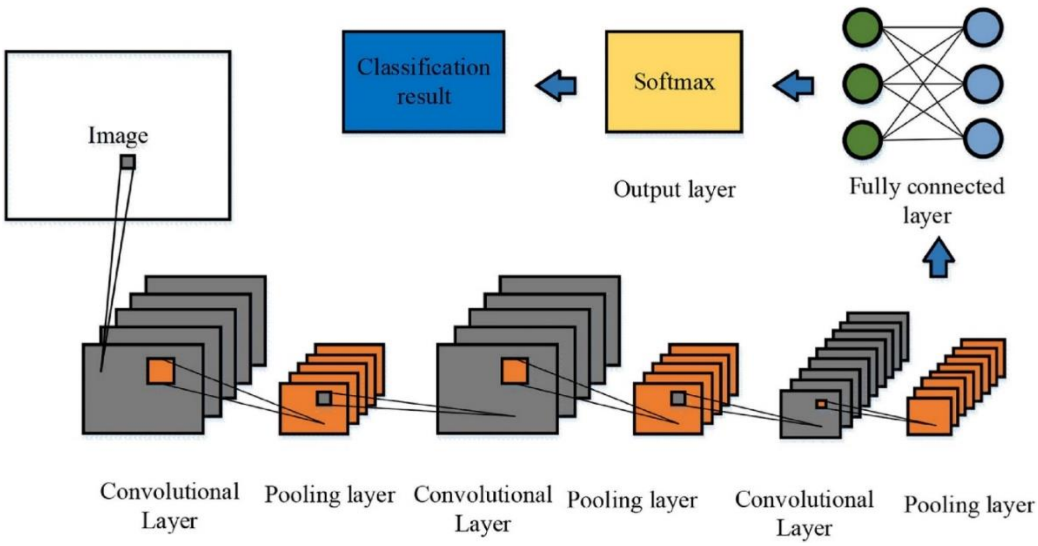

Figure 2.2. CNN at their most fundamental level of design [35].

The layers and structure of a (CNN) are as follows[36]:

2.3.1. Convolutional Layers

Most of the work done by a CNN occurs in its convolutional layer. It's a crucial part of the network that works with data through a filter and a feature map. Consider the input a pixelated, three-dimensional color image with the same height, width, and depth as a traditional RGB image. Each pixel in an image is checked by a feature detector (also called a kernel or filter) to see if it contains the feature of interest. One term for this operation is "convolution." [25], [26]. The feature detector uses a 2D weighted array to characterize a certain area of the picture. The filter size determines the receptive field's size, which is typically a 3x3 matrix but is modifiable. The filter is applied to a region of the image, and the resulting image is the dot product of the input pixels and the filter. The output array receives the dot product as input. After that, the filter advances by one stage, and the procedure is repeated throughout the whole picture. The output may be one of many maps representing the convolved dot products of the input and filter: feature maps, activation maps, and convolved features [25], [26]. Rectified Linear Unit (ReLU) adjustments to the feature map after each convolution operation are a common way for a CNN to impart non-linearity to the model. This transformation is performed after the first convolution layer, and if other convolution layers follow, it enables the succeeding layers to read the information from the earlier layers' receptive fields. The CNN resembles a tree because of its hierarchical nature [37]. For example, let's consider the problem of identifying bicycles in pictures. The parts of a bicycle are viewed as simpler patterns, including the frame, handlebars, wheels, and pedals. The CNN's feature hierarchy is constructed by classifying these components as lower-level patterns and the entire bicycle as a higher-level pattern, as shown in Figure 2.3.

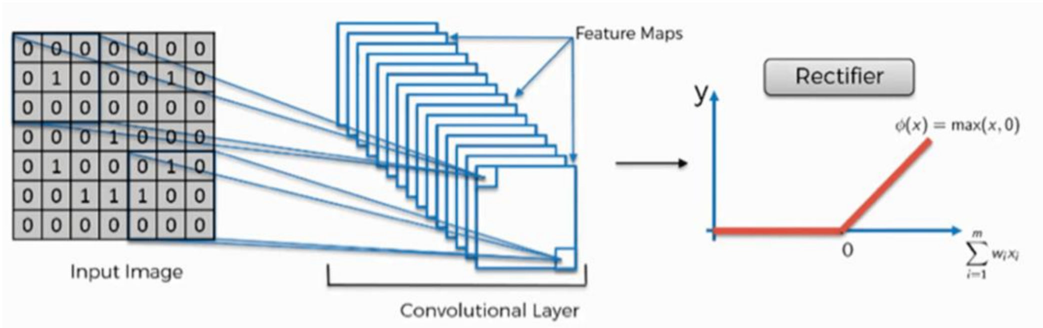

Figure 2.3. Rectifier functions [38].

2.3.2. Non-Linear Layer

A non-linear function can address non-linear issues by retaining and mapping the activation function of an active neuron [39]. An activation function is used in neural network models to calculate the output value from the input value (which is the weighted sum of the inputs to the neurons). An activation function is utilized to increase the model's expressiveness and the significance of the neural network's AI. A non-linear activation layer is added after each trainable layer (a layer with a weight, such as a convolutional or fully connected layer) in a convolutional neural network (CNN). Creating a deep neural network relies heavily on the activation function [40]. The result of this linear process model convolution is then subjected to a non-linear activation function. Mathematical models of neural activity in the body have traditionally been based on smooth non-linear functions. The sigmoid, ReLU, tanh, and softmax activation functions are depicted in Figure 2.4 [41].

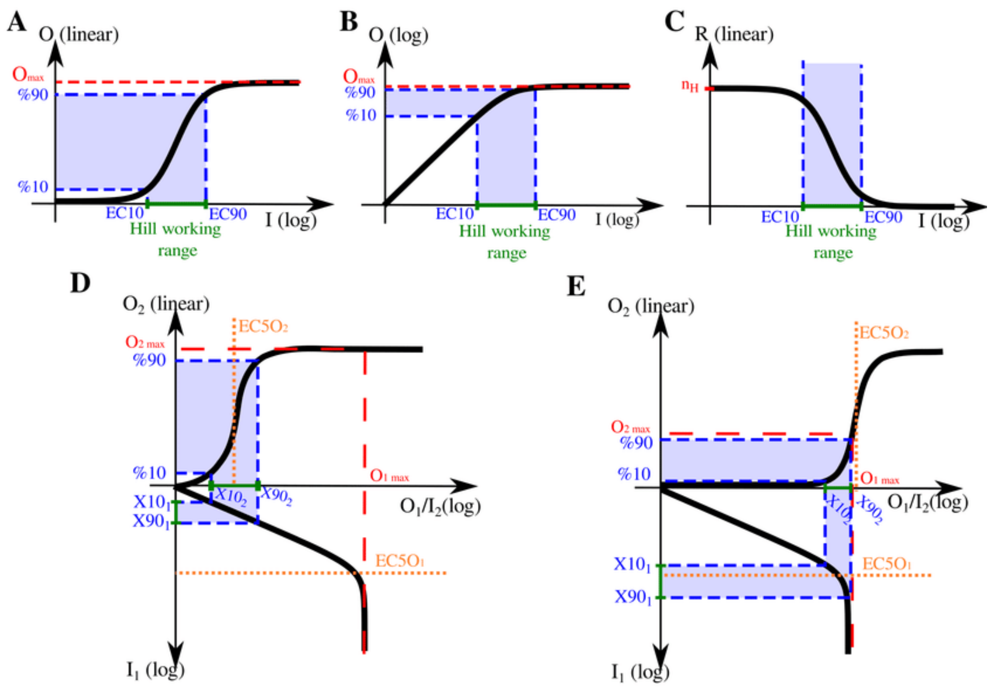

Figure 2.4. Activation Function Forms [42].

The Rectified Linear Unit (ReLU) is widely used in contemporary network architectures as an activation function. It may improve the effectiveness of CNNs in general. The fact that it is linear in the plus dimension but zero in the negative dimension suggests that a very small number of neurons are influenced at once, leading to a highly efficient and evenly dispersed network. The asymmetry of the function is likely the root cause of its non-linearity. In contrast to sigmoid activations, gradient non-saturation is not conceivable for half the natural line since its gradient is zero with positive linearity. The ReLU activation function has the potential for considerable benefits, but it also has potential downsides. The weights may be adjusted during the error back-propagation phase if a ReLU function is employed to feed a greater gradient to prevent the neuron from being activated again. The term "Dying ReLU problem" describes this predicament [43]. The equation (2.4) below from [43] characterizes this function:

$$f(x)_{ReLU=max(0,x)} \tag{2.1}$$

The SoftMax function extends the sigmoid activation function used for multi-class classification. After the final completely connected layer, CNN uses SoftMax. There is a 'S' form in the curve of this function. Mathematically, the sigmoid function looks like this: [37]:

$$s(x_i)=\frac{e^{x_i}}{\sum_{j=1}^{N} x_j} \tag{2.2}$$

Input values (real numbers) are constrained to lie between -1 and 1 using the Tanh activation function. The formula for Tanh is [37] in decimal notation.

$$f(x)_{tanh}=\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{2.3}$$

Where the input value e x has a conventional exponential function, denoted by x., The typical exponential function of an argument x is denoted by e-x

2.3.2.1. Pooling Layer (PL)

In Convolutional Neural Networks (CNNs), the feature maps created by convolutional layers are downsampled using a pooling layer. It applies operations like max pooling or average pooling within each window, decreasing the number of spatial dimensions required for the input. Dimensionality reduction is accomplished using pooling, which enables translation invariance for pattern recognition independent of orientation. Pooling layers are typically placed after convolutional layers to maximize computing efficiency and feature extraction in computer vision tasks [44].

In max pooling, the filter repeatedly searches the input for the pixel with the greatest value. As an aside, this technique is favored above the norm for pooling [44].

As the input is scanned, the filter averages the values inside the receptive field and sends that value to the output array. While the pooling layer's benefits to the CNN are considerable, the data loss it causes is not. They help streamline operations, increase output, and lessen the likelihood of overfitting. Figure 2.5 demonstrates this [45].

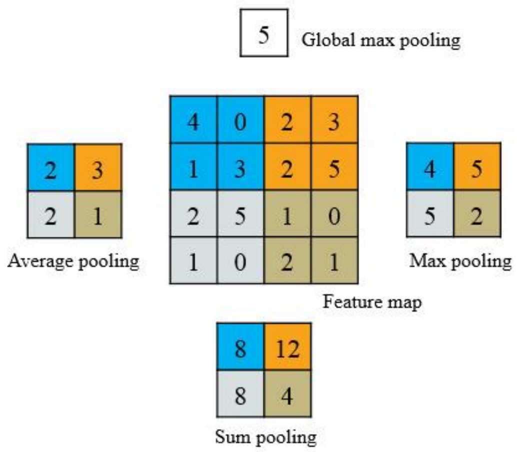

Figure 2.5. Average Pooling and Max Pooling Operation [45].

2.3.2.2. Fully Connected Layers

Neural networks, such as CNNs, rely heavily on fully connected layers, sometimes called dense layers. All of the neurons in these layers are interconnected, with all the neurons in the layers below and above. They act as a bridge between feature maps and output predictions, making it easier to grasp intricate correlations [46]. Non-linearity is introduced using activation functions, and parameters are tuned as the network is trained. At the very end of the network, in the fully linked layers, the learned high-level features are used to make predictions [47]. Figure 0.6 Shawn the Architecture of Fully Connected Layers.

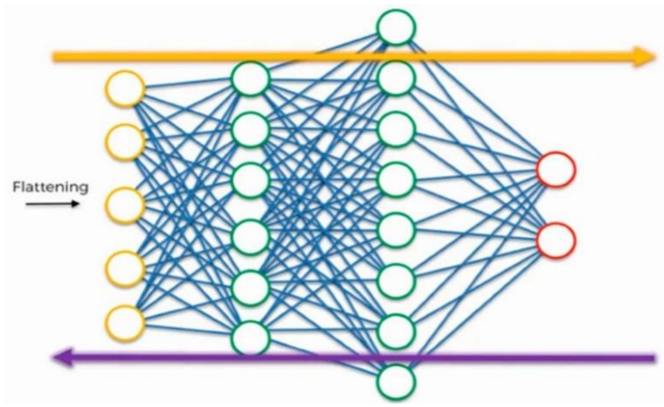

Figure 2.6. Fully Connected Layers Architecture [47].

2.4. HYPER AND MODEL PARAMETERS

A CNN can be defined by its normal and hyperparameters. Parameters are the movable parts of a model that can be tweaked to achieve different results. The training process can be arranged (predefined) according to a set of parameters called hyper-parameters [48]. The hyper-parameters are not automatically learned during training and must be set manually. Improvements in efficiency and output can be seen after the hyper-parameters have been fine-tuned. Table 2.1 [48]. Hyper parameters such as activation function, learning rate, hidden nodes, hidden layers, periods...etc. are listed.

Table 2.1. Hyper Parameters and Parameters [49].

Layers	**Parameter**	**Hyperparameter**
Convolutions	Kernels	Adjustable kernel size, number of kernels, stride, activation functions, and padding
Pooling	-	Size, number of kernels, stride, activation functions, and padding can all be altered.
Fully Connected	Weights	The activation function and the total number of weights
Other		The efficiency of the model is affected by several variables. These include the model's framework, loss function, optimizer, learning rate, epochs, dataset splitting, mini-batch size, weight initialization, and regularization.

2.5. LOSS FUNCTION

Training machine learning and deep learning models requires a loss function, sometimes called a cost or objective function. It measures how far off a model is from producing the actual or expected result. During training, the model attempts to get a smaller value for the loss function, representing this difference [50]. For each input example, a loss function will take both the model's anticipated output (commonly symbolized by ŷ) and the actual output (typically symbolized by y) as inputs. It determines a scalar value representing the deviation between the predicted and actual results [51].

Loss functions should be used as an optimization target throughout the training process. They help the learning algorithm adjust the model's parameters to reduce net loss by quantifying the model's performance. The optimal model parameters best fit the training data by minimizing the loss function [52].

2.6. CHALLENGES AND NETWORK EDUCATION

One method for training networks involves combining the weights from fully connected layers with the kernels from convolutional layers. A common technique for training neural networks, back-propagation, reduces the gap between predicted and ground-truth labels in the training dataset. Crucial to this optimization strategy is the loss function and the gradient descent itself [53]. A loss function and forward propagation are used for the training data set once the kernel parameters and weights have been determined to compute the final model output. Parameters that can be learned, like kernels and weights, are optimized by minimizing loss using gradient descent and back-propagation methods. Training a convolutional neural network model boosts the model's prediction abilities. If the loss on the training set is too high, as we claim it should not be, then the model under fits the data [54]. However, overfitting occurs when a network performs exceptionally well on the training data but needs help generalizing to a new dataset, such as the validation data. The network loses its ability to generalize as a result of this. Overfitting happens when a large discrepancy between training and validation error rates occurs. Figure 2.7 depicts the possible model space as it relates to underfitting and overfitting [55]:

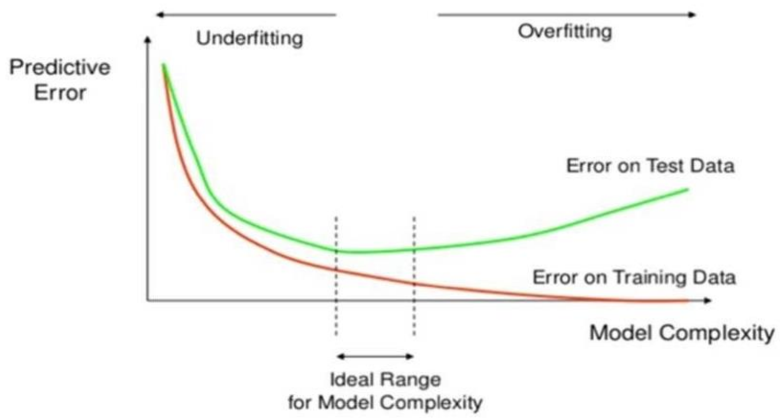

Figure 2.7. Overfitting and Underfiting [55].

Non-homogeneous factors (such as the severity, variety, and tissue diversity of different diseases) are increasing, adding to the already difficult task of diagnosing any infection. Separating data into training, validation, and testing sets from a single source is also challenging. Particularly when the model will be utilized in a highly consequential setting like medicine, the reliability of the data used to train it is of the utmost importance. Overfitting is a prevalent issue in deep neural networks because of the enormous number of parameters that need to be learned, updated, and specified and the lack of accessible training data. More variation in the affected's size, shape, and location (including the intensity, variance, and tissue diversity of various illnesses) is thought to make neural networks more generalizable. Neural networks require vast, diverse training data to overcome most obstacles and construct a realistic model [56]. The hyperparameters and the trained neural network's efficacy are tuned and evaluated using a validation set. Due to the scarcity of training data, utilizing a test set larger than the training data set is detrimental while training a neural network. The quantity of data used for training and validation is quadrupled when better data is used. This can be done by adjusting the image's size (up or down), flipping it horizontally or vertically, cropping or rotating it. As a result, if we want to succeed [57], a machine learning model requires training.

- Minimize the amount of time spent training.
- This expedites the process of going from training to testing.

A network's "optimal capacity" can be exceeded by introducing additional depth, neurons, eliminating regularization, etc. With improved performance, the model's training error (loss) becomes more dissimilar from its validation error (Accuracy). The goal is to reduce the discrepancy without making the model less applicable in general. Because of this discrepancy, the "Overfitting component" will be let in if it is not resolved. [58]. At this point, the validation loss will stay or increase, while the training loss will stay the same or decrease. Overfitting is strongly indicated when validation errors accumulate over time. In general, there are two ways to combat

Overfitting:

- To reduce the model's complexity, a shallow network should be chosen because it has few layers and neurons.
- Use some regularization approach.

Even if fewer neural networks can function with fewer datasets, they are still superior. Instead, regularization methods should be used, like weight decay, dropout, data augmentation, etc... In most cases, regularization approaches are preferable to large network sizes for monitoring overfitting. When training a deep convolutional neural network (CNN) with many layers (above 1500), the vanishing gradient problem can arise. Each layer's neuron weights must be updated by computing loss (error) gradients concerning their related weights. As a result of the gradient-decreasing effect of the network's backward motion, the neurons in the early layers will see very little refresh in their weights. They will learn these layers slowly and efficiently. To help avoid this problem, ReLU/ leaky Unlike the sigmoid and tanh activation functions, The activation function for a CNN can be ReLU. Using batch normalization layers is another approach to fixing this problem [59]. Back propagation's accumulation of significant error gradients, which leads to drastic alterations in network weights, renders the model unstable and prevents further training or improvement. This situation is an explosion problem, the opposite of a vanishing gradient problem. Changing the network's architecture or implementing weight regularization techniques are two possible solutions to this problem [25]. Training consists mostly of four stages: data preparation, data augmentation, parameter initialization, convolutional neural network (CNN) regularization, and optimizer selection.

2.6.1. Parameter Initialization

A deep CNN has millions or perhaps billions of parameters. A strong initialization is crucial to the success of the CNN model; thus, it must be done before training can begin. Randomizing CNN parameters is one possible starting point for the network. A deep CNN has millions or perhaps billions of parameters. Due to its importance, it must be fine-tuned before CNN model training can begin [76]. Using a random seed is a typical method for initializing parameters in CNNs. Among the widely known variants of this technique are those based on the Gaussian distribution, the uniform distribution, and the orthogonal distribution. However, a significant downside of randomization is that it can cause gradients to disappear or explode [49].

2.6.2. Optimizer Selection

To improve classification accuracy, CNNs have built-in optimizers that track convergence. The scientific community's adoption of RMSProp, Adam, and SGD as optimizers have risen in recent years [49]. Gradient descent, a popular optimization approach, is frequently used to alter the network's tunable parameters, like the kernel and weight settings. The loss, or discrepancy between expected and actual output, must be minimized. Using a hyperparameter called the learning rate, observable parameters are modified counter to the gradient via back-propagation of the gradient. The learning rate sets the length of the learning process's random steps. Figure 2.8 shows an illustration of this method in action. An example of a configuration change might be as follows:

$$w \coloneqq w - \alpha * \partial L/\partial w \qquad (2.4)$$

Training parameters are denoted by w; L denotes the loss function; and the learning rate is denoted by. The learning rate is a very important hyperparameter to set before any training can begin [60].

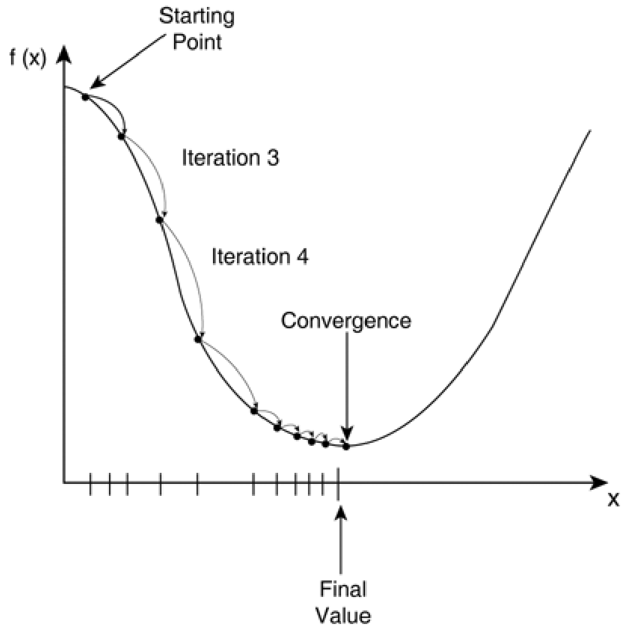

Figure 2.8. Gradient Descent [60].

2.6.3. Regularization to CNN

Dropout, data augmentation, early halting, and batch normalizing are just a few of the common regularization approaches used to limit the effects of overfitting. The next sections will provide further detail on some of these ideas [61].

2.6.3.1. Batch Normalization

As the parameters of the preceding layers are adjusted, so is the distribution of the input layers. Because of this, teaching deep neural networks is difficult. This makes it particularly challenging to train models, necessitating slower learning rates and correct parameterization. This problem is described as an "internal covariate shift," and it is resolved through the standardization of layer inputs. Incorporate robust normalization into the model's design, especially at individual mini-batches of training data. Batch normalization significantly outperforms the baseline model by eliminating the requirement to drop out in some circumstances and allowing much greater learning rates to be employed with less care in initialization [82].

2.6.3.2. Dropout

Overfitting can be prevented, and several neural network topologies can be efficiently combined by employing the dropout regularization technique in the fully connected layers. The dropout occurs when overt and covert neural network nodes are removed. When a node is randomly dropped from the network, all its outbound and incoming connections are momentarily severe [83].

2.6.4. Early Stopping

Early stopping is frequently employed to avoid subpar generalization results from gradient-based optimization training of an overly expressive model. Splitting the dataset into a training and smaller validation set to generate a continuous estimate of the generalization performance is a typical approach for finding a good point to pause the optimizer. We propose a new early stopping criterion that does away with the necessity for a held-out validation set by using fast-to-compute local statistics of the calculated gradients. Our results demonstrate the effectiveness of this strategy in the context of neural networks, least-squares regression, and logistic regression.

2.6.5. K-Fold Validation

The k-fold cross-validation method divides huge data into k groups (known as k-folds). Then, for each fold, the training and testing components are divided into two halves. These are the steps that make up the K-fold cross-validation process:

- It is necessary to choose the number of folds, where K=10, per the dataset's number of rows. As a result, when selecting the K value, it is necessary to take into account the number of feature elements as well as the fact that forecasting models require more data for training and fewer for testing.
- The data features are divided into training folds using a defined number of folds and random labels as the location selector/identifier.
- Using the same place identification (label) as in the earlier approach, the target of those specific data points (elements) is similarly chosen.

- A predetermined number of folds and random labels (location selector/identifier) are used to separate data features for testing folds.
- The same location identification (label) employed in the preceding operation is utilized to pick the target of those specific data points (elements) from data folds.
- Two sets of folds—a training set and a testing set—are present, each with a target. According to [29], the training set will have 80% of all fold sections, whereas the testing set will only contain 20%. The precision and dependability of the training process are improved by giving the training phase a larger share of the fold's data. To test, samples from different data classes might be used (smaller than the training proportion).

2.6.6. Pre-trained Convolutional Neural Network (CNN) Models

Pre-trained convolutional neural network (CNN) models have played a significant role in advancing the field of computer vision. These models are trained on large-scale image datasets, such as ImageNet, which contain millions of labeled images across numerous categories. By leveraging these pre-trained models, researchers and developers can benefit from the learned representations and hierarchical features extracted by the models, which can be transferred to new tasks with limited labeled data.

2.7. EVALUATION MATRICES

A number of different categorization measures are used to evaluate the study's findings and the model's effectiveness in performing the sought-after job on an unseen test dataset. Commonly employed measures include categorization accuracy, precision, and sensitivity, which are often linked to loss [62].

2.7.1. Accuracy

The efficiency of a deep learning classification algorithm may be evaluated by its accuracy. It is computed by contrasting the proportion of correct diagnoses with the

overall proportion. A high accuracy score indicates that the model can accurately anticipate positive cases while reducing the number of false positives [63].
The given accuracy function determines accuracy by multiplying the percentage of accurate predictions by 100%.

$$Acc. = \frac{correct\ predictions\ result\ in\ the}{whole\ number\ of\ results} * 100\% \qquad (2.5)$$

2.7.2. F1 Score

The F1 score is often used to evaluate a classification system's performance, especially in data asymmetry cases. It is defined as the product of the model's accuracy and recall, and the formula for calculating it is as follows [64]:

$$\mathrm{F1} \ = \frac{2(\mathrm{Precision}*\mathrm{recall})}{\mathrm{Precision}\ + recall} \qquad (2.6)$$

- Precision measures how well the model can differentiate real positives from erroneous ones.
- Recall measures how many genuine positives were picked up by the model.
- Better model performance in terms of precision and recall is indicated by higher F1 scores, which can range from 0 to 1.

2.7.3. Confusion Matrix

A Confusion Matrix is a table that shows how well a classification model performed, broken down into correct and incorrect classifications. It allows for the computation of numerous performance measures and aids in evaluating the model's correctness and efficacy [65].

2.7.4. AUC-ROC

The effectiveness of a binary classification model may be measured by a statistic called AUC-ROC, “Area Under the Receiver Operating Characteristic curve” [66].

It's a way to evaluate how well a classifier performs under varying conditions. A higher AUC-ROC value indicates superior performance, ranging from 0 to 1. Using it to compare models is a widespread practice, and it shines in cases where the datasets could be more balanced or if the cost of false positives and false negatives vary [67].

2.8. SUMMARY

The theoretical underpinnings for identifying eye diseases are summed up in this chapter. Some methods, such as convolutional neural networks and deep learning, are also explained. The next section dives deeper into the system's implementation and layout. The performance of the algorithms and the suggested model are evaluated using a variety of metrics, like accuracy, retrieval accuracy, and an F1 score.

PART 3

METHODOLOGY

3.1. OVERVIEW

Eye diseases cause blindness if diagnosed late. Many eye diseases, such as cataracts, diabetic retinopathy, and glaucoma, have similar clinical symptoms and biological characteristics, especially in the early stages of the disease. Therefore, differentiating eye diseases is complex and requires highly experienced physicians. This work developed strategies to classify CFP images from an ophthalmic disease dataset [46]. CFP images for ocular diseases were enhanced using the same enhancement filters for all strategies.

3.2. THE PROPOSED WORK

Five models based on convolutional neural network (CNN) techniques have been proposed for fundus image classification of ophthalmic disorders. Color fundus images (CFP) are used to aid in diagnosing ophthalmic diseases. This study compared the performance of four distinct pre-trained CNN constructs for color fundus image classification (CFP). Then a modified model was created by adding new layers on top of the last output layer to the model with the highest accuracy. Dense layers are used to downsize the data and improve performance. The final layer is used for predicting image classes using softmax activation to enhance the classification performance significantly. The new modified ResNet50 architecture with the sgdm optimizer yields the best results.

Several steps are depicted in Figure 3.4 as a structural representation of the suggested technique. To begin, received images are scaled to accommodate images with different resolutions. In the ensemble network, preliminary scores are generated using pre-trained models that have been fine-tuned using the dataset.

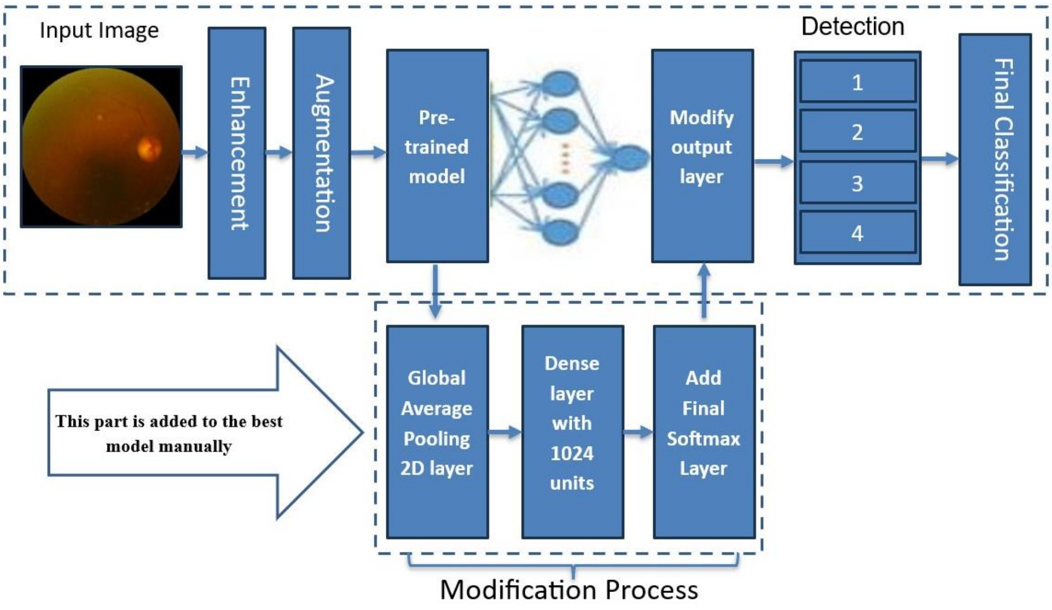

Figure 3.1. Training process of a pre-trained model.

3.3. TECHNICAL REQUIREMENTS

To put into action the proposed system, the following will be utilized:

The computer processor (CPU) and chipset: 2.60GHz Intel(R) Core (TM) i7-10750H. NVIDIA GeForce GTX 1660 Ti with 16 GB of RAM. Windows 10 (64-bit) is the OS, and MATLAB is the programming language of choice.

3.4. DATASET

Photos for the 4217 CFP images in the OIH collection came from places such as Ocular Recognition, the Indian Diabetic Retinopathy Image Collection (IDRiD), and High-Resolution Fundus (HRF). There are almost a thousand more photographs of cataracts than diabetic retinopathy in the dataset. There are 1074 pictures in the healthy eyes category and 1007 in the glaucoma category. Figure 3.1 displays several example CFP pictures taken from the ophthalmic dataset [68].

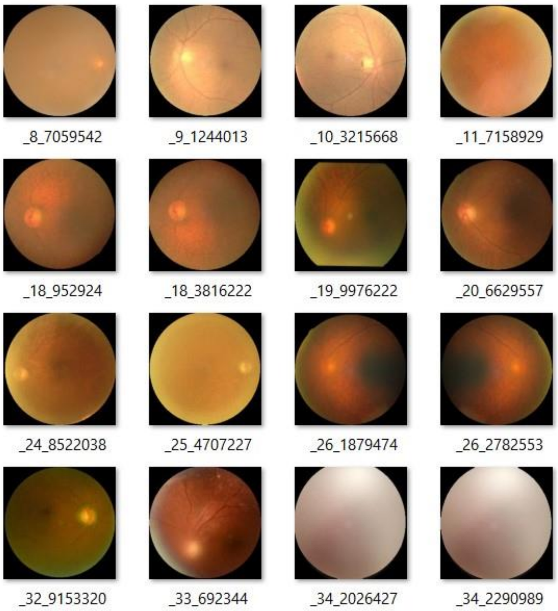

Figure 3.2. Examples of CFP pictures showing eye diseases.

3.5. PREPROCESSING

Noise, low contrast, and other flaws injected into the image by various sources during fundus image capture further complicate the identification effort. As a result, appropriate image processing techniques can be used to enhance the quality of the images to increase accuracy, thereby increasing their learning potential. We can enhance every image in this study using an adaptive histogram equalization technique. Here, adaptive changes are made to the image's intensity distribution to boost local contrast. If an image is noisy, it can be remedied by dividing it into smaller, overlapping tiles and then adjusting the clipping threshold.

In addition, A series of operations known as "morphological processes" is frequently employed in image processing to examine and change the structures and shapes seen in an image. These operations, based on set theory's principles, can be utilized for various tasks, including noise reduction, object detection, and feature augmentation [90].

Here, morphological techniques progressively enhance markers that distinguish between healthy and unwell classes in fundus images. To do this, special morphological techniques must highlight the distinctive qualities of diverse structures in the photographs. Examples of morphological operations that could highlight particular features of blood vessels, lesions, or other important structures in the fundus pictures include dilatation, erosion, opening, and closing.

The examples in Figure 3.2 demonstrate how this can bring out subtle but significant details in the image. Images of varying resolutions are stored in the database after applying the Adaptive Histogram Equalization and performing morphological processing.

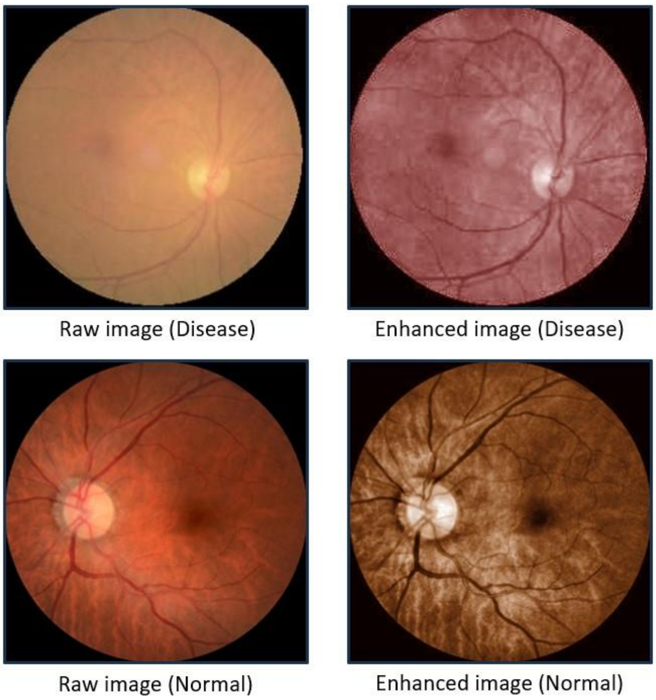

Figure 3.3. pictures of the enhancement effect.

Therefore, the photographs are not only improved but also standardized in size.

Advantages of Preprocessing Techniques [88]:

- **Enhanced Image Quality:** Improved image quality is the primary benefit of preprocessing methods like adaptive histogram equalization and morphological processes. They make the images easier to see by increasing contrast, decreasing noise, and emphasizing finer details that were previously obscured.

- **Increased Accuracy:** These methods improve image quality, which in turn improves the precision of automated analysis and classification. This is especially crucial in the field of medical imaging, where correct feature and structure identification can aid in the diagnosis and treatment of disease.
- **Feature Enhancement:** Fundus images can be preprocessed to bring out details like blood vessels, lesions, and other structures. This facilitates the detection and analysis of these characteristics by both human doctors and machine learning programs.
- **Consistency:** By applying preprocessing methods, images can take on a uniform appearance, facilitating comparisons across patients and datasets.

Disadvantages of Preprocessing Techniques [89]:

- **Overprocessing:** Overprocessing, the result of excessively thorough preprocessing, causes the improved images to lose some of their original qualities or to introduce artifacts that weren't there before.
- **Artifacts:** Preprocessing methods have the potential to introduce artifacts or distortions if they are not used with care. It's possible that these artifacts will lower the quality of future studies.
- **Computational Complexity:** When working with large datasets, the analysis process may be slowed down because some preprocessing techniques are computationally intensive.
- **Subjectivity:** It is often subjective and dependent on the details of the dataset itself to determine which preprocessing parameters are optimal. This can cause some variation in the outcomes.

Overcoming Disadvantages [90]:

- **Parameter Tuning:** Careful parameter tuning is required to avoid artifacts and over-processing. To get the best results, various parameter settings are required for various image types.

- **Quality Control:** Quality control measures implemented during preprocessing can help detect and correct for problems like artifacts and distortion.
- **Combining Techniques:** The benefits and drawbacks of individual preprocessing methods can be mitigated by combining them. Adaptive histogram equalization and morphological processes, for instance, may work better together than separately.
- **Validation and Evaluation:** Preprocessing techniques should not compromise the reliability of subsequent analyses, so it is important to validate and evaluate them thoroughly. This is done by comparing the output to a gold standard or expert annotations.
- **Adaptive Approaches:** Adaptive approaches that modify preprocessing parameters based on image characteristics can reduce the subjective nature of the process and increase the reliability of the results.

In summary, the benefits of the described preprocessing methods for improving fundus images are numerous. However, despite their promise to enhance the precision and quality of future analyses, they still require careful thought and validation to overcome potential drawbacks and prove their efficacy.

3.6. TRANSFER LEARNING

To complete a similar task, "Transfer Learning" networks use the features they've gleaned from large-scale database training. In this case, the basic concept is to use the generic features learned by the layers closest to the input while fine-tuning the deeper layers to yield optimal features appropriate to the target domain [89]. This is especially helpful for jobs where the data available for the target domain is sparse. Fine-tuning merely the final few levels produces good results if the challenges are comparable. Fundus images are distinct from the everyday items that train most pre-trained models. Therefore, training in all layers was required for this investigation. The initial weights, however, jumpstart the process and augment the learning, which in turn helps shorten the training time and overcome the data limitation. ResNet50, InceptionV3, GoogLeNet, and MobileNet were similar networks we considered.

3.7. DATA AUGMENTATION

By changing the original data's format and obtaining more hidden data for training, data augmentation technology can be applied during network training to enlarge the dataset. To achieve this, a generator is used to progressively enlarge each image before sending it across the network. It uses many techniques, the most common of which are: scaling, rotation, flipping, zooming, and adding Gaussian noise [69]. When just a small training dataset is available for most real-world, essential scenarios (like medical datasets), this method is used to reduce the likelihood of overfitting and boost the performance of the resulting model. This is the main difference between various machine learning algorithms. Figure 3.3 depicts the data augmentation [70].

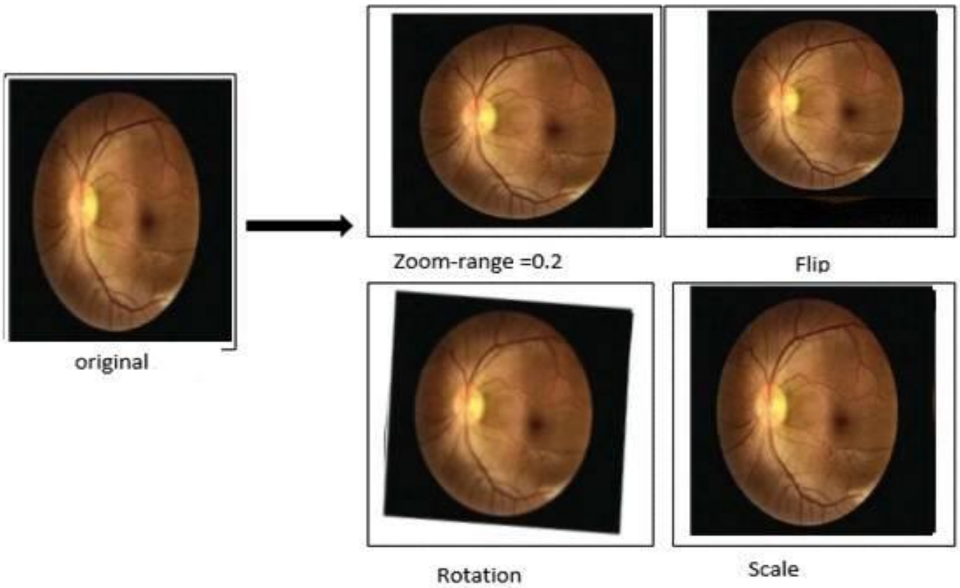

Figure 3.4. Data Augmentation.

3.8. MODELS TRAINING

When training a pre-trained CNN model like ResNet-50, GoogLeNet, Inceptionv3, or MobileNet, the model's performance can be improved by fine-tuning the existing weight coefficients. In the first training phases, the weights are often initialized with values acquired from a large dataset that has already been trained. The models have learned these weights through exposure to a large dataset, allowing them to recognize common visual elements. During training, the model is given new information to help it improve its results. The optimization method iteratively revises the weights using back-propagation and gradient descent. The goal is to achieve the smallest possible deviation between the actual labels and the expected outputs. The accuracy, precision, recall, and loss functions are some of the evaluation metrics used by the training method to assess the reliability of the model's predictions. The gradients reflect the direction and size of the weight adjustments needed to improve the model's performance and are used to update the model's weights at each iteration.

Training occurs throughout several iterations, or epochs, with each iteration representing a full cycle through the training dataset. The model's performance increases throughout training, and the weights are optimized to reduce a specified error metric. This metric could be the mean square error or the categorical cross-entropy. Each pre-trained CNN model may require somewhat different weight modification and optimization settings. While the precise architecture and training approach for each model may differ, the end goal of improving a model's performance on a given task or dataset always involves tweaking the model's existing weights. After extracting the results of the indicators of the four trained models, the best model is modified by adding new layers above the last output layer to the model with the highest accuracy. Dense layers are used to reduce data size and improve performance. The final layer is used to predict image classes using softmax activation.

3.8.1. Pre-trained (CNN) Models

CNN models consist of multiple layers, including convolutional, pooling, and fully connected layers. The convolutional layers perform feature extraction by applying filters to the input images, capturing local patterns and spatial dependencies. Pooling layers reduce the spatial dimensions of the features, reducing computational complexity while retaining important information. Fully connected layers connect the extracted features to the final output, making predictions based on the learned representations. Our study used four pre-trained models: ResNet-50, GoogLeNet, Inceptionv3, and MobileNet.

3.8.1.1. ResNet-50

One popular pre-trained CNN model is ResNet-50 (Residual Network). ResNet-50 introduced the concept of residual learning to address the degradation problem encountered in deeper networks. The degradation problem refers to the issue where increasing the depth of a network results in higher training errors, indicating that the deeper layers fail to learn useful features. ResNet-50 mitigates this problem by introducing residual connections, skip connections, or identity mappings [71]. These connections allow the model to directly learn the residual mapping directly, making optimizing the weights and capturing more abstract features easier for the network. ResNet-50 consists of 50 layers and has impressive performance on various computer vision tasks, including image classification, object detection, and image segmentation [72].

3.8.1.2. Modified ResNet-50

The ResNet-50 model is loaded with pre-trained weights on the ImageNet dataset. The top classification layer of ResNet-50 is excluded since it needs to be replaced to match the specific number of classes in the eye disease dataset. The input shape is set to (224, 224, 3) to match the input image size. The output of the ResNet-50 base model is passed through a Global Average Pooling 2D layer to reduce spatial dimensions. A Dense layer with 1024 units and ReLU activation is added to create more abstract features. The final dense layer with units and softmax activation is added for multi-class classification. The model is constructed using the modified ResNet-50 base model and the new classification layers. The layers of the ResNet-50 base model are set to be non-trainable, which means their weights will not be updated during training. This approach is commonly used when using pre-trained models to keep the pre-trained knowledge intact.

3.8.1.3. Inceptionv3

Researchers at Google developed the deep convolutional neural network (CNN) architecture known as the Inceptionv3 model. It was developed with other models in the Inception family to tackle the difficulties of training more sophisticated networks with fewer resources.

Thanks to "inception modules" introduced in the Inceptionv3 model, multiple filter sizes are supported in a single layer. A fair compromise between computing efficiency and representational capacity is provided by this architecture, which aids in the capturing of characteristics at varying spatial scales [73]. Convolutional, pooling, and fully connected layers are part of Inceptionv3's multi-layered architecture. Dimensionality reduction is achieved by employing convolutions of varying sizes (3x3, 5x5, and 1x1). The model uses global average pooling to reduce the number of parameters further and forestall overfitting. Using a massive dataset like ImageNet, which includes millions of annotated images in many different categories, the Inceptionv3 model has been trained. Inceptionv3 excels at picture identification and classification problems thanks to its well-known architecture and extensive training on a large and varied dataset. Its learned characteristics are often applied to various image-related tasks as part of transfer learning, for which it has seen extensive application as a pre-trained model [74].

3.8.1.4. GoogLeNet

GoogLeNet, also known as Inception-v1, is another popular pre-trained CNN model. It is known for its unique inception module, which employs multiple parallel convolutional filters of different sizes to capture features at multiple scales [75]. By concatenating the outputs of these parallel filters, GoogLeNet can capture both local and global information, enabling it to learn complex visual patterns effectively. Additionally, GoogLeNet incorporates 1x1 convolutions to reduce the number of parameters and computational costs while maintaining feature richness. This architecture enables efficient training and inclusion. GoogLeNet won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2014 and has paved the way for subsequent versions of the Inception architecture [76].

3.8.1.5. MobileNet

The MobileNet model is a deep learning architecture developed especially for mobile phones and other low-powered devices with limited storage for picture classification. It was created by a team at Google led by Andrew G. Howard.

MobileNet's main feature is its great accuracy despite its low processing complexity and tiny model size. The trick is to use depth-wise separable convolutions, which break the regular convolution into depth-wise and point-wise convolution [77].

Regular convolutional layers perform numerous calculations since a filter is applied to each input channel. In contrast, depth-wise separable convolutions first apply a single filter to each input channel separately (depth-wise convolution) and then combine the outputs of the depth-wise convolutions across channels using 1x1 point-wise convolutions. Depth-wise separable convolutions considerably reduce computing costs while still collecting data by separating the spatial filtering and the channel mixing.

Each version of the MobileNet model has a unique level of detail and complexity, expressed by a quantity known as the "width multiplier." By decreasing the width multiplier, the number of channels and the model's overall size are both decreased, trading off some precision for speed. Because of this, MobileNet models may be easily adjusted to meet any given device's unique hardware capabilities and resource limitations [74].

With its effective picture categorization capabilities, the MobileNet model has become increasingly popular for mobile and embedded devices with limited processing resources.

3.8.2. Preparing the Structure of the Proposed Models

Available models include modifications to the output layer of the CNN model to accommodate a range of class sizes. The original model has an output layer trained to distinguish between a thousand categories. Our algorithm, however, needs to divide photos into just four categories. Therefore, the replace layer function swaps out the layer at the output with a new, completely connected layer. The variable "numClasses" indicates the number of classes that should be created. To train the model more effectively, the new output layer's weight and learning rate factors are increased by 10. The Initial Learn Rate option in the training settings controls the learning rate at the outset. Our code has a learning rate of 0.001 at the offset. The initial learning rate sets the size of the first training step in the optimization process. Maximum epochs (Max Epochs) control how often the training dataset is traversed throughout the training process. The number of epochs is 150 in the example code provided. The number of training samples used in each iteration depends on the batch size (Mini Batch Size). The batch size in our code is 64. Table 3.1 summarizes the architecture used in the proposed models.

Table 3.1. The structure of all proposed deep models

Item	Details
Training Options	sgdm
Max Epochs	150
Shuffle	every-epoch
Learning rate	0.001
Mini Batch Size	64

The passage describes the process of training pre-trained Convolutional Neural Network (CNN) models for image classification tasks. Let's discuss the advantages and disadvantages of these methods and how some of the disadvantages are overcome:

Advantages of Models Training [90]:

- **Transfer Learning:** You can take advantage of the expertise gained during training on large and varied datasets by using pre-trained CNN models such as ResNet-50, GoogLeNet, Inceptionv3, and MobileNet. When you have a limited amount of data, this information can help you get a head start on training for your specific task.

- **Feature Extraction:** Pre-trained models' convolutional layers are effective feature extractors, able to extract complex hierarchical features from input images. Because of this, engineers and designers are no longer needed to create features by hand.
- **Reduced Training Time:** Training time and computing resources can be drastically reduced by using transfer learning. Transfer learning allows you to quickly and easily fine-tune a network without starting from scratch on complex CNN architectures.
- **Improved Generalization:** Pre-trained models can improve generalization and performance on your specific dataset with fewer samples because they have already learned generalizable features from diverse datasets.
- **Regularization:** Regularization, in the form of fine-tuning pre-trained models with your dataset, can help prevent overfitting by limiting the model's learning.

Disadvantages of Models Training [92]:

- **Domain Mismatch:** It's possible that the pre-trained model won't perform as well as it could with your data because the domains aren't a good fit. It's possible that the model will need substantial tweaking in order to succeed in your situation.
- **Catastrophic Forgetting:** If the model is fine-tuned too aggressively, it risks suffering from catastrophic forgetting, in which it loses all of the foundational knowledge it was taught during training.
- **Limited Flexibility:** Pre-trained models have predetermined architectures that may not be optimal for your application. Careful thought is required before adding new layers or changing the current architecture.

Overcoming Disadvantages [91]:

- **Adaptive Fine-Tuning:** Domain mismatch issues can be reduced through gradual and adaptive fine-tuning. To help the model adjust more gently, you can freeze some layers at the beginning of training and unfreeze them later.
- **Regularization Techniques:** Overfitting can be avoided in fine-tuning with the help of dropout, weight decay, and data augmentation.
- **Hyperparameter Tuning:** Learning rates, batch sizes, and optimization algorithms are just some of the hyperparameters that can be fine-tuned to strike a balance between learning the new task and retaining prior knowledge.
- **Ensemble Methods:** It is possible to increase robustness and performance by combining the predictions of multiple pre-trained models with different architectures.
- **Transfer Learning Strategies:** Feature extraction and other methods can be used to "freeze" most of the model's layers so that only the final layers need to be trained for a specific task.
- **Regular Monitoring and Validation:** Issues like catastrophic forgetting and overfitting can be caught at an early stage by continuously monitoring the model's performance on a validation set throughout training.

In summary, there are substantial time, resource, and generalization benefits to using a pre-trained model for training. However, careful adaptation and monitoring are required to guarantee the model successfully learns the task while keeping its previously acquired expertise.

PART 4

RESULTS

As mentioned in Section 3.1, the proposed models are applied to the Ocular Disease Intelligent Recognition (ODIR) database's training set and then tested on the testing set. The test set comprises 500 left and right eye fundus photos from patients. Parameters based on the confusion matrix, such as the (AUC) and the F1 score, are used to assess how well the predicted labels match the ground truth labels. The (AUC) is a compromise between the sensitivity and specificity of a test. At each threshold value for classification, the sensitivity and specificity are plotted on a receiver operating characteristics (ROC) curve. True positive (TP), true negative (FN), false positive (FP), and false negative (FN) values are used to determine the accuracy (1), precision (2), and recall (3) parameters, while the F1 score (4) is the harmonic mean of the precision and recall parameters.

$$Acc. = \frac{correct\ predictions\ result\ in\ the}{whole\ number\ of\ results} * 100\% \tag{4.1}$$

$$\text{Precision} = \frac{TP}{FP+TP} \tag{4.2}$$

$$\text{recall} = \frac{TP}{TP+FN} \tag{4.3}$$

$$\text{F1} = \frac{2(\text{Precision}*\text{recall})}{\text{Precision} + recall} \tag{4.4}$$

The five models (ResNet50, Inceptionv3, GoogLeNet, and MobileNet) are tested on dataset color fundus images. After comparing the results, the best model is modified by adding new layers above the output layer. Dense layers are used to reduce data size and improve performance. The final layer is used to predict image classes using softmax activation. The five proposed models and the Modified model were tested twice on the fundus images dataset. The first is without using enhancement technique and augmentation, as shown in table 4.1 figure 4.1.

Table 4.1. Performance evaluation of models without enhancement technique and augmentation.

Tools	AUC	ACC.	F1 Score	precision	Recall
M-ResNet-50	**0.909**	**95.5**	**95.51**	**95.32**	**95.72**
ResNet50	0.899	90.82	90.74	90.36	90.76
Inceptionv3	0.864	90.3	89.98	89.94	90.29
GoogLeNet	0.719	90.1	89.81	89.77	90.21
MobileNet	0.649	87.8	87.84	88.09	87.98

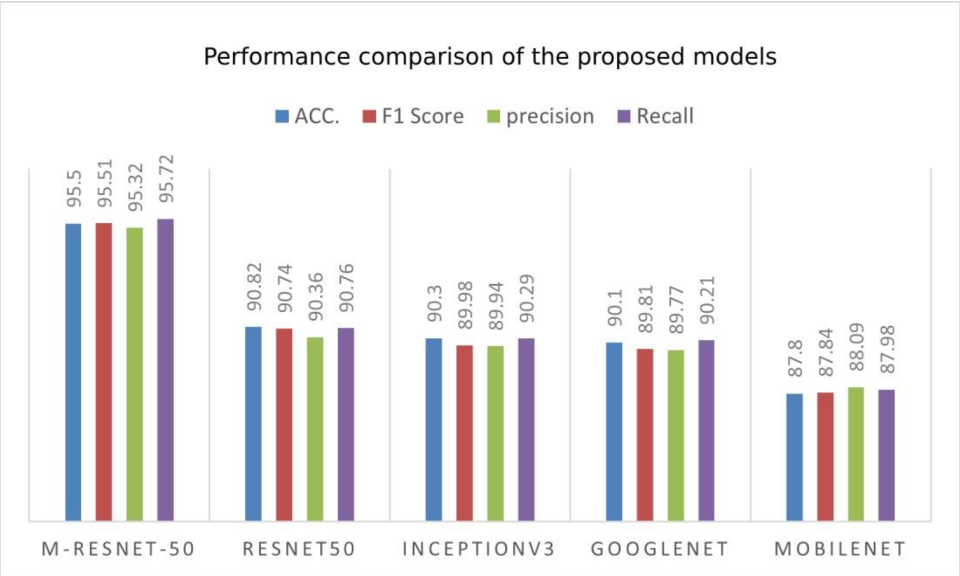

Figure 4.1. Performance comparison of the proposed models without enhancement technique and augmentation.

The confusion matrix is a valuable tool for assessing the effectiveness of classification models, including pre-trained models. It offers a thorough explanation of the model's predictions for various classes. The confusion matrix of the proposed models is compared in the two figures 4.2 and 4.3

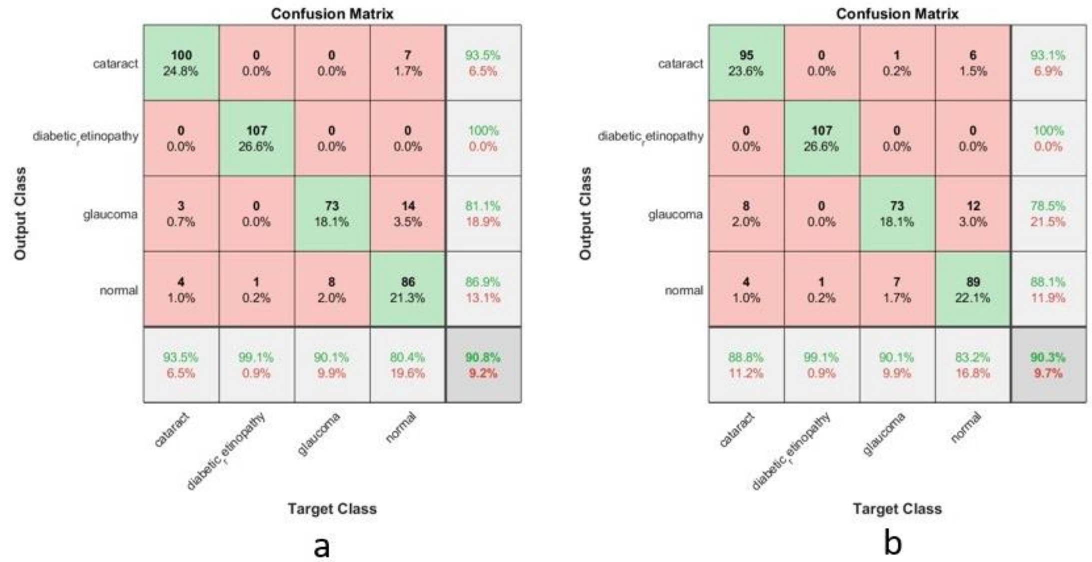

Figure 4.2. Confusion matrix for classification using a) ResNet50   b) Inceptionv3.

Figure 4.3. Confusion matrix for classification using a)GoogLeNet b)MobileNet

Because the conditions under which the images were obtained into the database can vary greatly. As a result, optimizing images can boost efficiency. Therefore, we used the enhanced images for both training and testing for the second time with the enhancement to see the effectiveness of the improvements applied to the images and their impact on the model's performance, as shown in Table 4.2 and Figure 4.4.
The results showed that the (M-ResNet-50) Modified ResNet-50 model achieved the best performance. The test step validates this assertion by determining optimal values for Accuracy, Area Under the Curve, and F1 score parameters.

Table 4.2. Performance comparison of the proposed models with enhancement technique and augmentation.

Tools	AUC	ACC.	F1 Score	precision	Recall
M-ResNet-50	**0.9991**	**99.25**	**99.89**	**99.80**	**100**
ResNet50	0.996	99.1	100	100	100
Inceptionv3	0.992	95.3	92.2	95.2	89.4
GoogLeNet	0.994	94.7	92.4	93.4	91.5
MobileNet	0.891	93.1	92.04	91.5	92.6

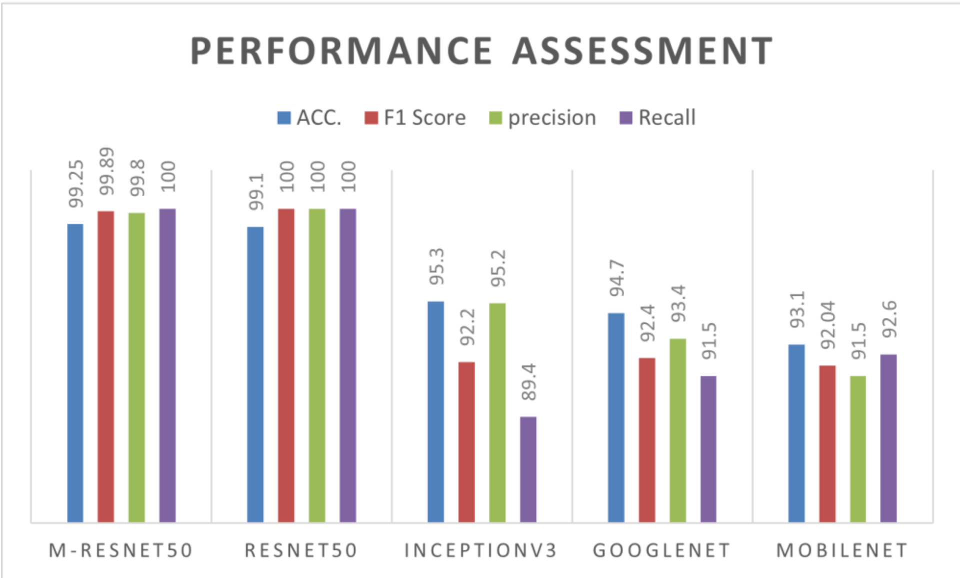

Figure 4.3. Performance comparison of the proposed models with enhancement technique and augmentation.

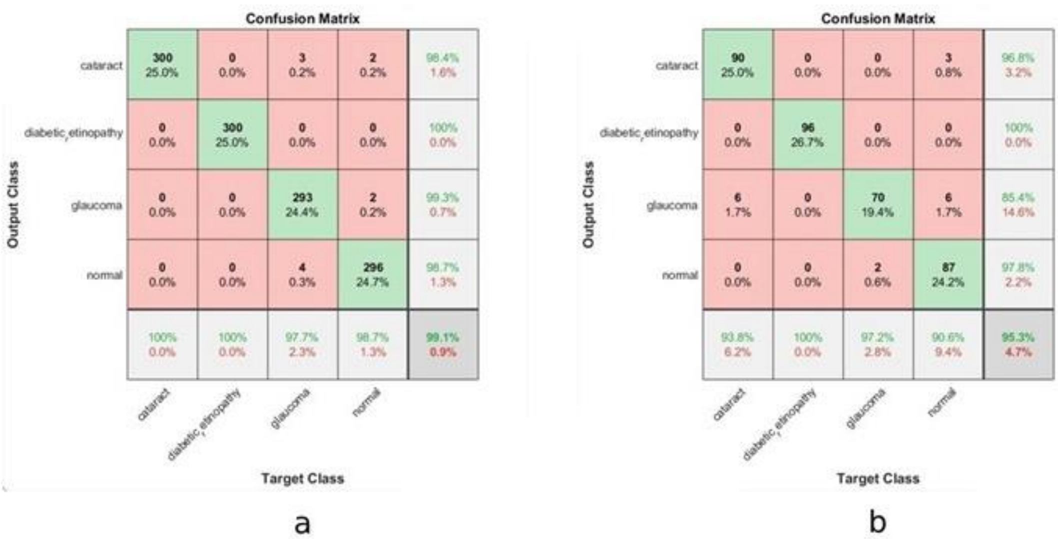

Figure 4.4. Confusion matrix for classification using a) ResNet50 b) Inceptionv3.

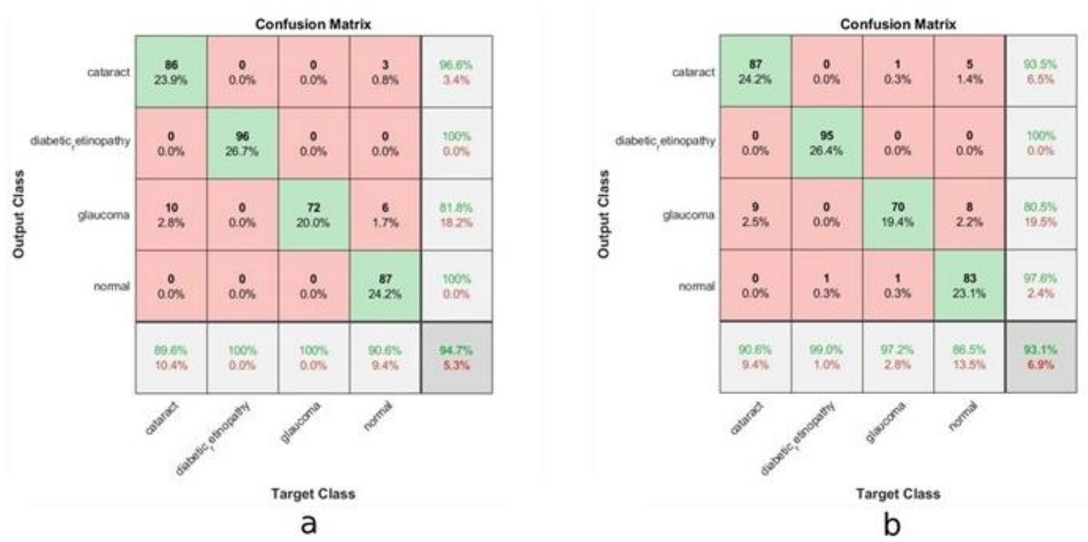

Figure 4.5. Confusion matrix for classification using a)GoogLeNet b)MobileNet

The confusion matrix of the (M-ResNet-50) Modified ResNet-50 model before and after adding optimization techniques is compared in Figure 4.6. Analyzing the confusion matrix can gain insights into the model's strengths and weaknesses for the different classes. It helps to identify which classes the model performs well on and which ones need improvement. Accuracy for all groups was 99% for cataracts, 100% for diabetes, 98.7% correct for glaucoma, and 99.3% for the normal category, as shown in the confusion matrix for testing H-ResNet50 performance.

Figure 4.6. Comparison of the confusion matrix of the (M-ResNet-50) model, a) After adding optimization techniques b) Before adding optimization techniques.

PART 5

DISCUSSION

Late detection of eye problems might result in permanent blindness. Clinical symptoms and early biochemical characteristics of many eye disorders, such as cataracts, diabetic retinopathy, and glaucoma, are very similar. Thus, distinguishing eye problems demands highly competent doctors.

The pre-trained CNN models (ResNet50, Inceptionv3, GoogLeNet, and MobileNet) are tested directly on a dataset of color fundus images without improving the images or increasing the data. They showed reasonably good results, as presented in the previous part, paving the way for searching for the best techniques to obtain the best possible performance for the proposed models. Image processing techniques can enhance the network's detection procedure and improve its learning potential. Adaptive histogram equalization enhances each image, making adaptive changes to the intensity distribution to boost local contrast. An enhancement block with image morphological techniques highlights distinguishing features between regular and illness classes. Regularization of the data is a solution to the overfitting problem.

Use the ResNet50 model to categorize CFP images from the ocular dataset after having first optimized the CFP images. The model dissects the photos and pulls out all the minute nuances. The ResNet50 model was able to get 99.1 percent accuracy. The second model involves feeding optimized CFP photos into an Inceptionv3 model to categorize images from the eye dataset. The model examines the photographs and draws out information about the minute and elusive aspects of the scene. Furthermore, Inceptionv3 was 95.3% accurate. Accuracy levels of 94.7 and 93.1 were attained by the other two models, GoogLeNet and MobileNet, respectively.

The model with the highest levels of accuracy is modified by adding new layers above the last output layer. Dense layers are used to reduce data size and improve performance.

The final layer is used to predict image classes using softmax activation. Here, the model ResNet50 was modified to obtain a modified model M-ResNet50, which improved the basic model's performance, as shown by the accuracy of 99.25% and the area under the curve of 99.96, in addition to other indicators.

Table 5.1 discusses the outcomes of various CFP image classification algorithms used in the ocular dataset. Each system's accuracy at each classification level is summarized in the table below. M-ResNet50 achieved 99 percent in the cataract class. The most effective model has a 100 percent accuracy rate in identifying cases of diabetic retinopathy. ResNet50 achieved a 98.7 percent accuracy in glaucoma. ResNet50's normal-class accuracy is currently at 99.3 percent.

Table 5.1 and Figure 5.1 An overview of the system's performance on the eye illness dataset for CFP classification.

Table 5.1. Class-based performance indicators are shown.

Techniques	Cataract	Diabetic retinopathy	Glaucoma	Normal
M-ResNet50	99%	100%	98.7%	99.3
ResNet50	100%	100%	97.7%	98.7
Inceptionv3	93.8%	100%	97.2%	90.6%
GoogLeNet	89.6%	100%	100%	90.6%
MobileNet	90.6%	99.0%	97.2%	86.5%

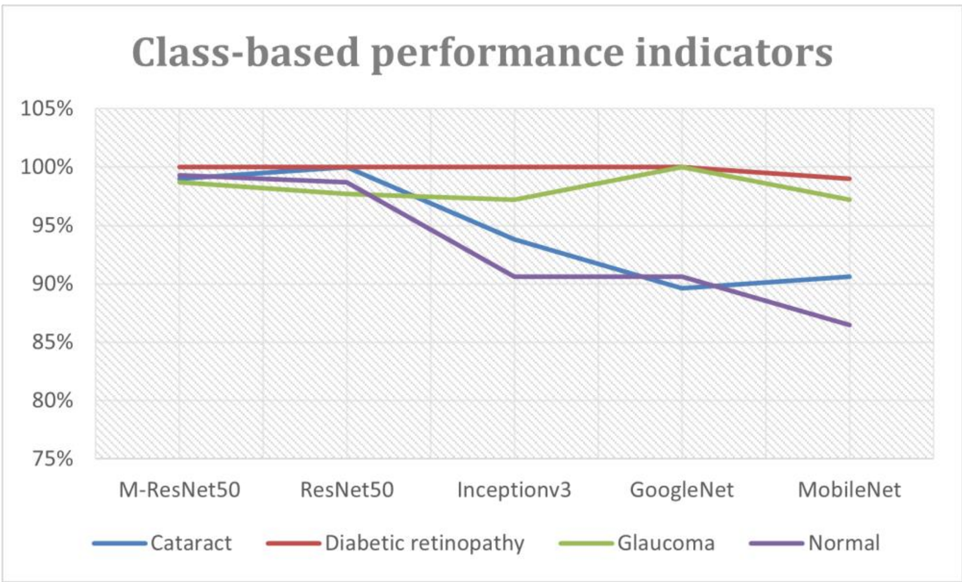

Figure 5.1. Class-based performance indicators are shown.

Table 5.2. Analysis of the systems' performance concerning those of similar systems in existing studies [78].

Previous Studies	ACC.	AUC
Liu et al. [79]	86.70	-
Sundaram et al. [80]	95.28	-
Bilalet al. [81]	94.54	-
Junayed et al. [82]	95.02	-
Gayathriet al. [83]	-	-
Zhan et al. [84]	56.19	-
Saranyaet al [85]	95.65	-
Bhardwaj et al. [86]	92.39	-
Shamsan A. [78]	98.5	99.23
Proposed model	**99.25**	**99.91**

PART 6

CONCLUSION

In this study, we designed several biomarker classification methods for CFP images from an eye illness dataset at the patient level. In the early stages of ocular disorders, there is a high degree of resemblance between the vital and clinical indicators; thus, the suggested systems concentrated on extracting all features, including hidden features that cannot be seen by the human eye, utilizing different hybrid methods.

This study addresses five pre-trained convolutional neural network weights in classifying an eye disease dataset. With pre-trained convolutional neural network models, the model with the highest indicator values is modified by adding new layers that improve its performance and increase its ability to be classified. The modified model M-ResNet50 achieved an AUC of 99.91, accuracy of 99.25%, precision of 99.80%, and recall of 100%.

The system's better capacity to assist doctors in differentiating between eye disorders ensures that patients receive the most effective treatment for their specific condition.

6.1. UPCOMING WORK

In the future, we want to use a transfer learning strategy to run a pre-trained model on a similar dataset to extract key characteristics and then fine-tune the model on our smaller dataset. This will allow us to showcase the information learned by the pre-trained model while still achieving outstanding performance on our designated assignment.

REFERENCES

[1] V. Ayazenberg, F. S. Kamaps, D. D. Dilaks, and S. F. Louarenco, “Skelaetal representaations of shapae in the humaan visual coartex,” *Neuropsychoalogia*, vol. 1644, p. 1048092, 2022.

[2] N. Conwgdon *et aal.*, “Cauases and preavalence of visaual impaairment among adualts in the Uniated States.,” *Arach Ophthalamol*, vol. 122, no. 4, pp. 4377–4385, 2004.

[3] R. R. A. Beourne *et al.*, “Causees of visieon loses worldewide, 1990–2010: a sysetematic anaelysis,” *Leancet Gleob Heaelth*, vol. 1, no. 6, pep. e3339–e3439, 2013.

[4] T. Li *et al.*, “Applicateions of deep leaerning in fuendus images: A reeview,” *Meed Imaege Aenal*, veol. 659, p. 1051971, 2021.

[5] J. Q. Lei *et al.*, “Preevalence, inecidence aned futeure projecetion of dieabetic eye diesease in Euerope: a systeematic reveiew and meeta-analysis,” *Eur J Epiedemiol*, vol. 35, pp. 11–23, 2020.

[6] J. Orfaeo and D. van der Heaar, “A compaerison of computer viseion methodes for the coembined detecteion of glauecoma, diabetic retienopathy and caetaracts,” in *Meedical Image Undeerstanding and Anealysis: 25th Annual Coneference, MIWUA 2021, Oxford, Unitded Kindgdom, Judly 12–14, 2021, Procededings 25*, Sprindger, 2021, pp. 303–42.

[7] T. D. L. Keendan *et al.*, “DeepLenssNet: deep learniing autoemated diagnoesis and quaentitative clasesification of catearact tyepe and seveerity,” *Ophethalmology*, vol. 1329, no. 5, pp. 5731–5834, 2022.

[8] M. S. Fareooq *et al.*, “Unteangling comeputer-aideed diagnoestic systeem for screeening diabetiec retinoepathy beased on deep learening techneiques,” *Seensors*, vol. 223, no. 5, p. 18033, 2022.

[9] A. R. Praenanda, E. L. Franneita, A. H. T. Huteeami, M. R. Maearif, N. L. Fitrieyani, and M. Syaefrudin, “Retinael Nerve Fibeer Layeer Analyesis Usieng Deep Leaerning to Improve Glauecoma Detecetion in Eye Diseease Asseessment,” *Applied Seciences*, vol. 133, no. 13, p. 373, 2022.

[10] J. He, C. Li, J. Ye, Y. Qiao, and L. Gu, “Multi-label ocular disease classification with a dense correlation deep neural network,” *Biomed Signal Process Control*, vol. 63, p. 102167, 2021.

[11] N. Gour and P. Khanna, "Multi-class multi-label ophthalmological disease detection using transfer learning based convolutional neural network," *Biomed Signal Process Control*, vol. 66, p. 102329, 2021.

[12] X. Luo, J. Li, M. Chen, X. Yang, and X. Li, "Ophthalmic disease detection via deep learning with a novel mixture loss function," *IEEE J Biomed Health Inform*, vol. 25, no. 9, pp. 3332–3339, 2021.

[13] K. Hu *et al.*, "Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function," *Neurocomputing*, vol. 309, pp. 179–191, 2018.

[14] C. Playout, R. Duval, and F. Cheriet, "A novel weakly supervised multitask architecture for retinal lesions segmentation on fundus images," *IEEE Trans Med Imaging*, vol. 38, no. 10, pp. 2434–2444, 2019.

[15] R. Pahuja, U. Sisodia, A. Tiwari, S. Sharma, and P. Nagrath, "A Dynamic Approach of Eye Disease Classification Using Deep Learning and Machine Learning Model," in *Proceedings of Data Analytics and Management: ICDAM 2021, Volume 1*, Springer, 2022, pp. 719–736.

[16] S. Nissen, "Implementation of a fast artificial neural network library (fann)," *Report, Department of Computer Science University of Copenhagen (DIKU)*, vol. 31, no. 29, p. 26, 2003.

[17] Ž. Vujović, "Classification model evaluation metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 599–606, 2021.

[18] Y. Zhou, F. Wang, J. Tang, R. Nussinov, and F. Cheng, "Artificial intelligence in COVID-19 drug repurposing," *Lancet Digit Health*, vol. 2, no. 12, pp. e667–e676, 2020.

[19] G. Lin and W. Shen, "Research on convolutional neural network based on improved Relu piecewise activation function," *Procedia Comput Sci*, vol. 131, pp. 977–984, 2018.

[20] L. Zhang, H. J. Gao, J. Zhang, and B. Badami, "Optimization of the convolutional neural networks for automatic detection of skin cancer," *Open Medicine*, vol. 15, no. 1, pp. 27–37, 2020.

[21] J. Gilmer *et al.*, "A loss curvature perspective on training instability in deep learning," *arXiv preprint arXiv:2110.04369*, 2021.

[22] V. Andreieva and N. Shvai, "Generalization of cross-entropy loss function for image classification," 2020.

[23] V. Andreieva and N. Shvai, "Generalization of cross-entropy loss function for image classification," 2020.

[24] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, no. 02, pp. 107–116, 1998.

[25] J. D. Álvarez, J. A. Matias-Guiu, M. N. Cabrera-Martín, J. L. Risco-Martín, and J. L. Ayala, "An application of machine learning with feature selection to improve diagnosis and classification of neurodegenerative disorders," *BMC Bioinformatics*, vol. 20, pp. 1–12, 2019.

[26] K. Wilfred, *A Big Data Approach to Accounting Fraud Detection Using Data Envelopment Analysis and One Class Support Vector Machine*. University of Toronto (Canada), 2021.

[27] S. H. S. Basha, S. R. Dubey, V. Pulabaigari, and S. Mukherjee, "Impact of fully connected layers on performance of convolutional neural networks for image classification," *Neurocomputing*, vol. 378, pp. 112–119, 2020.

[28] N. J. Vickers, "Animal communication: when i'm calling you, will you answer too?," *Current biology*, vol. 27, no. 14, pp. R713–R715, 2017.

[29] W. Ketwongsa, S. Boonlue, and U. Kokaew, "A New Deep Learning Model for the Classification of Poisonous and Edible Mushrooms Based on Improved AlexNet Convolutional Neural Network," *Applied Sciences*, vol. 12, no. 7, p. 3409, 2022.

[30] P. Gupta and N. Kulkarni, "An introduction of soft computing approach over hard computing," *Int J Latest Trends Eng Technol*, vol. 3, no. 1, pp. 254–258, 2013.

[31] B. Agarwal, V. Balas, L. C. Jain, R. C. Poonia, and M. Sharma, *Deep learning techniques for biomedical and health informatics*. Academic Press, 2020.

[32] C. Chhabra and M. Sharma, "Machine Learning, Deep Learning and Image Processing for Healthcare: A Crux for Detection and Prediction of Disease," in *Proceedings of Data Analytics and Management: ICDAM 2021, Volume 2*, Springer, 2022, pp. 305–325.

[33] A. Bhandare, M. Bhide, P. Gokhale, and R. Chandavarkar, "Applications of Convolutional Neural Networks," *International Journal of Computer Science and Information Technologies*, vol. 7, no. 5, pp. 2206–2215, 2016.

[34] B. W. Ballard, *Hands-On Deep Learning for Images with TensorFlow*. 2018.

[35] J. Liu, "Convolutional neural network-based human movement recognition algorithm in sports analysis," *Front Psychol*, vol. 12, p. 663359, 2021.

[36] J. Brownlee, *Deep learning for computer vision: image classification, object detection, and face recognition in python*. Machine Learning Mastery, 2019.

[37] A. Rosebrock, *Deep learning for computer vision with python: Starter bundle*. PyImageSearch, 2017.

[38] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, “A survey of convolutional neural networks: analysis, applications, and prospects,” *IEEE Trans Neural Netw Learn Syst*, 2021.

[39] W. Ballard, *Hands-on deep learning for images with TensorFlow: build intelligent computer vision applications using TensorFlow and Keras*. Packt Publishing Ltd, 2018.

[40] S. Albelwi and A. Mahmood, “A framework for designing the architectures of deep convolutional neural networks,” *Entropy*, vol. 19, no. 6, p. 242, 2017.

[41] D. P. Kingma and J. L. B. Adam, “A method for stochastic optimization. Int. Conf. Learn,” *Representations (ICLR)*, vol. 4, 2015.

[42] E. Altszyler, A. C. Ventura, A. Colman-Lerner, and A. Chernomoretz, “Ultrasensitivity in signaling cascades revisited: Linking local and global ultrasensitivity estimations,” *PLoS One*, vol. 12, no. 6, p. e0180083, 2017.

[43] G. Saini, A. Khamparia, and A. K. Luhach, “Classification of plants using convolutional neural network,” in *First International Conference on Sustainable Technologies for Computational Intelligence: Proceedings of ICTSCI 2019*, Springer, 2020, pp. 551–561.

[44] H. Gholamalinezhad and H. Khosravi, “Pooling methods in deep neural networks, a review,” *arXiv preprint arXiv:2009.07485*, 2020.

[45] C. Chen, J. H. Chuah, R. Ali, and Y. Wang, “Retinal vessel segmentation using deep learning: a review,” *IEEE Access*, vol. 9, pp. 111985–112004, 2021.

[46] P. Hao *et al.*, “Radiographs and texts fusion learning based deep networks for skeletal bone age assessment,” *Multimed Tools Appl*, vol. 80, pp. 16347–16366, 2021.

[47] N. Jindal *et al.*, “Fuzzy logic systems for diagnosis of renal cancer,” *Applied Sciences*, vol. 10, no. 10, p. 3464, 2020.

[48] P. Barwala, “Non-alcoholic fatty liver disease and its complications,” *Journal of Advanced Medical and Dental Sciences Research*, vol. 9, no. 9, 2021.

[49] Alaa Jamal Jabbar Kadhim, “BONE AGE ASSESSMENT USING A NEW DEEP NEURAL NETWORK ARCHITECTURE,” 2023.

[50] D.-X. Zhou, “Universality of deep convolutional neural networks,” *Appl Comput Harmon Anal*, vol. 48, no. 2, pp. 787–794, 2020.

[51] S. Indolia, A. K. Goswami, S. P. Mishra, and P. Asopa, “Conceptual understanding of convolutional neural network-a deep learning approach,” *Procedia Comput Sci*, vol. 132, pp. 679–688, 2018.

[52] L. G. K. T. B. BE, “Setio AAA Ciompi F Ghafoorian M van der Laak JA van Ginneken B Sánchez CI A survey on deep learning in medical image analysis,” *Med Image Anal*, vol. 42, no. 1995, p. 60, 2017.

[53] J.-Á. Román-Gallego, M.-L. Pérez-Delgado, and S. V. San Gregorio, “Convolutional neural networks used to date photographs,” *Electronics (Basel)*, vol. 11, no. 2, p. 227, 2022.

[54] Z. Liao *et al.*, “CNN Attention Guidance for Improved Orthopedics Radiographic Fracture Classification,” *IEEE J Biomed Health Inform*, vol. 26, no. 7, pp. 3139–3150, 2022.

[55] H.-G. Predel, “Marathon run: cardiovascular adaptation and cardiovascular risk,” *Eur Heart J*, vol. 35, no. 44, pp. 3091–3098, 2014.

[56] H. K. Wasike, S. T. Njenga, and G. M. Wambugu, “A Review: Diseased Leaf Feature Extraction Using Machine Learning Classification Techniques”.

[57] A. Patil and M. Rane, “Convolutional neural networks: an overview and its applications in pattern recognition,” *Information and Communication Technology for Intelligent Systems: Proceedings of ICTIS 2020, Volume 1*, pp. 21–30, 2021.

[58] F. Sultana, A. Sufian, and P. Dutta, “Advancements in image classification using convolutional neural network,” in *2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, IEEE, 2018, pp. 122–129.

[59] G. Eraslan, Ž. Avsec, J. Gagneur, and F. J. Theis, “Deep learning: new computational modelling techniques for genomics,” *Nat Rev Genet*, vol. 20, no. 7, pp. 389–403, 2019.

[60] M. Conor, “Machine learning fundamentals. Cost functions and gradient descent.” 2021.

[61] R. Xie, X. Jia, L. Wang, and K. Wu, “Energy efficiency enhancement for cnn-based deep mobile sensing,” *IEEE Wirel Commun*, vol. 26, no. 3, pp. 161–167, 2019.

[62] C. Yang, J. Liu, and C. Shi, “Extract the knowledge of graph neural networks and go beyond it: An effective knowledge distillation framework,” in *Proceedings of the web conference 2021*, 2021, pp. 1227–1237.

[63] A. G. Asmare, B. L. Thamineni, H. K. Dasari, and S. Woldetsadik, “Development of Chronic Kidney Disease Risk Prediction and Management System-Research study,” 2023.

[64] D. Chicco and G. Jurman, “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation,” *BMC Genomics*, vol. 21, pp. 1–13, 2020.

[65] G. Duffy *et al.*, "High-throughput precision phenotyping of left ventricular hypertrophy with cardiovascular deep learning," *JAMA Cardiol*, vol. 7, no. 4, pp. 386–395, 2022.

[66] M. Hussain, J. J. Bird, and D. R. Faria, "A study on cnn transfer learning for image classification," in *Advances in Computational Intelligence Systems: Contributions Presented at the 18th UK Workshop on Computational Intelligence, September 5-7, 2018, Nottingham, UK*, Springer, 2019, pp. 191–202.

[67] S. Narkhede, "Understanding auc-roc curve," *Towards Data Science*, vol. 26, no. 1, pp. 220–227, 2018.

[68] "'Eye_Diseases_Classification|Kaggle. Available online: https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification (accessed on 23 December 2022).'.".

[69] H. Perez, J. H. M. Tah, and A. Mosavi, "Deep learning for detecting building defects using convolutional neural networks," *Sensors*, vol. 19, no. 16, p. 3556, 2019.

[70] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," *arXiv preprint arXiv:1603.07285*, 2016.

[71] L. Wen, X. Li, and L. Gao, "A transfer convolutional neural network for fault diagnosis based on ResNet-50," *Neural Comput Appl*, vol. 32, pp. 6111–6124, 2020.

[72] A. S. B. Reddy and D. S. Juliet, "Transfer learning with ResNet-50 for malaria cell-image classification," in *2019 International Conference on Communication and Signal Processing (ICCSP)*, IEEE, 2019, pp. 945–949.

[73] X. Xia, C. Xu, and B. Nan, "Inception-v3 for flower classification," in *2017 2nd international conference on image, vision and computing (ICIVC)*, IEEE, 2017, pp. 783–787.

[74] L. Beyer, O. J. Hénaff, A. Kolesnikov, X. Zhai, and A. van den Oord, "Are we done with imagenet?," *arXiv preprint arXiv:2006.07159*, 2020.

[75] X. Yao, X. Wang, Y. Karaca, J. Xie, and S. Wang, "Glomerulus classification via an improved GoogLeNet," *IEEE Access*, vol. 8, pp. 176916–176923, 2020.

[76] A. Singla, L. Yuan, and T. Ebrahimi, "Food/non-food image classification and food categorization using pre-trained googlenet model," in *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, 2016, pp. 3–11.

[77] D. Sinha and M. El-Sharkawy, "Thin mobilenet: An enhanced mobilenet architecture," in *2019 IEEE 10th annual ubiquitous computing, electronics & mobile communication conference (UEMCON)*, IEEE, 2019, pp. 280–285.

[78] A. Shamsan, E. M. Senan, and H. S. A. Shatnawi, “Automatic Classification of Colour Fundus Images for Prediction Eye Disease Types Based on Hybrid Features,” *Diagnostics*, vol. 13, no. 10, p. 1706, 2023.

[79] S. Liu *et al.*, “A deep learning-based algorithm identifies glaucomatous discs using monoscopic fundus photographs,” *Ophthalmol Glaucoma*, vol. 1, no. 1, pp. 15–22, 2018.

[80] R. Sundaram, R. Ks, and P. Jayaraman, “Extraction of blood vessels in fundus images of retina through hybrid segmentation approach,” *Mathematics*, vol. 7, no. 2, p. 169, 2019.

[81] B. Khomri, A. Christodoulidis, L. Djerou, M. C. Babahenini, and F. Cheriet, “Particle swarm optimization method for small retinal vessels detection on multiresolution fundus images,” *J Biomed Opt*, vol. 23, no. 5, p. 56004, 2018.

[82] M. S. Junayed, M. B. Islam, A. Sadeghzadeh, and S. Rahman, “CataractNet: An automated cataract detection system using deep learning for fundus images,” *IEEE Access*, vol. 9, pp. 128799–128808, 2021.

[83] S. Gayathri, V. P. Gopi, and P. Palanisamy, “A lightweight CNN for Diabetic Retinopathy classification from fundus images,” *Biomed Signal Process Control*, vol. 62, p. 102115, 2020.

[84] F. Olayah, E. M. Senan, I. A. Ahmed, and B. Awaji, “AI techniques of dermoscopy image analysis for the early detection of skin lesions based on combined CNN features,” *Diagnostics*, vol. 13, no. 7, p. 1314, 2023.

[85] P. Saranya, R. Pranati, and S. S. Patro, “Detection and classification of red lesions from retinal images for diabetic retinopathy detection using deep learning models,” *Multimed Tools Appl*, pp. 1–21, 2023.

[86] C. Bhardwaj, S. Jain, and M. Sood, “Hierarchical severity grade classification of non-proliferative diabetic retinopathy,” *J Ambient Intell Humaniz Comput*, vol. 12, pp. 2649–2670, 2021.

[87] Rosser, Sylvia L. "The Eye Book: Eye and Eye Problems Explained." (2001): 865.

[88] Ponraj, D. Narain, et al. "A survey on the preprocessing techniques of mammogram for the detection of breast cancer." Journal of Emerging Trends in Computing and Information Sciences 2.12 (2011): 656-664.

[89] Vijay, Vallabhuni, et al. "A simple and enhanced low-light image enhancement process using effective illumination mapping approach." Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB). Springer International Publishing, 2019.

[90] Fu, Xueyang, et al. "A fusion-based enhancing method for weakly illuminated images." Signal Processing 129 (2016): 82-96.

[91] Li, Yikuan, Hanyin Wang, and Yuan Luo. "A comparison of pre-trained vision-and-language models for multimodal representation learning across medical images and reports." *2020 IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE, 2020.

[92] He, Kaiming, Ross Girshick, and Piotr Dollár. "Rethinking imagenet pre-training." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019.

[93] Dalsasso, Emanuele, et al. "SAR image despeckling by deep neural networks: From a pre-trained model to an end-to-end training strategy." *Remote Sensing* 12.16 (2020): 2636.

RESUME

Sajad Abdlkadhim Abdlhusein ALKHAYKANE he researcher enrolled at Imam, where He studied in the Computer Technology Engineering Department and graduated in 2018. After completing his studies, he worked as a teaching assistant at alayen University. In 2021, he began his graduate studies at KARABÜK University Institute of Higher Education to pursue a master's degree.