Deep Learning based Automated Diagnosis of Ocular Toxoplasmosis in Fundus Images using Convolutional Neural Network

by

Prionto Kumar Choudhury
19301089
Asma Akter Anika
19301038
Sumaiya Rahman Ramisa
19301147
Arsi Zaman
19301103
Rizvee Rifat Chowdhury
19101502

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
School of Data and Sciences
BRAC University
January 2024

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Prionto

Prionto Kumar Choudhury
19301089

Asma Akter Anika

Asma Akter Anika
19301038

Sumaiya Rahman Ramisa

Sumaiya Rahman Ramisa
19301147

Arsi Zaman

Arsi Zaman
19301103

Rizvee Rifat Chowdhury

Rizvee Rifat Chowdhury
19101502

Approval

The thesis/project titled "Deep Learning based Automated Diagnosis of Ocular Toxoplasmosis in Fundus Images using Convolutional Neural Network" submitted by

1. Prionto Kumar Choudhury (19301089)
2. Asma Akter Anika (19301038)
3. Sumaiya Rahman Ramisa (19301147)
4. Arsi Zaman (19301103)
5. Rizvee Rifat Chowdhury (19101502)

Of Spring, 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on January 18, 2023.

Examining Committee:

Supervisor:
(Member)

Images.jpg

Dr.Md. Golam Rabiul Alam

Professor

Department of Computer Science and Engineering

Brac University

Co-supervisor:
(Member)

Tanzim

Md. Tanzim Reza

Lecturer

Department of Computer Science and Engineering

Brac University

Program Coordinator:
(Member)

Images.jpg

Dr. Md. Golam Rabiul Alam

Professor

Department of Computer Science and Engineering

Brac University

Head of Department:
(Chair)

Images.jpg

Dr. Sadia Hamid Kazi

Chairperson and Associate Professor

Department of Computer Science and Engineering

Brac University

Abstract

Ocular toxoplasmosis is a disease caused by getting infected with Toxoplasma gondii.Untreated, ocular toxoplasmosis may lead to persistent infection and worsening symptoms. This can be costly enough if it needs to be diagnosed clinically. PCR has become more widely used for detecting parasite DNA in ocular samples. A computer based algorithm is now prefarable for viewing and reading accurately . So to address this issue, we have introduced an automated diagnosis that can identify the presence of ocular toxoplasmosis in the eye from the fundus image using convolutional neural networks (CNN). To foresee the presence of ocular toxoplasmosis, various binary cross entropy models are used. For classification, pretrained models such as VGG-19, MaskNet, ResNet50, and MobileNet are applied throughly to the fundus image dataset. having incorporated all these models, we acquired our best accuracy from our own customized CNN model, which is 95%.
Keywords: PCR;Binary cross entropy;Convulation Neural Network;VGG-19,MaskNet; ResNet50; MobileNet and Fusion model.

Acknowledgement

We begin by expressing our gratitude to the Almighty Allah for preserving our physical and mental well-being, allowing us to execute our thesis work titled " Deep Learning based Automated Diagnosis of Ocular Toxoplasmosis in Fundus Images using Convolutional Neural Network " seamlessly and within the designated time frame. We also acknowledge the divine endowment of knowledge, skills, and determination that empowered us to accomplish our task.
Our heartfelt appreciation goes out to our dedicated supervisor, Dr.Md.Golam Rabiul Alam, PhD, an Professor in Computer Science and Engineering and Md. Tanzim Reza sir for their invaluable guidance, unwavering support, continuous motivation, and constructive suggestions throughout our thesis journey. His enduring assistance has been instrumental in comprehensively developing our work and ideas.
The successful outcome of this study is not solely the result of individual toil, but rather a product of the collective efforts of numerous individuals. We extend our gratitude to our parents, whose unwavering hope and ongoing support have enabled us to pursue our career aspirations. Without their unwavering support, completing our thesis work would have been an uphill battle.

Table of Contents

List of Figures

List of Tables

Nomenclature

CNN Convolutional Neural Network

CV Computer Vision

$LSTM$ Long Short Term Memory

NLP Natural Language Processing

OT Ocular Toxoplasmosis

$RMSprop$ Root Mean Square propagation

SGD Stochastic Gradient Descent

$T.gondii$ Toxoplasma gondii

TP True Positive

VGG Visual Geometry Group

Chapter 1

Introduction

Ocular toxoplasmosis, caused by Toxoplasma gondii, is a significant eye infection with potential risks to vision. It is broadly classified into congenital and acquired forms, the former arising from maternal-fetal transmission and the latter from contaminated food or water. Research suggests that millions of people worldwide encounter ocular toxoplasmosis annually.Auto diagnosis system by computer is preferable now to quickly spot different ocular toxoplasmosis signs in eye pictures, just like skilled eye doctors do with retinal scans.The primary diagnostic approach involves inspecting distinct abnormalities fundus photographs. Deep learning, particularly Convolutional Neural Networks (CNNs), has significantly advanced disease understanding in the medical field. CNNs utilize fundus images to measure the likelihood of ocular toxoplasmosis and generate a heatmap highlighting the indicative region. Identification of ocular toxoplasmosis prompts further specification based on the dataset.Its depend on how large the data set is then the result will be more accurate.In the Ocular Toxoplasmosis Detection Challenge, we will introduce random transformations to the training dataset, validation, and test sets, ensuring their independence.In order to generate more accurate results in the image there will be two classes healthy and unhealthy.We will train the dataset in fusion model and ensemble model to get more better results than than the pre trained existing model.

1.1 Motivation

Our research is motivated by the urgent need to enhance ocular toxoplasmosis disease detection through advanced deep learning techniques.The disease's prevalence and impact on ocular health highlight the need for efficient detection methods.However, existing studies have left critical gaps in our understanding of efficient detection methods, particularly in the realm of deep learning for ocular toxoplasmosis. This paper aims to bridge these gaps by developing and implementing innovative deep learning models for accurate disease detection. Through this research, we aspire to contribute crucial insights to ophthalmology and address the current limitations in ocular toxoplasmosis detection knowledge.

1.2 Problem Statement

Ocular toxoplasmosis is a popular and possibly harmful ocular infection that has been linked to the protozoan parasite Toxoplasma gondii. The quick and precise diagnosis of this condition is required for the successful execution of efficient treatments and for the betterment of patient outcomes. However, the manual analysis of fundus pictures to identify lesions related to toxoplasmosis can be a laborious process that is prone to human error. Hence, there is a need for an automated and capable diagnostic system that is capable of precisely identifying ocular toxoplasmosis lesions in images of the fundus.

This study aims to develop an automated Ocular Toxoplasmosis detection system based on Convolutional Neural Network (CNN) technology and Deep Learning principles. The objective of the system is to examine fundus images with the goal of accurately identifying lesions associated with toxoplasmosis, while effectively distinguishing them from other ocular disorders. The Convolutional Neural Network (CNN) model will undergo training using a broad and diverse dataset of fundus images. This dataset will consist of a wide range of eye conditions, including both healthy eyes and those affected by toxoplasmosis. The system that is suggested should show the following features:

1. **Early Detection:** The automated diagnostic system has to have the capability of recognizing ocular toxoplasmosis lesions in its earliest stages, allowing rapid intervention and treatment.
2. **Precise Diagnosis:** The CNN model is expected to exhibit a high level of sensitivity and specificity in successfully identifying toxoplasmosis lesions, thereby minimizing the both false positive and false negative probability.
3. **Efficiency:** The automated system is predicted to demonstrate efficient processing of fundus images, enabling healthcare professionals to receive prompt results that aid in making timely and informed decisions.
4. **Localization and Progression Monitoring:** The CNN model will probably not only correctly identify toxoplasmosis lesions but also offer precise localization of the affected areas within the fundus images. Additionally, the system must have the capability to track the progression of the disease throughout its development.

The concept for Personalized Therapy Recommendations: The suggested automated diagnosis system includes the potential for seamless integration of genetic data and electronic medical records, that allows the delivery of specific therapy suggestions for each patient on an individual on an ongoing basis.

In order to achieve those goals, the research will focus on tackling challenges regarding the quality of the data set, optimisation. of algorithms, and interpret ability of the model. Furthermore, the execution of such a system in clinical practice will include regulatory and ethical considerations.

The efficient development and implementation of a Deep Learning-based Automated Diagnosis system for Ocular Toxoplasmosis has the possibility to greatly enhance ophthalmic care, improve patient outcomes, and lay foundations for similar applications in the automated diagnosis of multiple ocular and medical disorders.

1.3 Research Objective

The aim of this research is to create a system that automates diagnostics for Ocular Toxoplasmosis in Fundus Images applying Deep Learning and Convolutional Neural Network (CNN) technology. This study intends to address the challenges of manual diagnosis, such as individuality and complicated analysis, by employing deep learning algorithms.
The study's objective emphasizes on building an extensive set of fundus images that includes both healthy eyes and eyes affected by toxoplasmosis. It will develop and optimize a CNN model for precise detection and diagnosis of ocular toxoplasmosis lesions. The system will prioritize accomplishing high levels of sensitivity and specificity in order to minimize the chance of false positives and false negatives.
The automated system will offer early detection and precise localization of ocular toxoplasmosis lesions, enabling intervention in a timely manner. Additionally, this study looks at techniques for monitoring the progression of diseases over a period of time by utilizing long-term fundus image data.
The CNN model will undergo an in-depth assessment using suitable metrics and cross validation techniques. It will be compared to manual diagnoses carried out by experienced ophthalmologists. The model's decisions will be increased to enhance interpret ability and explain ability, which will give information regarding lesion identification.
The study investigates the potential integration of genetic data and electronic medical records for personalized therapy recommendations. Data privacy and regulatory compliance will be considered for potential clinical implementation.
The study is academically significant as it contributes to the fields of ophthalmology and medical imaging by demonstrating the efficacy as well as practicality of using Deep Learning based Automated Diagnosis for Ocular Toxoplasmosis. The possible effect of the automated system on ocular toxoplasmosis diagnosis and management is significant, as it can enhance accuracy, efficiency, and interpret ability. This possesses the potential of resulting in improved patient care and outcomes.

1.4 Paper Orientation

The primary purpose of this chapter is to enlighten the readers about toxoplasmosis. It briefly outlines the study's objectives and problem description.The essay's remaining sections are organized as follows: The background material for this investigation is outlined in Chapter 2, along with the concept of CNN and an explanation of why it is suitable for this research.Chapter 3 contains a literature review of a few prior research that classified healthy and unhealthy tissues in toxoplasmosis patients using ML and DL. The data set, its analysis, its reprocessing, Model Evaluation for use in the study are the main topics of Chapter 4. Chapter 5 of this article presents an introduction to the models utilized in the study, while Chapter 6 delves into a thorough examination of their findings and conclusions.

Chapter 2

Background

2.1 Toxoplasmosis

Toxoplasmosis is a parasitic illness caused by the protozoan worm Toxoplasma gondii. Globally, Toxoplasma gondii (T. gondii) infection is very common. According to research, a third of all people on this planet are chronically infected, up to 80% of whom have no signs of illness[1]. It is possible for this parasite to infect a variety of warm-blooded creatures, including people. Cats and other fields are Toxoplasma gondii's main hosts. Cats are the only hosts where the parasite may reproduce sexually; the parasite goes through various life stages in its life cycle.

2.2 Toxoplasmosis Life Cycles

The single-celled protozoan parasite Toxoplasma gondii is the source of the parasitic infection known as toxoplasmosis. Numerous warm-blooded species, including humans, are susceptible to infection[2,3] . The sexual and asexual stages of Toxoplasma gondii's life cycle are its two primary stages, and normally two separate hosts are involved in each stage.

2.2.1 Sexual Stage

The family Felidae, which includes domestic cats and other animals, is the specific host for Toxoplasma gondii. The parasite enters its sexual stage in the cat, which results in the development of oocysts in the intestine. This process is known as sexual reproduction.

2.2.2 Asexual Stage

Toxoplasma gondii uses warm-blooded creatures like humans and rodents and birds as intermediate hosts. Within these hosts, the parasite enters its asexual stage. Oocysts produced in cat feces or under cooked meat harboring tissue cysts containing the parasite are the two main ways that the parasite is transmitted to intermediate hosts.[29]

Life Cycle of Asexual Stage

The oocysts released when an intermediate host consumes them from contaminated food, water, or soil produce sporozoites that infiltrate the host's cells. These sporozoites subsequently go through a number of asexual divisions and become tachyzoites. The parasite's quickly reproducing form is known as a tachyzoite. Tachyzoites can spread throughout the body, infecting the host's tissues severely. In this acute stage, healthy people may experience very mild flu-like symptoms or none at all.[29] The infection, however, can be serious and even lethal in those with compromised immune systems or while pregnant. Some of the tachyzoites have the ability to develop into cyst-forming bradyzoites in reaction to the immune system of the host.[29] These cysts are exceptionally hardy and can survive in the host's tissues, notably in the muscles and brain. Latent toxoplasmosis is the name for this persistent, dormant stage of the infection.

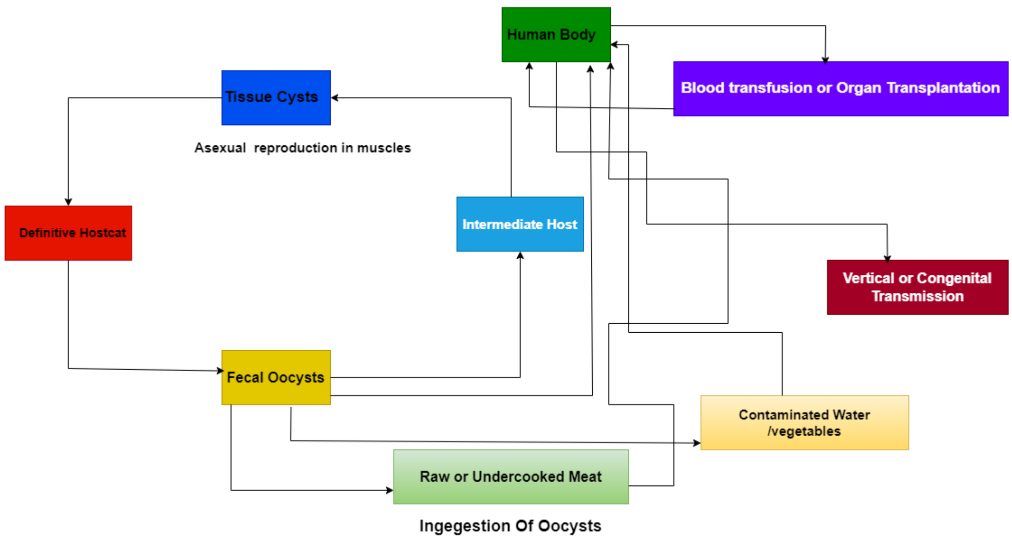


Figure 2.1: Life Cycle

2.3 Way Of Infected

Ocular toxoplasmosis is the most frequent cause of posterior uveitis worldwide, yet little is known about the frequency with which these lesions really develop in patients, how they impact their vision, or which traits make them more likely to have a poor prognosis. The main means of transmission of the illness are water, oocyst-infected fruits and vegetables, raw or undercooked meat with tissue cysts (with brady-zoites), and even vertical contact[2]. Most infected individuals remain asymptomatic for the entirety of their lives, and eye infection, a classic indicator of the illness, can present with a variety of symptoms, ranging from none at all to blindness.

2.4 Convolutional Neural Network

Convolutional Neural Network (CNN) is a deep learning-based neural network model that pro-ceases data using a grid pattern, or images. It was developed by analyzing how the visual cortex of animals is organized [1, 2], which enables it to distinguish between low-level and high-level feature patterns. Convolutional neural networks (CNNs) are made up of three layers: pooling, convolution, and fully connected layers. The convolution and pooling layer's job is to extract features, and it can be repeated a number of times to extract more features. The fully connected layers get the extracted features so they can map them onto the output for tasks like categorization. The outputs of CNN get progressively more sophisticated as it adds additional layers. In order to improve the outputs' consistency with the "ground truth" labels, the model is trained using a variety of optimisation algorithms, such as back propagation, gradient descent etc.

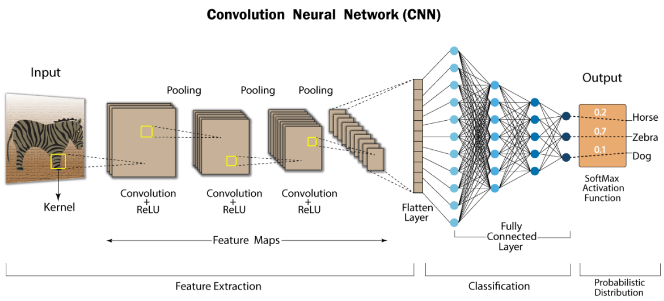


Figure 2.2: CNN[63]

2.5 Building Block of CNN Architecture

Several convolutional and pooling layers joined to fully connected layers make up a standard CNN. A typical design often consists of one or more fully linked layers that are followed by repetitions of numerous convolution layers and pooling layers. Forward propagation is the process of transforming input data into output.

2.5.1 Convolution Layer

The convolution layer is crucial in the extraction of features utilising linear and nonlinear mathematical operations like convolution operation and activation. Convolution is a mathematical operation that includes joining two functions to produce a third function. Two sets of data are combined. The tensor input data is processed using CNN using a small array of integers known as a kernel or filter. The kernel and also tensor components are combined to form an element-wise outcomes that is specific to each spot of the tensor, resulting in a feature map made up of outputs with discrete places in the output tensor. A number of various extra kernels are applied to the input tensor in order to extract different features from the data sets.

Size and number of kernels are the convolution operation's two primary arguments. The most typical kernel size is 33, but 55 and 77 are also frequent. Conventionally, convolutional methods do not permit the kernel's centre to overflow the input tensor's outermost element, which results in a loss of data in the feature map. This problem can be solved by using a method known as zero padding.[36] A stride is typically one and is defined as the distance between two successive kernel positions. A larger one is occasionally used to downscale the feature maps. Weight sharing, which describes how the kernels are distributed among different picture locations, is a crucial component of the convolutional procedure. Having the ability for other kernels to use the local features found by one kernel as variables avoids the need to repeatedly identify local features.
Downsampling and a pooling operation can produce a larger feature map and allow for the learning of spatial hierarchies of feature patterns. Additionally, by lowering the number of necessary parameters, the model's effectiveness can be improved

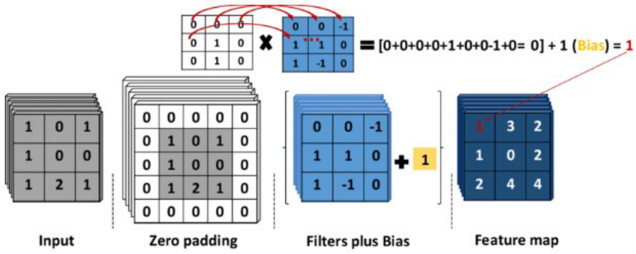


Figure 2.3: Convolution Layer

2.5.2 Non-Linear Activation Function

Later, a non-linear activation function receives the feature map the convolution layer produced. The ReLU is the most typical activation function.

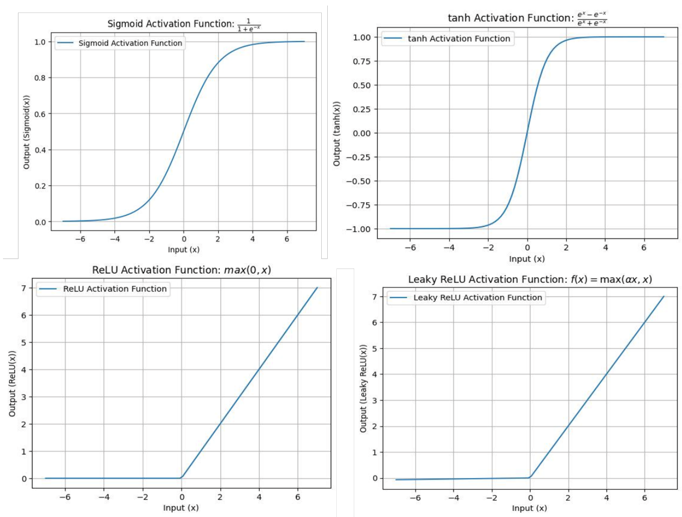


Figure 2.4: Non-Linear Activation Function

2.5.3 Pooling Layer and Max Pooling

The pooling layer is used to lower the size of the feature map in order to produce a translation consistency that can detect minute deviations and distortions and assists in reducing the amount of values learned throughout training. Max pooling, which partitions the feature map into collections of patches, chooses the maximum value for every patch, and ignores the remaining values, is the recommended pooling procedure. The size and stride of the most popular max pooling filter are 2 and 2, respectively. In this scenario, the feature map is down sampled by a factor of 2.

2.5.4 Fully Connected Layer

One or more fully connected layers, in which every input is coupled to every output by a learn able weight, are connected to the outputs of the convolutional and pooling layers. After the convolution layer has extracted the features and the pooling layer has down sampled them, the features are then passed on to fully connected layers, which generate the final outputs, such as classification probabilities, etc. The final fully linked layer's output nodes are determined by the number of classes. A nonlinear function, such as ReLU, as previously mentioned, follows each layer.

2.5.5 Last Layer Activation Function

The activation function employed in the final layer differs from the activation functions used in the earlier levels, and it depends on the task.

2.6 Training Network

A network is trained to identify a set of distinct convolutional kernels and fully connected layer weights that provide outputs with the least amount of variance from the input labelled dataset.

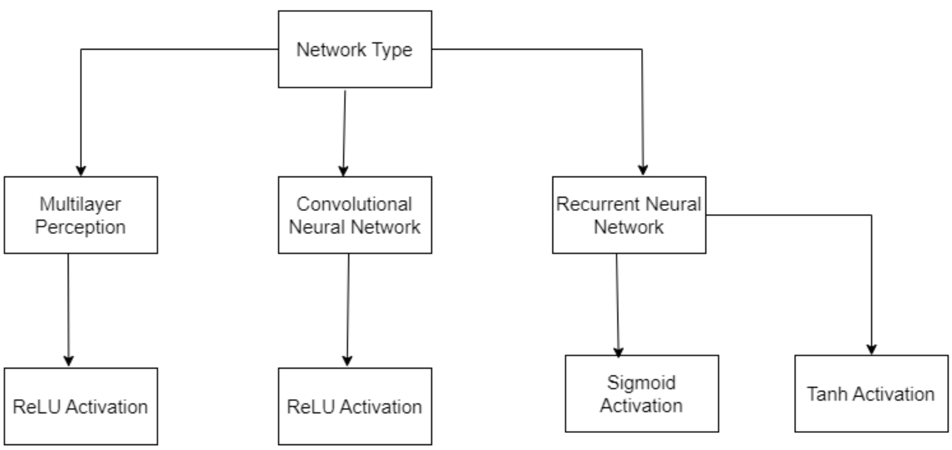


Figure 2.5: Choose Hidden layer in Activation Functions

The primary approach for training neural networks with hidden layers, backpropagation, also uses the gradient descent optimization process and the loss function.

2.6.1 Loss Function

The cost is calculated by the loss function via forward propagation based on the difference between the actual output and the network output. The most popular loss function for multi class classification is cross-entropy, while regression to continuous values typically makes use of mean squared error. In this investigation, the binary cross entropy was used.

2.6.2 Gradient Descent

Gradient descent is an optimization strategy that constantly modifies the network's learn able parameters, such as kernels and weights, with the main objective of minimizing loss. However, the approach used in this research is called the Adaptive Moment Task Last Layer Activation Function. Binary classification Single-class Multi class Sigmoid[36] Classification Softmax Various classes are classified The Sigmoid Regression to Continuous Values Identity 11 Estimation (Adam) method represents a more significant improvement than the general gradient descent method.

2.6.3 Adam Optimizer

Adam is an adaptive learning rate optimization algorithm that has been compared to a hybrid of the optimization methods root mean square propagation (RMSprop) and stochastic gradient descent (SGD). This is due to the fact that it actually squares gradients similarly to RMSprop and modifies the learning rate by using the gradient's moving mean instead of its SGD with momentum counterpart.[39] various factors lead to various learning rates because of the method's adaptability. Adam additionally retains the exponentially decaying average of the previous gradients. The typical values for 1 and 2 are 0.9 and 0.999 respectively[38].

2.6.4 Data and Ground Truth Labels

The most crucial components in every machine learning algorithm, including deep learning, are the data sets and the ground truth labels. In actuality, the data set and ground truth label of any such method and models determine their success. Because of this, it is imperative to carefully choose the data sets and ground truth labels, but doing so is costly and time-consuming. There are several widely available, high-quality sources for medical photos. However, the model requires data sets with specified ground truth labels in order to be used for a certain topic or function; as a result, more care must be given.Data sets typically fall into one of three categories: training, validation, or test. As the name implies, the network is trained using a training set, and learn able parameters are updated back into the network via back propagation while loss values are determined by forward propagation. Throughout the entire training process, the validation set is utilized for fine-tuning the hyper parameters and performing model selection. The completed model or network is tested on the test set, and its performance is assessed after being fine-tuned using training and assessment data sets. It is noteworthy that test and assessment sets are maintained separate.

2.7 Over-fitting

Because of over-fitting, the signal in a model has been replaced by statistical regularities specific to the training set when it performs badly on a fresh data set. In other words, rather than learning the signal, it learns the noise or supplementary data that is specific to the data set. Even if there are effective ways to limit over-fitting, having more training data is unquestionably the best strategy. There are alternative methods available because this is not always feasible, including regularization with dropout, weight decay, data augmentation, etc.Dropout is a regularization method that reduces the model's sensitivity to certain network weights by randomly choosing some activation's during training and setting them to 0. The "dropout" regularization method involves choosing a number of activation's at random.The regularization method known as "dropout" comprises changing a number of randomly selected activation's to 0 during training in order to reduce the sensitivity of the model to particular network weights[39].

2.8 Transfer Learning

Despite the fact that large data sets are ideal for model training, they are hard to come by. One way to solve this problem is to use transfer learning, which first trains the network model on a huge data set like ImageNet before reusing it for the target topic of interest. The essential idea is that features can be learned from a large data set and then shared amongst data sets that initially appear to be completely unrelated. Due to its ability to alter the generic features that have been learned from data sets, deep learning has the advantage of functioning effectively with small data sets in a range of domain tasks. models like AlexNet, VGG, ResNet, and others are examples of this [38].Although there are numerous uses for pre-trained networks, static extraction of features will be the main emphasis of this research. To keep the residual network—which is composed of a series of layers for pooling and convolution called the convolutional base and used as a fixed feature extractor—fully linked layers from a neural network that has been trained in an extensive database must be removed. The new classifier can only be trained on a selected data set of interest by adding fully linked CNN layers to the static feature extractor.

2.9 Deep Learning and it's use in Ocular Toxoplasmosis

Artificial neural networks are used in deep learning, a kind of machine learning, to model and identify patterns in large amounts of complex data. It has multiple uses in many different industries, including healthcare. Deep learning has the potential to significantly affect various aspects of ocular toxoplasmosis.

1. **Diagnosis and Detection:** In order to identify symptoms of ocular toxoplasmosis, deep learning algorithms can be trained to examine medical images like fundus photos, OCT scans, or retinal imaging[29,31]. These algorithms can be trained to recognize particular disease-related patterns and lesions, assisting ophthalmologists in making precise and timely diagnosis.

2. **Segmentation:** Retinochoroiditis is one of many disorders and abnormalities in the eye that can result from ocular toxoplasmosis. In order to provide comprehensive information regarding the amount and location of the lesions, deep learning algorithms can be utilized to segment the affected areas in medical images. This can help in tracking the development of the illness and evaluating therapy results.[31]

3. **Prognosis and Disease Progression:** Deep learning algorithms are able to forecast the course and prognosis of the disease by examining longitudinal data from individuals with ocular toxoplasmosis.[2,5] These models can use a number of patient-specific variables and data from medical imaging to offer individualized insights on the progression of the disease.[31]

4. **Drug Development and Treatment Planning:** To help in the identification of drugs for the treatment of ocular toxoplasmosis, deep learning can be used to analyze huge data sets, including molecular data and drug-target interactions. Deep learning models can also help with treatment plan optimisation and therapy response prediction based on patient characteristics.[31]

5. **Risk Assessment:** Deep learning algorithms can be applied to determine how likely a community or individual is to contract ocular toxoplasmosis. These models can assist in identifying those who are more likely to contract the disease by taking into account a variety of risk factors, including age, immunological state, and geographic location.

6. **Public Health Surveillance:** Deep learning algorithms can be used to analyze data from a range of sources, including public health databases and electronic medical records, to track the prevalence and spread of ocular toxoplasmosis in different regions.[2,31] Health authorities may utilize this to support the implementation of targeted interventions and control measures.

Chapter 3

Literature Review

The author Chakravarthy et al.[21] of this paper focused on A contagious illness called ocular toxoplasmosis brought on by the parasite Toxoplasma gondii, which eats away at good retinal cells. OT diagnosis is a difficult process. Our method involves first developing a network to identify unwell and well-picture portions extracted from a fundus picture. The next level involves creating a a duplex input hybridized system that can take photos and the related heat maps and accurately detect photos that are unwell When recognizing unhealthful images, the hybrid CNN achieves preciseness that is superior to the patch level system by combining photos and patch level data.As part of the pre-processing, the contrast was adjusted using the contrast bounded dynamic histogram, and all photos in the two datasets were resized to (five hundred and twelve *five-hundred and twelve) pixels in parameters. They developed a deep CNN with a two-class output for proportion phase and photo phase classification of fundus photos using the VGG16 architecture programmed on the ImageNet records. For evaluation purposes, the model was programed for hundred epochs for each of the three sample ratios specified in the patch phase CNN. For the testing ratio of seventy/thirty, our model obtained an AUC of (0.949). When the hybridized approach's total results are taken into account, the (seventy/thirty) sampling ratio outperformed (fifty/fifty) and (thirty/seventy) in each of the assessment criteria.

In this study, The author Abeyrathna et al.[22] concentrate on the segmentation of fundus pictures with ocular scars and lesions brought on by OT, as well as the detection of all OT scars and lesions, as well as their exact boundaries.They first constructed a cutting-edge instance classification network based on Mask R-CNN for dividing OT lesions in corneal fundus photos. Second, they create a novel unsupervised learning-based methodology for expert led studies of ground truth labels for instance segmentation network optimization. The proposed method uses a pre-trained CNN to extract features from the ground truth. K-means clustering is then applied to the feature space to construct tiny clusters of predicted and real-world instances with related characteristics.We demonstrate that this strategy can enhance segmentation effectiveness by evaluating just thirty three percent misclassified examples and then build strategies for optimizing the networks on the basis of professional's suggestion on the misclassified occurrences. Additionally, studying sixty-six percent of those cases yields the similar progress as studying hundred percent of them, demonstrating a thirty four percent decrease in manual labor without sacrificing productivity.

Alam et al.[3]researched Ocular toxoplasmosis (OT) which is a very well known eye disease brought on by T. gondi that can impair eyesight, has not received much attention in research. The authors have developed a benchmark research that assesses the performance of current established networks employing transfer learning approaches to identify OT from fundus photos in order to address this issue. The effectiveness of transfer-learning based segmentation networks for segmenting infections in the photos has also been studied, and in-depth analyses of various feature extraction techniques have been carried out to determine the best technique for OT division and categorization of lesions. We have tested previously trained methods including (VGG16, MobileNetV2, InceptionV3, ResNet50, and DenseNet121) models for classification tasks. MobileNetV2 performed better than all other methods in terms of (Accuracy and Recall, and F1 Score), surpassing InceptionV3 by 0.7% in Acc. However, DenseNet121 outperformed MobileNetV2 in terms of Precision, coming in with a 0.1% advantage. This work has utilized U-Net architecture for the segmentation task.(MobileNetV2 and InceptionV3 and ResNet34, and VGG16) were used to upgrade various structure in order to use transfer learning. . When the Jaccard loss function is used during the training, MobileNetV2/UNet outperformed ResNet34 in terms of Acc and Dice Score by 0.5% and 2.1%, respectively.

Park et al [57] shows that ocular toxoplasmosis is a condition brought on by either congenital or acquired infection with Toxoplasma gondii.Once inside the retina, the parasite multiplies before rupturing host cells and invading neighboring cells to produce primary lesions. Sometimes the confined parasite can be activated by the host defenses in the primary scar, infecting a nearby lesion. The host immunity in the initial scar can occasionally activate the confined parasite, causing it to infect an adjacent lesion. The main symptom of ocular toxoplasmosis patients is blurred vision, which can be identified by looking for antibodies or parasite DNA. If untreated, ocular toxoplasmosis can sometimes result in vision loss 1 since it requires treatment with numerous drug combinations to get rid of the parasite and the accompanying inflammation.As a zoonotic pathogen, Toxoplasma gondii is a common, obligate intracellular parasite that affects both humans and warm-blooded animals. T. gondii 23 is thought to infect almost one-third of all people on the planet on a chronic basis. The prevalence of the illness and the causes of infection, however, varied between geographical areas with various toxoplasmic settings, such as climatic conditions, dietary customs, and cleanliness standards. Toxoplasmic retinochoroiditis, which accounts for 30-55.

The use of residuals Neural networks (ResNets) for the automatic diagnosis of ocular toxoplasmosis (OT) from fundus images is covered by Parra et al.[4]. The authors emphasize the potential uses of ResNets in ophthalmology and give an outline of the present state of research in this area. They also talk about the difficulties in using ResNets to diagnose OT, and they make suggestions for future studies. The authors next give a thorough review of recent findings in this area, emphasizing ResNets' potential for use in OT diagnosis. They go over how to segment OT lesions using ResNets and how to categorize fundus images into normal and pathological categories. They also go through how ResNets may be used to identify OT lesions in fundus pictures. Finally, the authors talk about the difficulties in applying ResNets to OT diagnosis. They emphasize the significance of huge data sets and high-quality data for machine learning research. A data set of samples is used to fine-tune a pre-trained residual neural network. The results demonstrate that the suggested model is quite promising, with sensitivity and specificity rates of 94% and 93% respectively.

The use of machine learning in ocular imaging modalities is covered by Tong et al[5]. The authors emphasize the potential uses of machine learning in ophthalmology and give an outline of the present state of research in this area. They also talk about the difficulties in applying machine learning to ocular imaging and offer suggestions for future studies.The authors then give an outline of the various ocular imaging modalities and the difficulties each modality faces. Additionally, they talk through the various machine learning methods that are frequently employed in Ophthalmology. They go into how machine learning is used to diagnose glaucoma, age-related macular degeneration, and diabetic retinopathy, among other eye conditions. Additionally, they go over the use of machine learning to retinal image analysis and the identification of anomalies in the optic nerve head.Finally, the authors talk about the difficulties in using machine learning to ocular imaging studies. They emphasize the significance of huge datasets and high-quality data for machine learning research. They also talk about how data gathering and analysis procedures need to be standardized.

Medina et al.[6] show that UWF imaging is a powerful tool for detecting and monitoring retinal illnesses, with the potential to improve clinical evaluation quality and broaden the use of tele ophthalmology in providing efficient eye care.The author compares UWF fundus photography to traditional ophthalmoscopy for diagnosing and classifying various retinal disorders.The results showed that graders had high agreement in categorizing retinal disorders using both UWF imaging and ophthalmoscopy. Apart from that, inter-rater agreement was nearly flawless, demonstrating that UWF imaging is a trustworthy alternative to ophthalmoscopy for identifying severe retinal disorders.UWF imaging offers benefits like accurate diagnosis, monitoring of retinal conditions, detailed assessments without discomfort, and validation for teleophthalmology applications, directing patients to appropriate clinical care pathways.

The incorporation of Retina Image Analysis by Jamal A et al [7] revolutionizes in the field of ophthalmology, allowing optometrists to increase their diagnostic capabilities and improve patient care, notably in non-invasive diagnosis and treatment planning. RIA is intended to aid in the study of retinal pictures, with a special emphasis on vascular discovery within the retina. This study's main objective was to give optometrists a powerful tool for examining retinal pictures through the use of an advance interface that offers a variety of options for image processing, analysis, and storage.The discovered vessels are displayed on the retinal picture by RIA for further investigation.Detecting irregularities in retinal pictures can lead to timely actions that could save a patient's sight.RIA improves diagnostic accuracy while decreasing the chance of misdiagnosis.RIA provides valuable insights into retina condition, aiding in better treatment planning and serving educational purposes, though it does not prescribe treatment or medication.

M. S. Khan et al.[8] in this study have tried to put up an innovative method to tackle the issue of significantly imbalanced data in the classification of eye diseases. Instead of addressing the initial multi-class classification problem, the researchers opted to convert it into a series of binary classification tasks, utilizing data-sets that were balanced in terms of class distribution. The execution of this strategic maneuver facilitated the deep learning model, particularly VGG-19, in obtaining an impressive level of accuracy when discerning between typical eye conditions and specific ailments such as myopia, cataract, and glaucoma. The implementation of this strategy intended to address the root cause of class imbalance had a significant role in the achievement of positive outcomes for their automated ocular disease detection system. This study has some problems, such as the possibility of an imbalance in the data,its limited effectiveness to other diseases and data-sets, its reliance on data quality, the lack of clinical validation, the challenges of understanding the model's decisions, ethical concerns, and the need for more research and enhancement in areas like ocular image segmentation and class imbalance handling using generative adversarial networks (GANs).

Computer-aided diagnostic (CAD) systems have started on an exciting new path in the effort to protect eye health. This paper introduces a multi-label convolutional neural network (ML-CNN) system based on multi-label classification (MLC) that is capable of simultaneously detecting a multitude of ocular diseases from color fundus images, a significant advancement over existing systems which tend to focus on individual ocular diseases.Here the researchers(E. AbdelMaksoud)[9] innovated the three pillars of pre-processing,modeling and prediction.Their ML-CNN displays exceptional performance, achieving 94.3 percent accuracy, 91.5 percent precision, and 96.7 percent AUC with the help of its three convolution layers, max pooling, dropout layers, and fully linked layers. Beyond its technological prowess, this study stands out for its optimistic outlook on a future where a large, balanced ML data set, com- bined with meticulous manual splitting, promises to further improve diagnostic accuracy and eliminate over-fitting, opening exciting avenues for future research in the field of ocular health.

Using Deep Learning (DL) models, authors Rodrigo Parra, Verena Ojeda, and their team [10]approach the problem of establishing trust in ocular toxoplasmosis (OT) diagnoses. The authors present an unusual method for evaluating the trustworthiness of DL models by comparing their decision criteria to those of ophthalmologists. Their results demonstrate the importance of trust in model selection alongside more conventional measures, especially for medical applications. To further the adoption of DL in the medical community, potential future enhancements include validation by ophthalmologists, alternate attribution approaches, and comparisons with traditional ML models.

Adithi D. Chakravarthy et.al[18]This study makes a substantial contribution to the field of medical image segmentation by offering an effective method for improving the accuracy of instance segmentation networks while lowering the manual workload for medical specialists. This research tackles the problem of enhancing the accuracy of instance segmentation networks in the field of medical picture segmentation, notably in the context of retinal fundus images comprising lesions and scars caused by ocular toxoplasmosis (OT). Focusing on the Mask R-CNN architecture, the authors suggest a method that combines directed finetuning and feature grouping to improve the performance of instance segmentation networks. The suggested strategy significantly improves model performance. Notably, the mean average precision (mAP) at a 0.5 IoU threshold increases by 20% while the mask average intersection over union (IoU) improves by 7%, proving the efficiency of the method. When compared to traditional methods, the research significantly minimizes the strain on medical specialists by reducing their involvement in the analysis process. The authors also intend to investigate how their method might be expanded to improve segmentation models without the use of proposals.

M. et al.[11] proposed employing a deep learning algorithm to differentiate between ocular toxoplasmosis (OT) lesions and normal fundus photographs. They collected fundus images of eyes with OT lesions from multiple uveitis facilities, annotated the images, performed patch-level classifications, and then generated a probability heat map using a sliding window protocol. They created a dual-input hybrid CNN model for detecting OT fundus images by integrating the heat map and patch features. Using metrics such as AUC, sensitivity, and specificity, the results showed that the model could be a useful diagnostic aid for ocular toxoplasmosis for clinicians.

In this scholarly work, M. Akil et al.[12] addresses the significant difficulty of detecting ocular pathology from fundus images in the healthcare industry. The complexity arises from the varying severity phases of pathologies, which can be identified by lesions with distinctive morphological characteristics. In addition, distinct pathologies may exhibit similar characteristics, and patients may be afflicted by multiple pathologies simultaneously. The process involves a complex multi-class classification strategy.The author underlines that methods based on deep learning outperform other approaches due to their ability to adapt network configurations to specific detection objectives. The study investigates these deep learning-based ocular pathology detection methods comprehensively, with a focus on lesion segmentation and pathology classification. The research explores the processing stages, neural network structures, hardware and software requirements, and experimental design principles for method evaluation.Nevertheless, the work identifies several challenges and variations within the field. Even among methods with similar objectives, there is substantial variation in processing techniques, network architectures, input data management, and performance evaluation techniques. The clinical context requires the simultaneous detection of multiple diseases during screening, which represents an immense task for the majority of deep learning-based methods.

The study[19] suggests an automated screening technique for eye illnesses utilizing fundus images that is based on deep learning. An automated screening technique for agerelated macular degeneration (AMD), glaucoma, diabetic retinopathy (DR), and a few other diseases is presented in this work.The system has a high Cohen's kappa score of 97.6 and a multiple disease classification accuracy of 97.0

Anneke Annassia Putri Siswadi [20] presents CAD models for detecting ocular abnormalities from single-color fundus photography. The first model focuses on detecting microaneurysms with high sensitivity and low number of FPI. The main challenge in MAs detection is the limited number of data with MA, causing severe data imbalance. The MAs detection consists of three main processes:

pre-processing, MAs candidate extraction, and MAs classification. The MAs classifiers are built using ensemble and cascade learning methods. The firstmodel uses the enhanced-green channel, background suppression image, and blue channel for MAs detection, while the second classifier uses cascade learning to reduce FPI with high sensitivity. The second model identifies 28 ocular abnormalities, including frequent and rare ones.The model adds a co-occurrence dependency factor, using linguistic features of labels as a semantic dictionary. Two approaches for multi-label detection with deep learning are proposed:CNN-based and Transformer-based semantic dictionary learning. The transformer-based approach achieves higher performance compared to CNN-based semantic dictionary learning.

J. G. Montoya et al.[24] discuss the diagnosis of toxoplasmic retinochoroiditis, highlighting the challenges in detecting a systemic immune response within a localized occurrence. The author's suggested using aqueous humor analysis for specific antibodies or parasitic DNA, but acknowledge the need for vitreous sampling. They also discuss the challenge of laboratory confirmation due to inter-individual variances in antibody production. They propose a tailored diagnostic algorithm for atypical clinical presentations and suggest that laboratorybased tests can improve clinical diagnoses. They also explore the enigmatic aspects of humoral immunity in toxoplasmosis, including the diagnostic window for false-negative results, differences in antibody detection, and the role of non-specific immune stimulators. They emphasize the need for further scholarly endeavors in this area.

J. E. Gomez-Marn et al.[25] conducted a study in Armenia-Quindo, Colom- ´ bia, to investigate the prevalence of retinochoroidal lesions caused by ocular toxoplasmosis and their association to risk factors. The researchers evaluated 161 people and discovered that 10.5% of them had retinochoroidal scars, indicating an old dormant Toxoplasma gondii infection. All 17 patients with these lesions tested positive for T. gondii antibodies. Bottled water intake was found to be a protective factor against T. gondii infection in this community. Despite such limitations, the study discovered a statistically significant protective factor in bottled water use. The risk factor assessment was based on patient interviews, which could create recall bias. To address this, a standardized questionnaire was utilized. The study emphasizes the significance of toxoplasmosis-related ocular lesions in Armenia-Quindo, Colombia, and advocates for promoting the consumption of boiled or bottled water as a significant preventive public health measure to reduce T. gondii infection and the subsequent onset of ocular toxoplasmosis.

The E. J. H. Goh et al.[1] of this comprehensive analysis proposes a multifaceted strategy to address the complex issues posed by ocular toxoplasmosis (OT). Their proposal calls for increased research into the specific mechanics of T. gondii's invasion of ocular tissues, with the ultimate goal of developing tailored pharmaceutical therapies to reduce the risk of ocular illness. Furthermore, the author emphasizes the critical advancement of diagnostic techniques, including the development of cutting-edge serotyping methodologies and specialized biomarkers meant to differentiate between distinct forms of OT. The review also highlights recent breakthroughs, such as localized therapy methods and the examination of interleukin-17A, as being important in improving diagnostic precision and therapeutic pathways for OT. Furthermore, the author highlights the possibility for customized treatment paradigms, precisely tuned to individual host susceptibilities and variations among parasite strains, to radically shift the landscape of addressing this vision-threatening condition. In summary, the author's suggestion is a comprehensive and unique strategy that combines an enhanced understanding of OT's pathophysiological complexities with inventive diagnostic and therapeutic methodologies, all with the ultimate goal of improving patient outcomes.

Clinical Manifestations of Ocular Toxoplasmosis" by Delair et al.[26] discuss the signs and treatments for ocular toxoplasmosis. The disease is caused by an infection with Toxoplasma gondii, either acquired or congenital. Once the parasite enters the retina, it multiplies within the host cells before rupturing them and invading adjacent cells to produce primary lesions. The primary symptom of ocular toxoplasmosis is impaired vision, which can be diagnosed by examining antibodies or parasite DNA. Vision loss may result from untreated ocular toxoplasmosis. Multiple treatment combinations are required to treat the parasite and its associated inflammation. The three distinct types of T. gondii are tachyzoites, tissue cysts (which contain bradyzoites), and oocysts. In approximately one-third of countries, T. gondii is believed to cause chronic infections in humans. However, the prevalence of the disease and the causes of infection varied between geographic regions with different toxoplasmic conditions, such as climate.

This comprehensive literature review explores the complex nature of ocular toxoplasmosis, focusing on its postnatally acquired form. The researcher (Kalogeropoulos)[27] synthesizes research from the PubMed database and Google Scholar, highlighting its clinical features, diagnostic methodologies, and therapeutic strategies. They suggest that identifying characteristic clinical findings is crucial for identifying the disease, and recommend laboratory confirmation through traditional antibody detection or PCR techniques. The author recommends conventional treatment regimens like oral pyrimethamine, sulfadiazine, and corticosteroids as the mainstay, while acknowledging alternative treatment modalities. The review concludes with a forward-looking perspective, envisioning future research in epidemiology, pathogenesis, diagnosis, and treatment to enhance our understanding of ocular toxoplasmosis and refine its management.

This extensive literature review highlights the utmost significance of tackling the zoonotic infection known as Toxoplasmosis, which is caused by the parasite Toxoplasma gondii, within the context of China. These people are particularly vulnerable to the negative impacts of infection, as seen by the significant consequences that can arise. Despite the scarcity of English journal publications from China,P. Zhou et al.[28] emphasize that the infection caused by T. gondii is a noteworthy public health issue inside the country. The present research provides an in-depth analysis of the clinical manifestations, modes of transmission, and prevalence of Toxoplasma gondii infection in the human population of China. Additionally, it provides a concise overview of the genetic characteristics observed in documented T. gondii isolates. The authors recommend the promotion of public education regarding the potential dangers linked to unhealthy dietary choices and lifestyle behaviors. They additionally suggest the expansion of serological screenings for certain demographic groups and the adoption of strategies aimed at improving both food safety and occupational safety. The authors conclude by offering comprehensive control strategies, which encompass the adoption of Hazard Analysis and Critical Control Points (HACCP) principles, educational initiatives aimed at mitigating environmental contamination, and the implementation of measures to ensure the safety of food consumptionand the inspection of meat for T. gondii contamination. These strategies are intended to effectively manage and diminish T. gondii infections in both human and animal populations within China.

The use of attention mechanisms in deep learning algorithms for computer vision and natural language processing, as well as their application in healthcare, is covered in the paper "A survey on attention mechanisms for medical applications: are we moving towards better algorithms?" by Tiago Gonçalves, Isabel Rio-Torto, Luís F. Teixeira, and Jaime S. Cardoso [39]. According to the kinds of jobs that may integrate many work streams within the medical domain, the research evaluates the use of attention mechanisms in machine learning techniques (such as Transformers) for multiple medical applications. With the use of an experimental case study on medical picture classification featuring three distinct use cases, the authors also suggest a critical examination of the assertions and possibilities of attention mechanisms put out in the literature. The article ends with a critical evaluation of the assertions and possibilities about attention mechanisms made in the literature, along with suggestions for future directions in medical research that could make use of these frameworks.

In order to perform convolution-free multilabel classification of satellite imagery for deforestation monitoring, a vision transformer model is used in this research. The author Kaselimi et al. [40] provide ForestViT, a multilabel vision transformer model that eliminates for any convolution operations found in widely used deep learning models for deforestation detection by utilizing the advantages of the self-attention mechanism. The model can simulate long-range spatial dependencies and handle imbalanced datasets. In addition, the authors demonstrate that the self-attention mechanism outperforms established CNN methods in deforestation monitoring, particularly when it comes to the dataset's less frequent classes. The identification and monitoring of deforestation has shown to be a successful application of remote sensing. Numerous pertinent research have shown that Landsat and MODIS images data have been widely employed for deforestation identification.

A novel method for large-scale vision models in remote sensing (RS) activities is put out by the authors of the study "Advancing Plain Vision Transformer Towards Remote Sensing Foundation Model" [41]The authors aim to offer massive vision models specifically designed for RS tasks, using simple vision transformers with approximately 100 million parameters. After looking into the performance of these massive models, the authors suggest replacing the original full attention in transformers with a new rotated varied-size window attention, which can drastically lower computational costs and memory footprints while improving object representation by extracting rich context from the generated diverse windows. Their models' performance on downstream classification and segmentation tasks is also competitive when measured against the state-of-the-art techniques. Additional tests demonstrate the benefits of their approaches in terms of data transmission efficiency and computational complexity.

The class-guided Swin Transformer (CG-Swin) is suggested in the study "Class-Guided Swin Transformer for Semantic Segmentation of Remote Sensing Imagery"[42] in order to segment remote sensing images semantically. Experimental results on ISPRS Vaihingen and Potsdam datasets show that the proposed method outperforms both recent Transformer-based and sophisticated CNN-based approaches on 10 benchmarks. The research proposes that CNN-based approaches achieve high representation capability by automatically learning semantic features, in contrast to conventional methods. The suggested technique uses a transformer-based encoder–decoder structure, introducing the class-guided Transformer block to build the decoder and the Swin Transformer backbone as the encoder.

In this paper, G.deng et al. [43] a deep learning model for VHR aerial image segmentation based on semantics termed Crisscross-Global Vision Transformers (CGVTs) is proposed[44]. For the purpose of resolving the confined receptive field problem, CGVTs take advantage of the transformer's capacity to acquire distant contextual data. A crisscross transformer encoder block (CC-TEB) and a worldwide squeeze transformer encoder block (GS-TEB) make up the two components of the model. Whereas GS-TEB improves worldwide representation of features ability, CC-TEB increases local feature representation ability and gets over the drawbacks of classic SA design. Using three superior resolution sensor segmentation of image datasets—Vaihingen, Zeebrugge, and LoveDA—the model has been verified. Regarding the semantic segmentation task on both datasets, the CGVTs performed very competitively in both quantitative and qualitative dimensions, coming in second place.

In this paper[44],the author(Kalogeropoulos et al.,) determines that The most frequent cause of infectious posterior uveitis is ocular toxoplasmosis, which is mostly diagnosed based on clinical symptoms.PCR has become more widely used for detecting parasite DNA in ocular samples, with increased sensitivity.. In many nations, toxoplasmic retinochoroiditis is a major cause of visual loss that lowers quality of life. Recent advancements in treatment and diagnostic methods have helped to prevent or minimize vision loss. Oral pyrimethamine, sulfadiazine, and corticosteroids are the current treatments, although alternative regimens work well as well. It is anticipated that more research will improve our knowledge of the epidemiology, pathophysiology, diagnosis, and management of this difficult clinical entity.

In this paper [45],the author(Zou et al.,2022) proposed a brand-new LESSFormer technique for classifying High-Selectivity Imagery (HSI) that incorporates HSI data attributes into the architecture of a vision transformer. Unlike single-dimension tokens, this approach creates low-level spectral-spatial tokens designed for HSI. Despite having little training samples, it is also able to model location. Extensive trials on three benchmark HSI datasets demonstrate the robustness and efficiency of the technique. Future research will explore self-supervised variants to reduce training data demand by extracting valuable semantic features from unlabeled HSI data.

The author (Zhang et al., 2019) of this paper [46]demonstrates in the case of semantic segmentation,among all the neural network models,deeplab series is one of the best as it includes channel attention module to accelerate the convergence speed of model training and enhance the accuracy of segmentation.Additionally, it provides insight into the segmentation efficacy of several semantic segmentation network models on a human-parsing sample data collection and demonstrates the superior segmentation impact of the suggested deep neural network design.In the topic of segmentation using semantics, there are two main lines of inquiry: deep learning-based image semantic segmentation and picture semantic segmentation based on visual graphics.The values of pixels in the picture are typically used for evaluation in approaches based on visual graphics. Because of its potent processing capacities, the advent of convolutional neural networks (CNN) offers a new avenue for semantic segmentation experiments.By conducting a number of tests, the author demonstrated experimental findings using Xception as the foundation, Enhanced Deeplabv3+ and ASPP, as well as When it comes to segmenting the human-parsing collection of dataset, SE-block performs the best.

The author (Chen et al., 2022) [47] explains that the core of a building extraction network in this study is a hierarchical vision transformer using shifting windows (swin), a potent multiscale feature learning module.For constructing the extraction process, we first created a generic framework that included a head network for feature fusion and refinement and a backbone for extracting multiscale properties. Next, using swin as the framework's foundation, we used channel- and spatial-wise augmentation in the head network. In comparison to the MAP-Net, the state-of-the-art (SOTA) methodology for constructing extraction from remote sensing pictures, our approach provides enhancements in both F1-score and intersection over union (IoU), according to data collected from experiments. The evaluations conducted on three publicly available datasets in this work show its benefits with regard to of all four widely used measures. They authors also discovered that extending the swin transformer's dimensions does not increase efficiency. Additionally, another discovery was found which is that the limited quantity and poor labeling quality of available training sets make it difficult to extract structures from satellite photos.

The paper[48] "SCViT: A Spatial-Channel Include Protecting Vision Transformer for Inaccessible Detecting Picture Scene Classification" presents a demonstration to progress the discriminative capacity of the Vision Transformer (ViT) show in farther detecting picture scene classification. The show presents a dynamic accumulation technique to extricate nearby basic highlights and a lightweight channel consideration module to consider diverse highlight channels. Exploratory studies indicate that SCViT progresses the discriminative capacity of ViT, particularly for complex spatial highlights, and can be decreased without noteworthy execution misfortune.

The paper [49]introduces a hybrid spatial-spectral convolutional vision transformer (SSViT) for hyperspectral change detection (HCD) tasks. The SSViT model combines the strengths of convolutional neural networks (CNN) and transformers to improve the effectiveness and efficiency of HCD. The authors focus on generating highly reliable pseudo sample data for training the SSViT model. They use different methods such as Euclidean distance, Chan-Vese segmentation, and sorting of change and no-change pixels to generate pseudo change sample data. The experimental results show that the proposed SSViT model outperforms the existing methods.

The paper [50]"Swin Transformer: Hierarchical Vision Transformer using Shifted Windows" introduces a novel hierarchical vision Transformer architecture called Swin Transformer. The key contributions of the paper include hierarchical representation, shifted window partitioning, linear computational complexity, and significant performance improvement in various vision tasks. The authors compare Swin Transformer with other architectures and demonstrate its superior accuracy and slightly faster speed. Overall, Swin Transformer achieves state-of-the-art performance on different vision benchmarks.

The author (Iyer et al.,2023) of this paper[51] demonstrated in the paper a patient was treated for acute focal toxoplasmosis chorioretinitis in the right eye with oral medication, intravitreal clindamycin, and intravitreal dexamethasone.. He showed up with a branch retinal artery occlusion (RAO) next to an inactive retinal scar nine months later. Paracentral acute middle maculopathy (PAMM) shows an arteriolar distribution, according to the results of widefield en face structural swept-source optical coherence tomography. After starting a daily aspirin regimen, PAMM disappeared after six months. The uncommon correlation between ocular toxoplasmosis and delayed onset RAO is underscored by this case, which is the first of its kind.

Chapter 4

Methodology

4.1 Workflow

The approach recommended for this thesis is illustrated in this section. The gathering, organization, and application of methods for pre-processing to the database mark the first steps of the workflow. To determine which model is most effective in detecting and binary classifying healthy and diseased eye images, the workflow consists of a standard CNN model and the transfer learning approach of four pre-trained CNN models (VGG16, VGG19, ResNet50, Mobilenet). The effectiveness of each of these models compared to using metrics such as accuracy, precision, recall, F1 score, Confusion Matrix, and AUC curve. The workflow for the methodology is shown in full in Figure 4.1 as a block diagram.

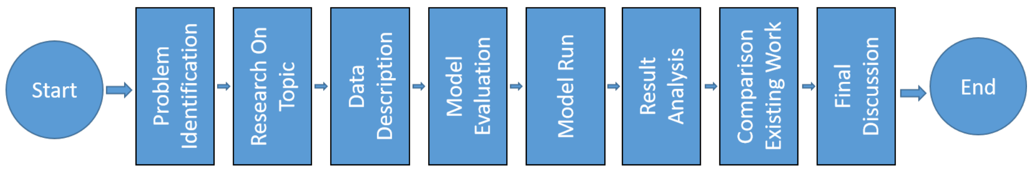


Figure 4.1: Workflow of Thesis

4.2 Data Description

Images of the eyes pictures were Collected at the **Hospital de Clnicas medical facility in Asuncion, Paraguay** and **Niños de Acosta Ñú General Pediatric Hospital** [57],are part of the Ocular Toxoplasmosis (OT) Fundus Images Data set. The data set is used to create models for toxoplasmosis in the eye that may be detected automatically. A prediction model used as a tool for OT diagnosis could help ophthalmologists, especially those with less experience, diagnose unusual cases and save time

4.3 Data Analysis

Two primary folders comprise the data set one containing all the photos that were gathered and the other including masks for lesions on eyes caused by ocular toxoplasmosis.

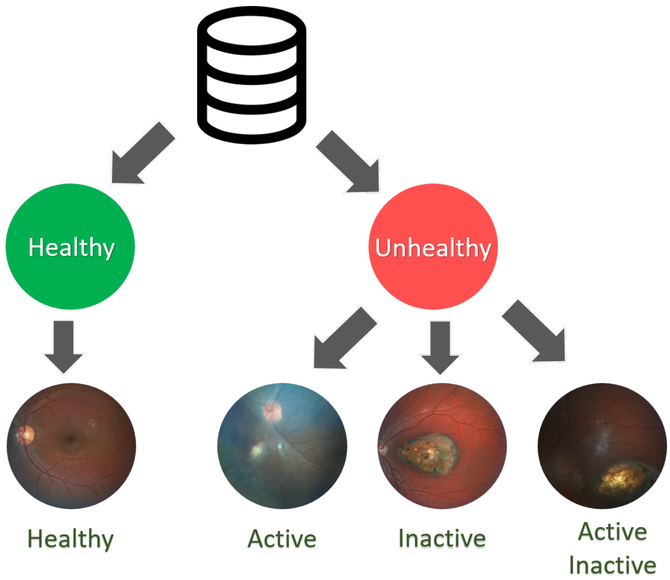


Figure 4.2: Data set Structure of Ocular Toxoplasmosis

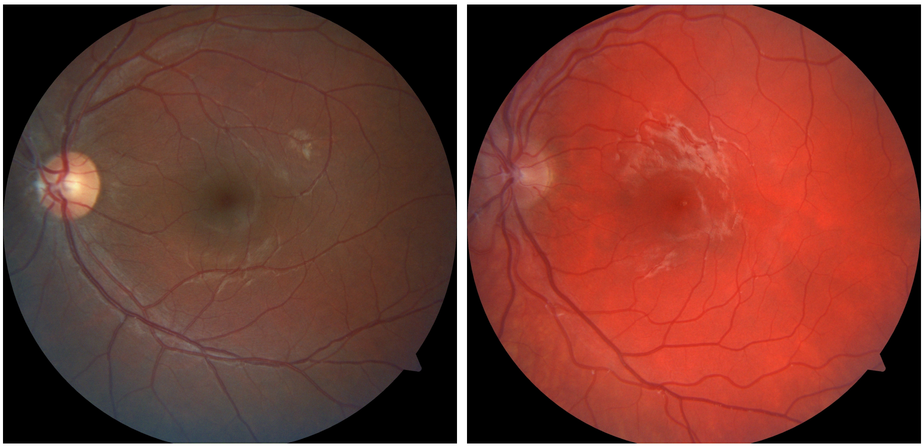
This folder contains the original eye images that are being studied. The eye itself, as well as any lesions or abnormalities, are likely to be included in these photos, along with other information. The healthy photographs and non-healthy images are both in the categorization folder. Images showing active, inactive and active inactive lesions carried on by ocular toxoplasmosis can be found in the non-healthy group.
Mask: This folder contains binary pictures (masks) with labels designating whether each pixel is a part of the lesion or not. A black-and-white map-like representation of the mask indicates which pixels in the associated eye image relate to the lesion or area of interest and which ones do not. Typically, lesions are represented by white (or 1) pixels, while healthy pixels are represented by black (or 0) pixels.

Importance of Masking

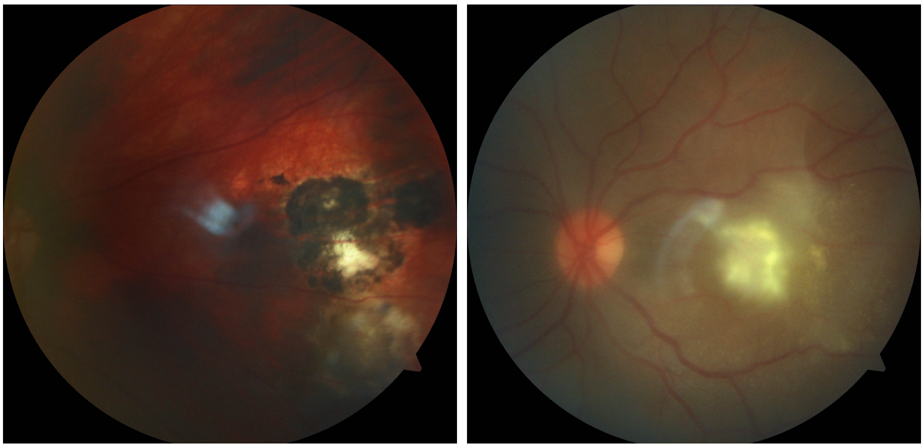
Masks are used for segment images. The method of segmenting an image is to divide it into various parts according to a set of criteria like Healthy, Active, Inactive, Active, Inactive. It is frequently used to isolate and detect certain characteristics, like lesions or tumours, within the source pictures in medical image analysis. The lesion area can be highlighted and extracted for additional examination or diagnosis by superimposing the mask over the original image.

Data Pre-processing

1. **Segmentation:** Given that this study's primary objective is to predict whether a patient have deases or not, we have train our split our dataset in

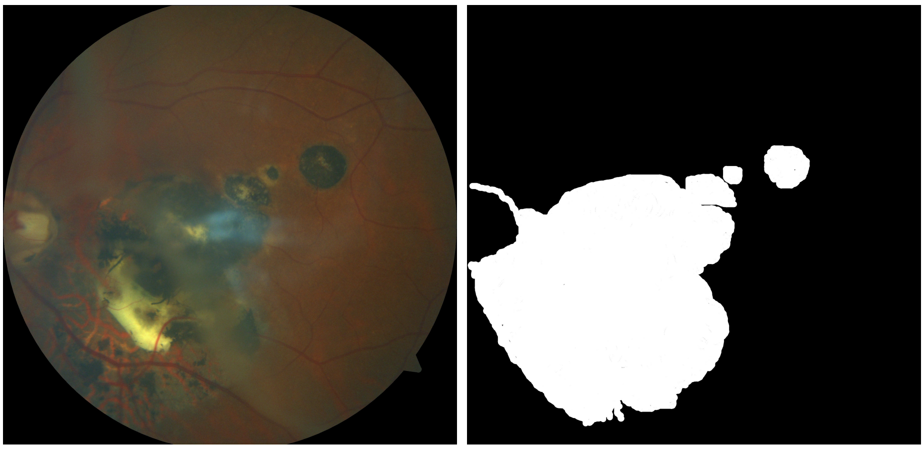


(a) Healthy (b) Healthy



(c) Disease (d) Disease

Figure 4.3: Example of Fundus Images from the data set



(a) 113 (b) 113_Mask

Figure 4.4: Images Masking.

to train, test. This spliting method helps to develop and evaluate deep learning models efficiently. It ensures to avoiding overfitting, fine-tuning hyperparameters, and evaluating the model's performance on unseen data to ensure that model performs well in actual situations. Consequently, the final dataset now included the initial 5200 photos divided into the 3 categories in a ratio of 4000:1200[39].

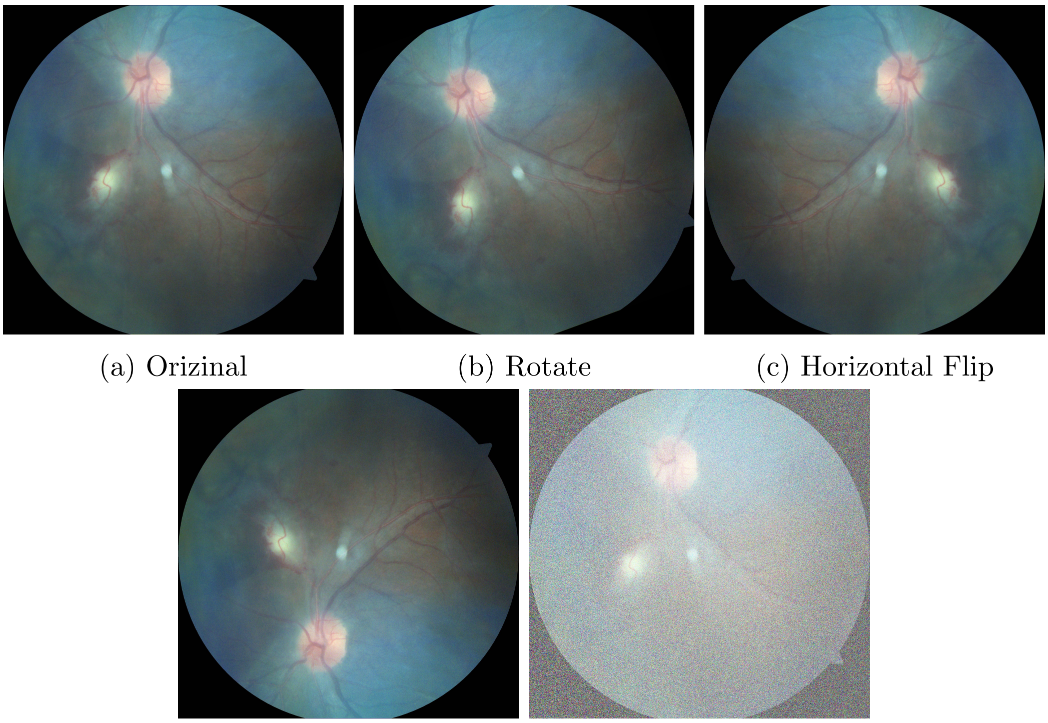
2. **Augmentation:** As was established in the previous literature study, one of the largest barriers to the advancement of research is dealing with small data sets. When given a small dataset to train on, a neural network model becomes overfitted and memorizes the data rather than the relationships. In order to be able to predict new data that wasn't accessible during training, a model must be able to generalize patterns from the training data. Thus, the primary objective of augmentation is to enhance the overall size of the dataset's photos.

    Resize Images: Resizing images in image datasets for deep learning is vital to maintain consistency, cut down on computing cost, increase model compatibility, and assure consistent and effective training. That's the reason photos were reduced from their original resolution of pixels to 224x224 pixels proportionately.

    Rotation: The photos were 20 degrees anticlockwise rotated. It makes the adjustments appear minimal.

    Fliping: Images are flipped both hozirontal nad vertical axis.

    Gausian Noise: In order to create Gaussian noise, it is usually done to sample random values from a Gaussian distribution and add them to an image's pixel values.



(a) Orizinal (b) Rotate (c) Horizontal Flip

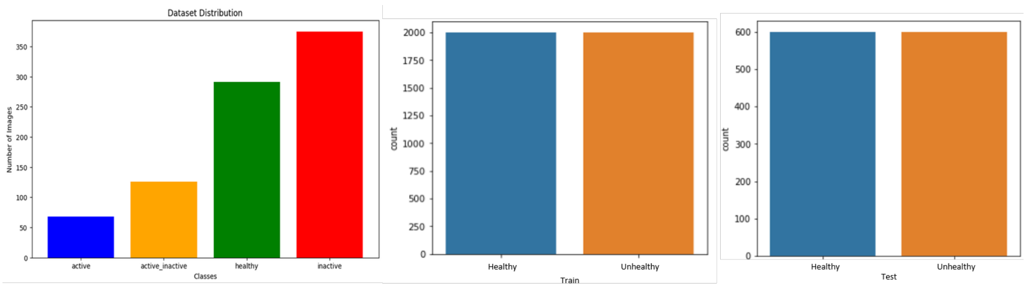
(d) Vertical Flip (e) Gausian Noisy

Figure 4.5: Before and after Augmentation

Augmentation will help to improve the diversity of training dataset, boosts model generalisation, and allows models to deal with variations in object orientation.

3. **Result after Augmentation:** We move multi-classification to binary-classification beacuse it leads to improved model performance and interpret-ability.

   When classes that are conceptually similar are combined, binary classification frequently produces a data set that is more balanced. By lowering the possibility of class imbalance problems in a multi-class scenario, this balance can have a positive effect on model training. In models where the majority class is predicted more often than the minority class, class imbalance can result in biased models.



(a) Before and After of Dataset

   Binary classification tasks require less complex models, resulting in faster training times and lower computational requirements. This simplicity makes the model more interpret-able and understandable. It simplifies model evaluation and interpretation, allowing for clearer interpretation of metrics like precision, recall, and F1 score, aiding in understanding model performance and identifying improvement areas.

4.4 Evaluation of Models

After finishing the preparation process detailed in Chapter 4 of that chapter, the datas et may now be put into CNN models to generate a relationship between the image features and the final interpretation. The data set was partitioned once more in a ratio of 70:30:10 for training, testing and validation respectively. Scikit-Learn's train test validation performed this process using the "Shuffling true" parameter.

4.4.1 VGG19:

1. A well-known architecture called VGG-19 is renowned for its efficiency in picture classification tasks.

2. Due to its simple design and simplicity, it acts as a great baseline for a variety of computer vision applications.

3. Transfer learning can be performed using pre-trained VGG-19 models, which saves time and resources as compared to training on big data sets.

4. Because of its depth and architecture, the VGG-19 is suited for a variety of image recognition tasks.

5. With the availability of pre-trained models and implementations, VGG-19 has broad community support, making project adoption easier.

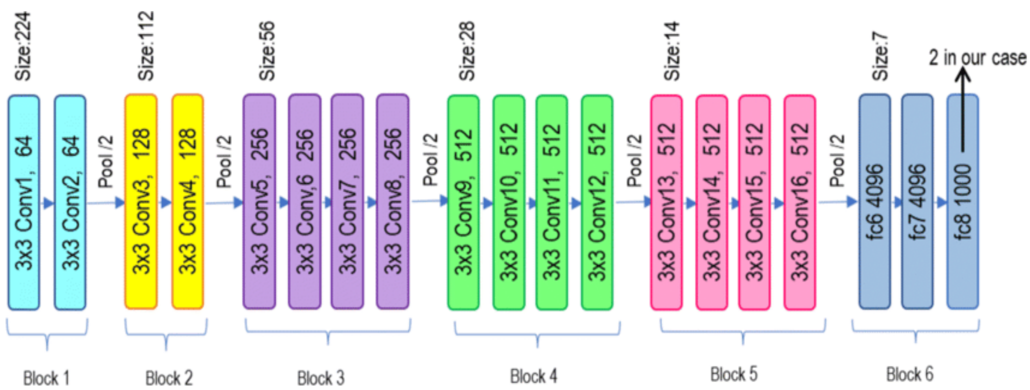


Figure 4.7: **VGG19[62]**

The formula of VGG19:
Convolutional Layer:

$$H_{\text{out}} = \left\lfloor \frac{H_{\text{in}} + 2P - F}{S} \right\rfloor + 1$$

$$W_{\text{out}} = \left\lfloor \frac{W_{\text{in}} + 2P - F}{S} \right\rfloor + 1$$

Max pooling Layer:

$$H_{\text{out}} = \left\lfloor \frac{H_{\text{in}} + F}{S} \right\rfloor + 1$$

$$W_{\text{out}} = \left\lfloor \frac{W_{\text{in}} + F}{S} \right\rfloor + 1$$

ReLU activation function:
ReLU(x)=max(0,x)

4.4.2 ResNet50:

In 2015 it was introduced by Kaiming He and his team. Later on it has became a very widely used architecture for different computer vision.
Residual block:

$$H_{\text{out}} = \left\lfloor \frac{H_{\text{in}} + 2P - F\text{conv1}}{S\text{conv1}} \right\rfloor + 1$$

$$W_{\text{out}} = \left\lfloor \frac{W_{\text{in}} + 2P - F\text{conv1}}{S\text{conv1}} \right\rfloor + 1$$

Pooling Layer:
Input=Hin*Win*Din
Output=1*1*Din

1. It has 50-layer deep neural network which helps to capture complex hierarchical features in data, making it ideal for high-level feature learning tasks.
2. It introduce a residual connection. It helps to bypass certain layer during training. This feature reduce the vanishing gradient problem.
3. As we know it is very deep neural network contain with 50 layer.
4. Its depth and remaining connections help it recognise objects and patterns in images with a high degree of precision.
5. It can pre-trained on large data set like ImageNet. Also can fine tune on smaller data sets for specific tasks and saving time and resources.
6. The ResNet-50 architecture exhibits adaptability and suitability for various computer vision applications, encompassing tasks such as picture classification, object recognition, and segmentation. Additionally, it offers reliable performance and substantial depth for a diverse array of applications.

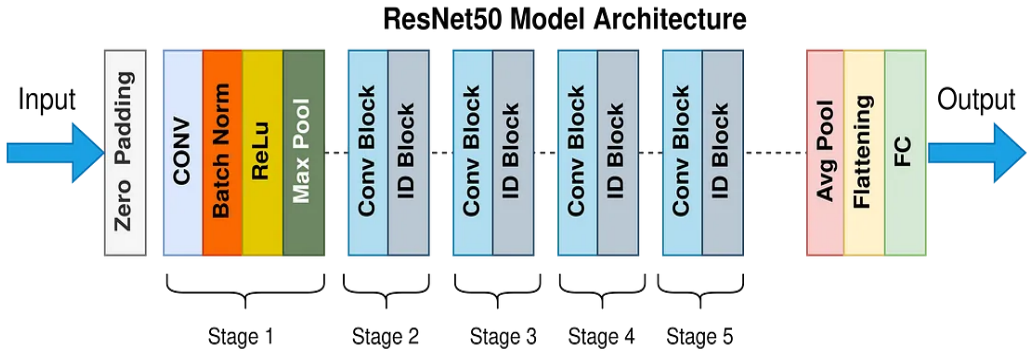


Figure 4.8: **ResNet-50**[60]

4.4.3 MobileNet:

It was introduced by Google and it is famous for its light weight and efficient design.
DepthWise Separable convolution:

$$H_{\text{out}} = \left\lfloor \frac{H_{\text{in}} + 2P\text{dw} - F\text{conv1}}{S\text{conv1}} \right\rfloor + 1$$

$$W_{\text{out}} = \left\lfloor \frac{W_{\text{in}} + 2P\text{dw} - F\text{conv1}}{S\text{conv1}} \right\rfloor + 1$$

Pointwise Convolution:

$$H_{\text{out}} = \left\lfloor \frac{H_{\text{in}} + 2P\text{dw} - 1}{} \right\rfloor + 1$$

$$W_{\text{out}} = \left\lfloor \frac{W_{\text{in}} + 2P\text{dw} - 1}{} \right\rfloor + 1$$

Global Average Pooling:

Input=H*W*D
Output=1*1*D

1. MobileNet models are extremely resource and memory efficient, making them excellent for deployment on mobile devices, embedded systems, and edge devices.

2. It reduce Complexity. So can run faster in larger data set and making them ideal for real-time application.

3. It has low inference latency which is crucial for applications where quick response times are required.

4. MobileNet is highly suited for mobile applications and settings with limited resources because it helps preserve battery life and lowers data usage.

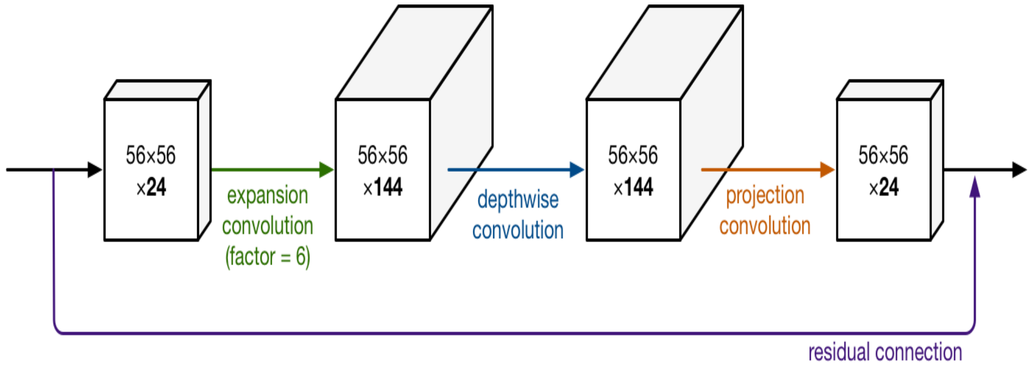


Figure 4.9: **MobileNet[59]**

4.4.4 VGG16:

VGG16 is a convolutional neural network (CNN) architecture with 16 layers, known for its simplicity and effectiveness in image classification tasks.

The formula of VGG16:

Convolutional Layer:

$$H_{\text{out}} = \left\lfloor \frac{H_{\text{in}} + 2P - F}{S} \right\rfloor + 1$$

$$W_{\text{out}} = \left\lfloor \frac{W_{\text{in}} + 2P - F}{S} \right\rfloor + 1$$

Max pooling Layer:

$$H_{\text{out}} = \left\lfloor \frac{H_{\text{in}} + F}{S} \right\rfloor + 1$$

$$W_{\text{out}} = \left\lfloor \frac{W_{\text{in}} + F}{S} \right\rfloor + 1$$

ReLU activation function:
ReLU(x)=max(0,x)

Resnet50:

$$\text{Input} = H_{\text{in}} * W_{\text{in}} * D_{\text{in}}$$

1. VGG16 has 16 layers that can learn lots of details in pictures, helping it understand and recognize different features.

2. It uses 3x3 filters for making its structure simple and easy to follow, helping us understand how it works.

3. VGG16 was trained on a big dataset called ImageNet, making it smart and ready to be used for recognizing things in new pictures.
4. Whether it's figuring out what's in a picture or finding objects, VGG16 can handle various jobs because of its balanced design.
5. VGG16 keeps things simple, making it easier for researchers and users to understand how it makes decisions in different situations.

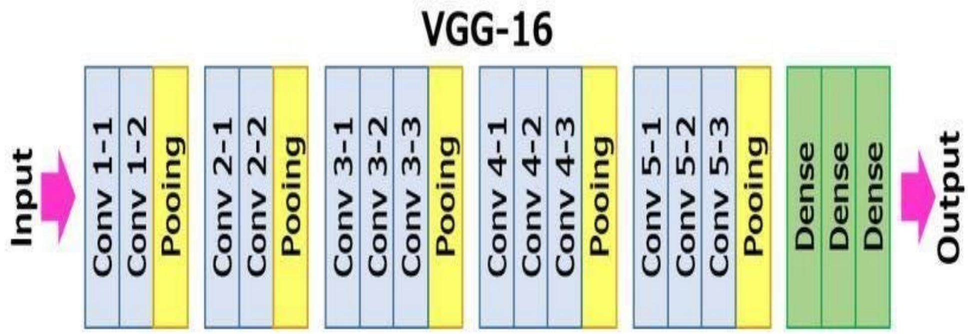


Figure 4.10: **VGG16[61]**

4.5 Ensemble Model

1. **Models**

    VGG16: Building convolutional layers are followed by fully connected layers, and this design is known for its simplicity. Hierarchical features are captured.

    ResNet50: Uses residual connections to lessen the issue of vanishing gradients. Capturing long-range dependencies, appropriate for deep networks.

    MobileNet: Using depth-wise separable convolutions, this efficient design is appropriate for environments with limited resources.

    VGG19: More in-depth and capable of capturing intricate patterns with more convolutional layers than VGG16.

2. **Input:** VGG16 receives the data. then, in that order, ResNet50, MobileNet, and VGG19. Every model has been modified to take in shape-based input (128, 128, 3).

3. **Architecture:** The outputs of the last layers of VGG16, ResNet50, MobileNet, and VGG19 are concatenated following individual training. For the final classification, a fresh set of dense layers is added to the concatenated output. The various features that each base model learns are combined in the ensemble model.

4. **Training Process:** Ensemble models improve overall predictive performance by utilizing the diversity of individual models. A wider variety of features and patterns can be captured by combining various architectures. When compared to individual models, it offers a more reliable and generalizable model. Reducing overfitting and enhancing the model's capacity to generalize to new data are two benefits of the ensemble approach. The ensemble model can adjust and fine-tune the combined features for improved performance by being trained over more epochs.

4.6 Proposed Model

1. **Convolutional 2D Layer:**

   - **1st Conv2D layer:** With a kernel size of (3, 3), the first convolutional layer, conv2d_3, generates 128 feature maps by processing the input images. After batch normalization, which normalizes the inputs of the layer, the ReLU activation function adds non-linearity to the model and improves training stability. 3,584 parameters are added by this layer to the overall model architecture.
   - **2nd Conv2D layer:** Conv2d_4, the second convolutional layer, creates 32 feature maps by using a (3, 3) kernel to further analyze the feature maps produced by the previous layer. ReLU activation and batch normalization are applied, just like in the layer above. There are 36,896 parameters in total for this layer.
   - **3rd Conv2D layer:** Using a (3, 3) kernel, the third convolutional layer, conv2d_5, completes the hierarchical feature extraction process and generates 64 feature maps. Once more, batch normalization and ReLU activation are used, adding 18,496 parameters to the model.
   - **4th Conv2D layer:** Lastly, 128 feature maps are produced by the fourth convolutional layer, conv2d_6, using a (3, 3) kernel to process the data. Applying batch normalization and ReLU activation, this layer increases the total number of parameters in the model by 73,856.

2. **Max Pooling Layer** After Conv2d Layer we used Max Pooling layer. In this model, the MaxPooling layers function to spatially down-sample the feature maps, lowering the dimensionality while keeping relevant data. These layers improve the model's capacity to identify important patterns and features by choosing the maximum values within local regions, which encourages efficient hierarchical representation learning.

3. **Batch Normalization:** Four batch normalization layers are present. We used the Batch Normalization layer following each convolution layer. In

essence, this layer normalizes activation, which speeds up and stabilizes our customized CNN model.

4. **Flatten Layer:** The model's Flatten layer flattens 2D feature maps into a 1D vector, facilitating data transfer from convolutional layers to fully connected layers for processing and classification.

5. **Dense Layer:** The model uses three dense layers for feature extraction and transformation, with the first providing a high-dimensional representation, the second providing a compact representation, and the final producing binary classification probabilities.

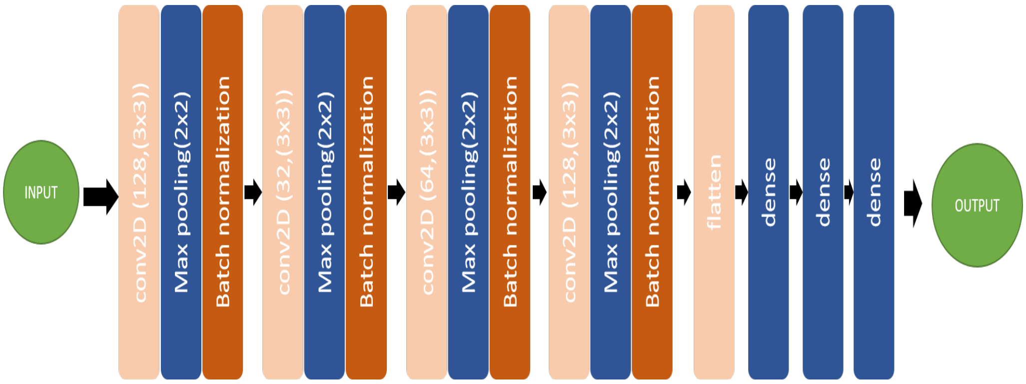


Figure 4.11: **Custom CNN Model**

Our customized CNN Model Structure is given which will make the structure more understandable.This model has a total of 1,330,722 parameters, where 1,330,018 are trainable parameters and 704 are non-trainable parameters.

Chapter 5

Result and Analysis

5.1 Performance Metrics

Parameters such as accuracy, recall, f1 score, precision, macro average, and weighted average were used to assess the model's performance. Some of the metrics' formulas are given as:
The formula for Precision is given by:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

The formula for Recall is given by:

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

The formula for F1 Score is given by:

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

The formula for accuracy is given by:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total Number of Predictions}}$$

5.1.1 Confusion Matrix

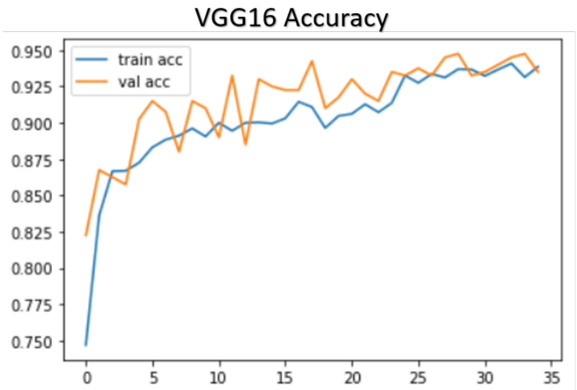
A confusion matrix is a table used to assess how well a categorization system performs.. A confusion matrix visualizes and summarizes the results of a classification algorithm. As a result, there are 4 possible results in a binary classification:
1. True Positive (TP): When the results match both expectations and reality, this is a true positive (TP).

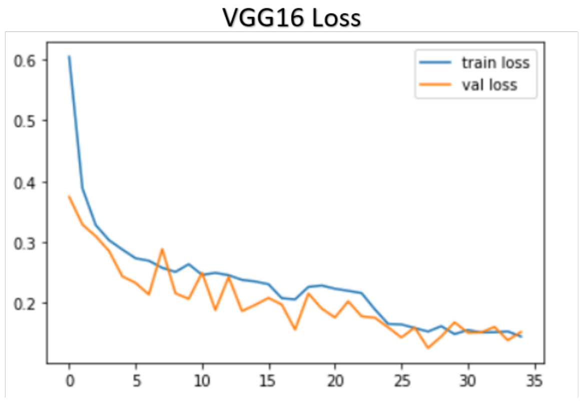
2. True Negative (TN): when the results are not what was anticipated or what happened as predicted.
3. False Positive (FP): Also referred to as a Type 1 error, this mistake type happens when a positive result is anticipated but the actual outcome is negative.
4. False Negative (FN): also known as Type 2 error, happens when a result is predicted to be negative but really turns out to be positive.

5.2 Model Evaluations

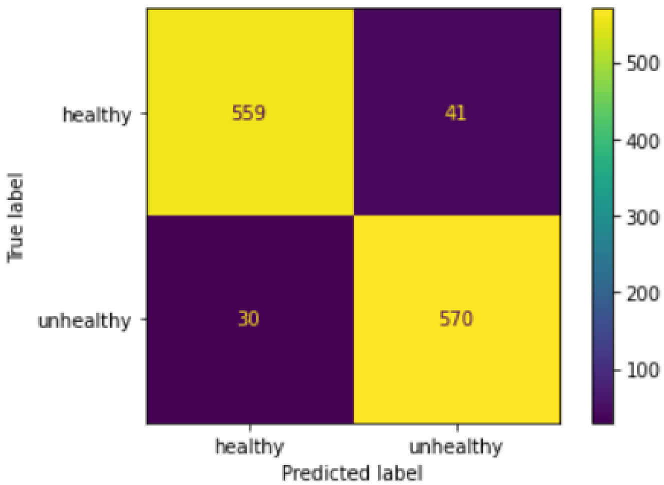
5.2.1 Performance Study on VGG16

With a beginning learning rate of 0.001, the VGG16 model was trained across 35 epochs, tracking metrics such as accuracy, precision, recall, and AUC, and optimizing weights. Over the course of the epochs, the model demonstrated progress in a number of performance indicators on the training and validation sets. Notable results include increased precision, recall, accuracy, and AUC. Validation loss decreased from 0.3747 to 0.1526, whereas validation accuracy increased from 0.8225 to 0.9350. Adaptively, the learning rate was adjusted throughout training: after the 23rd epoch, it was set to 0.0001, and after the 33rd epoch, it was set to 1e-05.



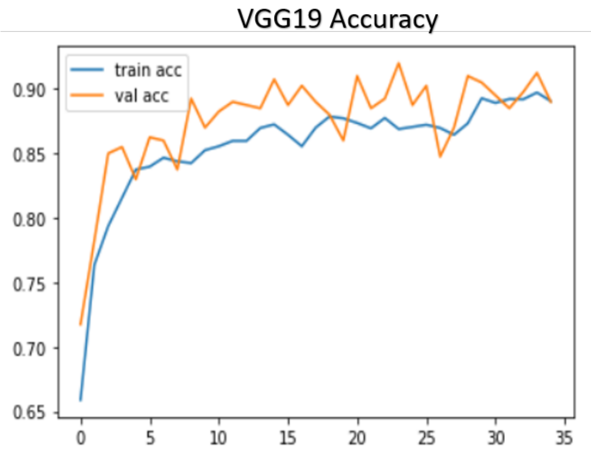


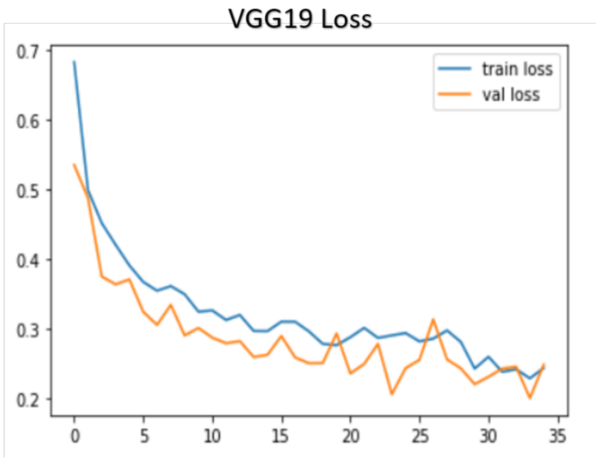
	precision	recall	f1-score	support
healthy	0.95	0.93	0.94	600
unhealthy	0.93	0.95	0.94	600
accuracy			0.94	1200
macro avg	0.94	0.94	0.94	1200
weighted avg	0.94	0.94	0.94	1200



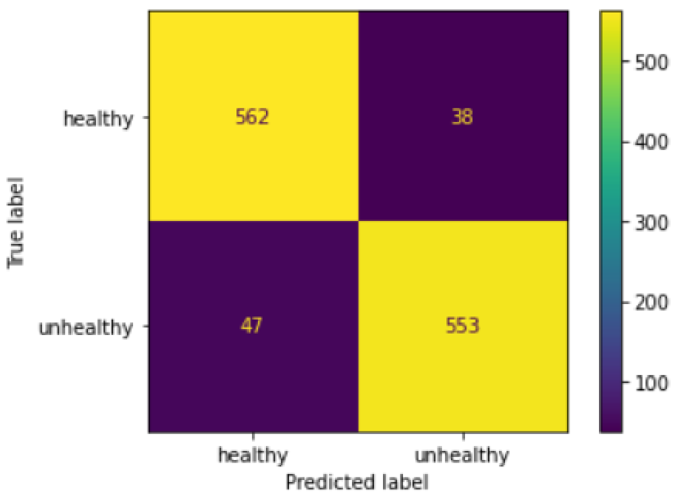
5.2.2 Performance Study on VGG19

35 epochs were used to train the VGG19 model. The model demonstrated an accuracy of 65.89% and a loss of 0.6833 in the first epoch. Accuracy, area under the curve (AUC), recall, precision, and both training and validation losses increased over the course of the next epochs. Among the noteworthy findings is the dynamic adjustment of the learning rate, which at epoch 29 drops to 0.0001. One popular method for maximizing model convergence is this adaptive learning rate. The validation set continuously showed improved metrics, indicating the improved capacity of the model to generalize to unknown data: accuracy, AUC, recall, and precision.



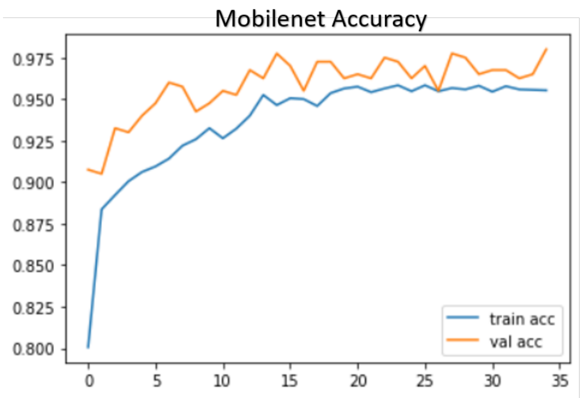


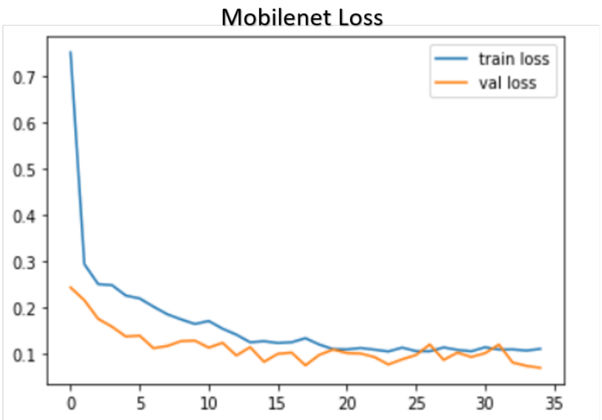
	precision	recall	f1-score	support
healthy	0.92	0.94	0.93	600
unhealthy	0.94	0.92	0.93	600
accuracy			0.93	1200
macro avg	0.93	0.93	0.93	1200
weighted avg	0.93	0.93	0.93	1200



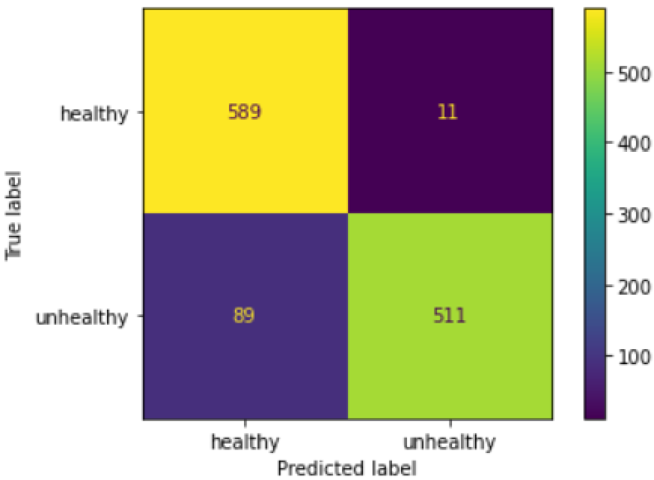
5.2.3 Performance Study on Mobilenet

Thirty-five epochs were dedicated to training the MobileNet model. Numerous performance metrics showed a steady improvement over these epochs. The model showed an accuracy of 80.06% and a loss of 0.7526 in the first epoch. Accuracy, area under the curve (AUC), recall, and precision all increased as training went on and the loss went down. One noteworthy finding is that at epoch 12, the learning rate dropped to 0.0001, suggesting the use of an adaptive learning rate strategy. This adjustment most likely made it possible for the model to converge successfully. Improved metrics like accuracy, AUC, recall, and precision were consistently shown in the validation set, indicating the model's potential for good generalization to untested data.



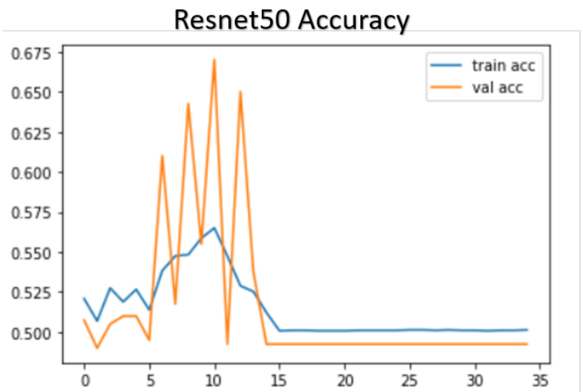


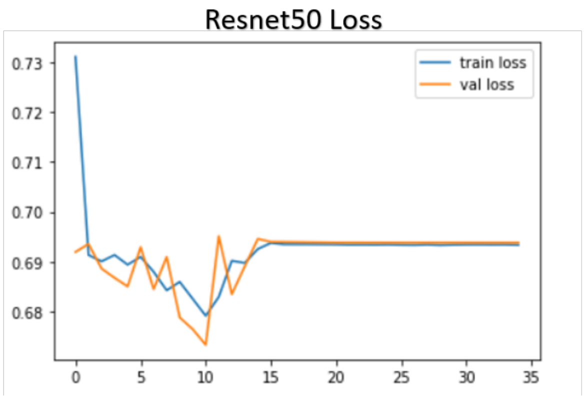
	precision	recall	f1-score	support
healthy	0.87	0.98	0.92	600
unhealthy	0.98	0.85	0.91	600
accuracy			0.92	1200
macro avg	0.92	0.92	0.92	1200
weighted avg	0.92	0.92	0.92	1200



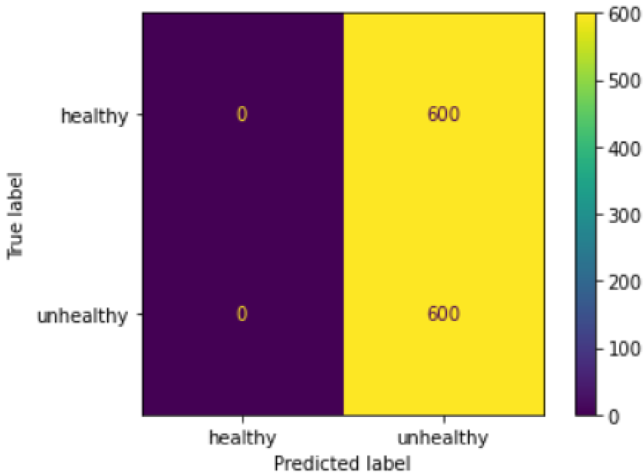
5.2.4 Performance Study on Resnet50

During the ResNet50 model's 35-epoch training process, the starting conditions demonstrated a 52.08% accuracy and a loss of 0.7311. As the model developed, the accuracy stayed around 50.14% and the loss progressively dropped to 0.6933. Starting at 0.001, the learning rate changed dynamically over the course of training, finally settling at a minimum of 1e-06. The validation metrics showed variations over the course of the epochs, with variations in the values of AUC, Recall, and Precision. The model's performance on the validation set plateaued despite learning rate reductions, suggesting that fine-tuning parameters may be necessary or that there may be difficulties in achieving further improvement.



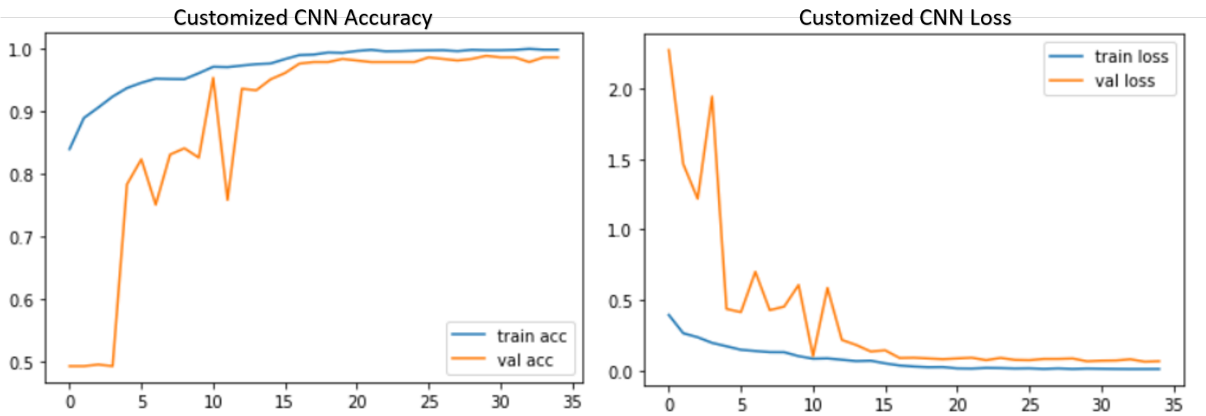


	precision	recall	f1-score	support
healthy	0.00	0.00	0.00	600
unhealthy	0.50	1.00	0.67	600
accuracy			0.50	1200
macro avg	0.25	0.50	0.33	1200
weighted avg	0.25	0.50	0.33	1200

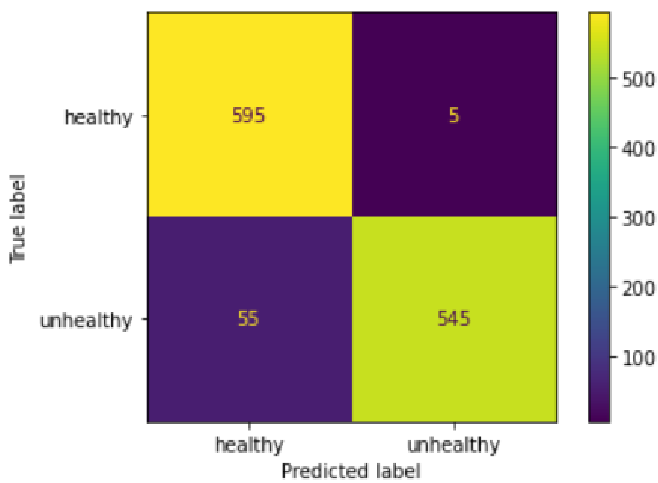


5.2.5 Performance Study on Custom CNN

Thirty-five epochs were dedicated to training the MobileNet model. Numerous performance metrics showed a steady improvement over these epochs. The model showed an accuracy of 80.06% and a loss of 0.7526 in the first epoch. Accuracy, area under the curve (AUC), recall, and precision all increased as training went on and the loss went down. One noteworthy finding is that at epoch 12, the learning rate dropped to 0.0001, suggesting the use of an adaptive learning rate strategy. This adjustment most likely made it possible for the model to converge successfully. Improved metrics like accuracy, AUC, recall, and precision were consistently shown in the validation set, indicating the model's potential for good generalization to untested data.

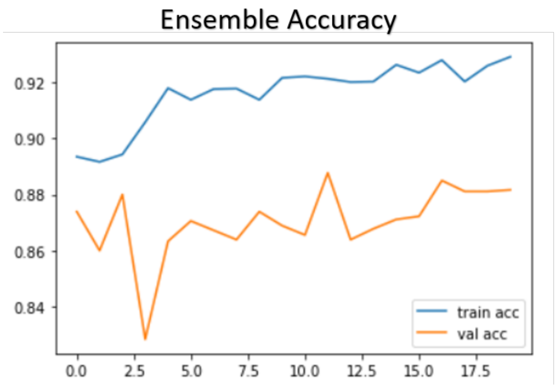


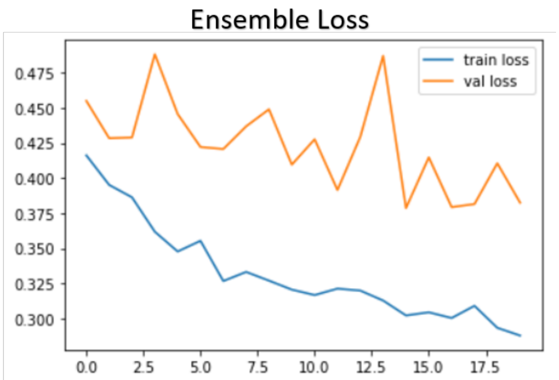
	precision	recall	f1-score	support
healthy	0.92	0.99	0.95	600
unhealthy	0.99	0.91	0.95	600
accuracy			0.95	1200
macro avg	0.95	0.95	0.95	1200
weighted avg	0.95	0.95	0.95	1200



5.2.6 Performance Study on Ensemble

Throughout the course of the model's 20 training epochs, a few significant findings can be noted. With a robust performance, the first epoch began with a training loss of 0.4162 and an accuracy of 89.35%. As training went on, the model showed steady progress, peaking at 92.92% accuracy by the 20th epoch and with a minimum loss of 0.2878.
The results for the validation set also showed a positive trend, with the accuracy staying relatively high at about 88.17% and the loss falling from 0.4550 to 0.3824. This suggests that the model's ability to generalize to new data is strong. The training and validation metrics gradually converged, indicating stability in the training process. The model's high accuracy on the validation set and learning to minimize errors suggest its effectiveness in producing precise predictions on fresh, untested data.





5.2.7 Results

Model	Accuracy	Loss
Customized CNN	95%	0.11
VGG16	94%	0.19
VGG19	93%	0.24
MobileNet	92%	0.13
Ensemble Model	92%	0.28
Resnet50	50%	0.70

Table 5.1: Accuracy and Loss

5.2.8 Discussions

A comparison of the performance of different models offers valuable information about how well they work for the task at hand. The Customized CNN performed better than the majority of the other models, exhibiting a low loss of 0.11 and a noteworthy accuracy of 95%. This implies that the architecture created especially for the task at hand showed exceptional learning capabilities and successfully identified the underlying patterns in the data.

VGG16 outperformed the other models with an impressive accuracy of 94%; VGG19 and MobileNet came in second and third, respectively, with accuracies of 93% and 92%. However, the accuracy rates of the Ensemble Model and Resnet50 were lower, at 92% and 50%, respectively, suggesting that these models might have trouble understanding the complexities of the dataset.

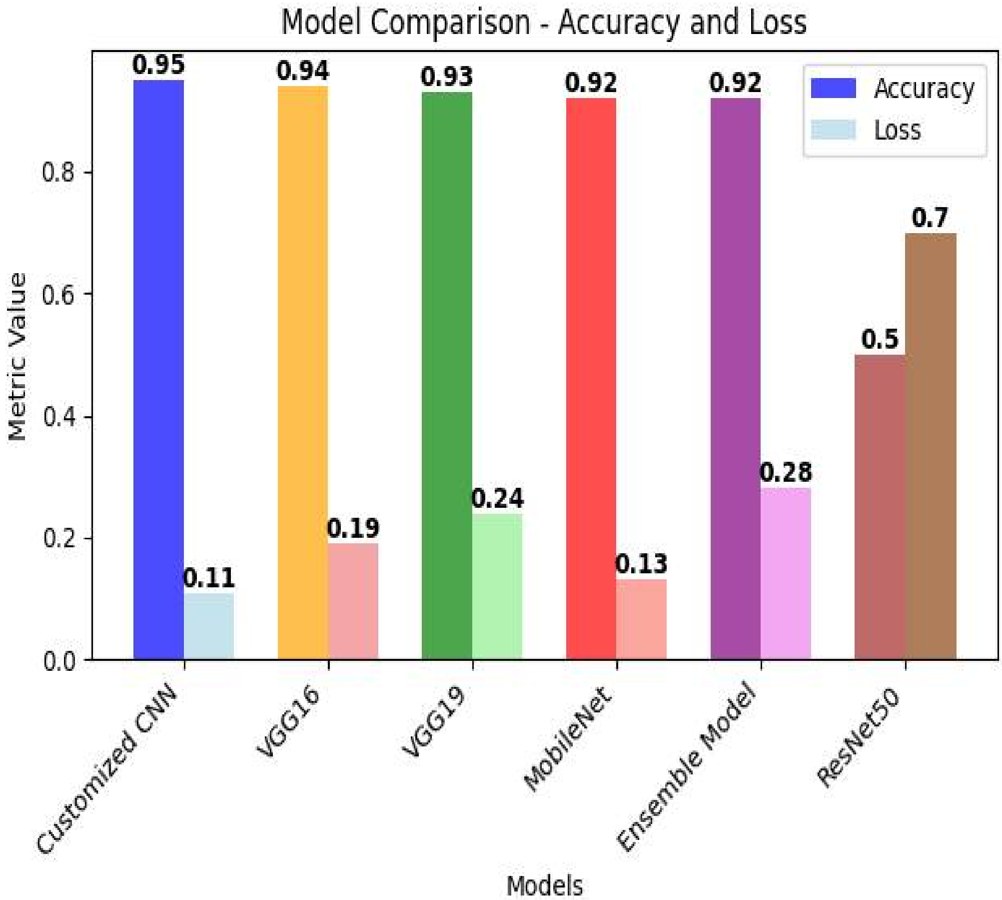


Figure 5.1: **Accuracy and Loss of Models**

The Customized CNN's superior accuracy when compared to all other models highlights how well the custom architecture extracts pertinent features and generates precise predictions. Even though VGG16 did well as well, the Customized CNN shows how important it is to modify the model architecture to fit the unique features of the dataset in order to achieve better performance. This emphasizes how important careful planning, design, and customization are to getting the best results possible for image classification tasks.

5.2.9 Comparison with Recent work

This thesis article compares the accuracy of ML-CNN and ResNets implementations on fundus image datasets. E.AbdelMaksoud et al.[9] achieved 94.3% accuracy for detecting ocular diseases from color fundus images, while Parra et al.[4] achieved 93.3% accuracy for identifying ocular toxoplasmosis. The CNN model achieved the highest accuracy of 95%.D.Abeyrathna[22] presents a segmentation network fine-tuning technique that employs feature clustering of predictions and labeled training instances to improve network performance using Mask R-CNN and achieves 66% accuracy.

Approaches	**Models**	**Accuracy**
Customized CNN	CNN%	95%
E.Abdel Maksoud et.al[9]	ML-CNN	94.3%
Parra et.al[2]	ResNets	93%
E.Abdel Maksoud	Mobilenet	92%
D.Abeyrathna et.al[22]	Mask R-CNN	66%

Table 5.2: Work Comparison

Chapter 6

Conclusion

6.1 Conclusion

Toxoplasma gondii is a protozoan parasite which leads to ocular toxoplasmosis, a form of eye infection. Ocular toxoplasmosis is a bacterial or viral illness of the eye. Therefore, if this eye condition is not treated in a timely manner, it may be fatal. We must first identify this illness if we want to treat it quickly. Serologic test reports are widely used in the diagnosis of this illness. Utilizing test data alone can make it extremely challenging to determine whether or not the patient actually has ocular toxoplasmosis disease. It can occur that despite the patient not having ocular toxoplasmosis, the doctor eventually recommends medication for the condition after noticing inaccuracies in the test results. Conversely, the patient might have ocular toxoplasmosis, but due to inaccuracies in test findings, the doctor does not treat the patient as a patient. As a result of receiving insufficient treatment, the patient will experience numerous physical issues. We used a customized CNN model and several pre-trained models on images that provided good accuracy to detect ocular toxoplasmosis with minimal error. Our ability to detect ocular toxoplasmosis will be improved if we increase the number of epochs. Our goal is to be able to diagnose this illness with a high degree of precision and fewer mistakes.

6.2 Limitation

Because our dataset is an image dataset, it takes a long time to run any CNN model, and the probability of achieving acceptable accuracy decreases with the number of epochs. Every model's accuracy graph shows a growing graph, as can be seen by looking at it. It follows that an increase of epochs will ultimately result in an improvement in accuracy.

6.3 Future Work Plan

In order to improve accuracy and F1 score later on, we wish to increase the number of epochs and patience for all models.

Bibliography

[1] Park, Y.-H., & Nam, H.-W. and " Clinical features and treatment of ocular toxoplasmosis," In *T*he Korean Journal of Parasitology, vol.51(4), pp.393-399,https://doi.org/10.3347/kjp.2013.51.4.393.

[2] Parra, R., Ojeda, V., Vázquez Noguera, J. L., García Torres, M., Mello Román, J. C., Villalba, C., Facon, J., Divina, F., Cardozo, O., Castillo, V. E., and Castro Matto, I, "Automatic diagnosis of ocular toxoplasmosis from fundus images with residual neural networks," In *(Studies in Health Technology and Informatics.* IOS Press,2011).

[3] Tong, Y., Lu, W., Yu, Y. and Shen, Y,"Application of machine learning in ophthalmic imaging modalities. Eye and Vision," In *(*Institute of Ophthalmology, University College London, London,2020),https://doi.org/10.1186/s40662-020-00183-6.

[4] B. L. Shoop, A. H. Sayles, and D. M. Litynski, "New devices for optoelectronics: smart pixels," in *Handbook of Fiber Optic Data Communications*, C. DeCusatis, D. Clement, E. Maass, and R. Lasky, eds. (Academic, 1997), pp. 705–758.

[5] R. E. Kalman,"Algebraic aspects of the generalized inverse of a rectangular matrix," in *Proceedings of Advanced Seminar on Generalized Inverse and Applications*, M. Z. Nashed, ed. (Academic, 1976), pp. 111–124.

[6] R. Craig and B. Gignac, "High-power 980-nm pump lasers," in *Optical Fiber Communication Conference*, Vol. 2 of 1996 OSA Technical Digest Series (Optical Society of America, 1996), paper ThG1.

[7] D. Steup and J. Weinzierl, "Resonant THz-meshes," presented at the Fourth International Workshop on THz Electronics, Erlangen-Tennenlohe, Germany, 5–6 Sept. 1996.

[8] khan and M.S,"Deep Learning for Ocular Disease Recognition : an Inner-Class balance," in *(*Computational Intelligence and Neuroscience,2022),pp. 1–12.

[9] O. Ouda ,E. Abdel Maksoud, A. a. A. El-Aziz and M. Elmogy, " Multiple ocular disease diagnosis using FundUS images based on Multi-Label Deep Learning Classification," in *(*Electronics, June.2022),vol .11,pp. 1966, doi: 10.3390/electronics11131966.

[10] R. Parra et. al ,"A Trust-Based Methodology to Evaluate Deep Learning Models for Automatic Diagnosis of Ocular Toxoplasmosis from Fundus

Images," in (Diagonistic,Oct.2021), pp. 1951,doi: 10.3390/diagnostics11111951.

[11] Hasanreisoglu M.," Ocular Toxoplasmosis Lesion Detection on Fundus Photograph using a Deep Learning Model," in (Arvojournals,June.2020),vol. 61.

[12] Y. Elloumi, M. Akil, and H. Boudegga, "Ocular Diseases Diagnosis in Fundus Images using a Deep Learning: Approaches, tools and Performance evaluation," In (Researchgate,2019).

[13] Midena, E., et al. "Ultra-Wide-Field Fundus Photography Compared to Ophthalmoscopy in Diagnosing and Classifying Major Retinal Diseases." Scientific Reports, vol. 12, no. 1, 11 Nov. 2022, https://doi.org/10.1038/s41598-022-23170-4. Accessed 17 Jan. 2023.

[14] Jamal A, Hazim Alkawaz M, Rehman A, Saba T. Retinal imaging analysis based on vessel detection. Microsc Res Tech. 2017 Jul;80(7):799-811. doi: 10.1002/jemt.22867. Epub 2017 Mar 13. PMID: 28294460.

[15] Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. In International Conference on Learning Representations (ICLR).

[16] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).

[17] Szegedy, C., et al. (2015). Going Deeper with Convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).

[18] Adithi D. Chakravarthy, Dilanga Abeyrathna, Mahadevan Subramaniam, Parvathi Chundi, Venkataramana Gadhamshetty, "Semantic Image Segmentation Using Scant Pixel Annotations", Machine Learning and Knowledge Extraction, vol.4, no.3, pp.621, 2022.

[19] Shrivastava and A. "Deep Learning Based Ocular Disease Classification using Retinal Fundus Images," In (arvo.journals,August 2021).

[20] Anneke Annassia and Putri Siswadi.,"Computer-aided-diagnosis for ocular abnormalities from a single color fundus photography with deep learning. Signal and Image Processing", In (University of Burgundy,2023).

[21] Chakravarthy, A. D., Abeyrathna, D., Subramaniam, M., Chundi, P., Halim, M. S., Hasanreisoglu, M., Nguyen and Q. D. " An Approach Towards Automatic Detection of Toxoplasmosis using Fundus Images," In IEEE 20th International Conference on Bioinformatics and Bioengineering (October28,2019),https://doi.org/10.1109/BIBE.2019.00134.

[22] D.Abeyrathna et al., "Directed Fine Tuning Using Feature Clustering for Instance Segmentation of Toxoplasmosis Fundus Images," In *I*EEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), Cincinnati, OH, USA, 2020, pp. 767-772, doi:10.1109/BIBE50027.2020.00130.

[23] Alam, S. S., Shuvo, S. B., Ali, S. N., Ahmed, F., and Chakma, A, " Benchmarking Deep Learning Frameworks for Automated Diagnosis of Ocular Toxoplasmosis: A Comprehensive Approach to Classification and Segmentation," In *p*rint arXiv:2305.10975,20,(May 23,2018),https://doi.org/10.48550/arXiv.2305.10975.

[24] J. G. Garweg, J. D. F. De Groot-Mijnes, and J. G. Montoya, "Diagnostic approach to ocular toxoplasmosis," Ocular Immunology and Inflammation, vol. 19, no. 4, pp. 255–261, Jul. 2011, doi: 10.3109/09273948.2011.595872.

[25] Gómez-Marín, J. E., Muñoz-Ortiz, J., Mejía-Oquendo, M., Arteaga-Rivera, J. Y., Rivera-Valdivia, N., Bohorquez-Granados, M. C., Velasco-Velásquez, S., Castaño-De-La-Torre, G., Acosta-Dávila, J. A., Garcia-López, L. L., Torres-Morales, E., Vargas, M., Valencia, J. D., Celis-Giraldo, D., De-La-Torre, A, "High Frequency of Ocular Toxoplasmosis in Quindío, Colombia and Risk Factors Related to the Infection,"(April 1,2021),https://doi.org/10.1016/j.heliyon.2021.e06659.

[26] Delair, E., Latkany, P., Noble, A. G., Rabiah, P., McLeod, R., & Brézin and A,"Clinical manifestations of ocular toxoplasmosis. Ocular Immunology and Inflammation," In *v*olume.19, pp.91-102,https://doi.org/10.3109/09273948.2011.564068.

[27] Kalogeropoulos, D., Sakkas, H., Mohammed, B., Vartholomatos, G., Malamos, K., Sreekantam, S., Kanavaros, P., Kalogeropoulos, C,"Ocular Toxoplasmosis: A Review of the Current Diagnostic and Therapeutic Approaches. International Ophthalmology," In *(*Researchgate,2021),https://doi.org/10.1007/s10792-021-01994-92.

[28] P. Zhou et al., "Toxoplasma gondii infection in humans in China," In *(*Parasites and Vectors,2011),https://doi.org/10.1186/1756-3305-4-165.

[29] T. Masters, *Practical Neural Network Recipes in C++* (Academic, 1993).

[30] B. L. Shoop, A. H. Sayles, and D. M. Litynski, "New devices for optoelectronics: smart pixels," in *Handbook of Fiber Optic Data Communications*, C. DeCusatis, D. Clement, E. Maass, and R. Lasky, eds. (Academic, 1997), pp. 705–758.

[31] R. E. Kalman,"Algebraic aspects of the generalized inverse of a rectangular matrix," in *Proceedings of Advanced Seminar on Generalized Inverse and Applications*, M. Z. Nashed, ed. (Academic, 1976), pp. 111–124.

[32] A. Howard, M. Zhu, B. Chen, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," In *(*Researchgate,2017).

[33] He, K., Gkioxari, G., Dollár, P. and Girshick, R.(2017). Mask R-CNN. In arXiv [cs.CV]. http://arxiv.org/abs/1703.06870.

[34] N.d.). Machinelearningmastery.com. Retrieved September 18, 2023, from https://machinelearningmastery.com/how-to-train-an-object-detection-model-with-keras.

[35] K. O'Shea, "An introduction to convolutional neural networks," arXiv.org, Nov. 26, 2015. https://arxiv.org/abs/1511.08458

[36] R. Yamashita, M. Nishio, R. K. Gian, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," Insights Into Imaging, vol. 9, no. 4, pp. 611–629, Jun. 2018, doi: 10.1007/s13244-018-0639-9

[37] D. P. Kingma, "Adam: A method for stochastic optimization," arXiv.org, Dec. 22, 2014. https://arxiv.org/abs/1412.6980

[38] "A brief introduction to OpenCV," in *I*EEE Conference Publication — IEEE Xplore, May 01, 2012. https://ieeexplore.ieee.org/document/6240859

[39] Gonçalves, T., Rio-Torto, I., Teixeira, L. F., Cardoso, J. S. (2022). "A survey on attention mechanisms for medical applications: are we moving towards better algorithms?"" In *a*rXiv [cs.CV]. http://arxiv.org/abs/2204.12406

[40] Kaselimi, M., Voulodimos, A., Daskalopoulos, I., Doulamis, N., Doulamis, A. (2023). A vision transformer model for convolution-free multilabel classification of satellite imagery in deforestation monitoring. *I*EEE Transactions on Neural Networks and Learning Systems, 34(7), 3299–3307. https://doi.org/10.1109/tnnls.2022.3144791

[41] D. Wang et al., "Advancing Plain Vision Transformer Toward Remote Sensing Foundation Model," in *I*EEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1-15, 2023, Art no. 5607315, doi: 10.1109/TGRS.2022.3222818.

[42] X. Meng, Y. Yang, L. Wang, T. Wang, R. Li and C. Zhang, "Class-Guided Swin Transformer for Semantic Segmentation of Remote Sensing Imagery," in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 6517505, doi: 10.1109/LGRS.2022.3215200.

[43] G. Deng, Z. Wu, M. Xu, C. Wang, Z. Wang and Z. Lu, "Crisscross-Global Vision Transformers Model for Very High Resolution Aerial Image Semantic Segmentation," in IEEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1-19, 2023, Art no. 4404019, doi: 10.1109/TGRS.2023.3276172.

[44] Kalogeropoulos D, Sakkas H, Mohammed B, Vartholomatos G, Malamos K, Sreekantam S, Kanavaros P, Kalogeropoulos C. Ocular toxoplasmosis: a review of the current diagnostic and therapeutic approaches. Int Ophthalmol. 2022 Jan;42(1):295-321. doi: 10.1007/s10792-021-01994-9. Epub 2021 Aug 9. PMID: 34370174; PMCID: PMC8351587.

[45] J. Zou, W. He and H. Zhang, "LESSFormer: Local-Enhanced Spectral-Spatial Transformer for Hyperspectral Image Classification," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-16, 2022, Art no. 5535416, doi: 10.1109/TGRS.2022.3196771.

[46] Zhang, K., Liu, X., Chen, Y. (2019, December 20-22). Research on Semantic Segmentation of Portraits Based on Improved Deeplabv3 +. *I*OP Conference Series: Materials Science and Engineering, 806. 10.1088/1757-899X/806/1/012057

[47] X. Chen, C. Qiu, W. Guo, A. Yu, X. Tong and M. Schmitt, "Multiscale Feature Learning by Transformer for Building Extraction From Satellite Images," in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 2503605, doi: 10.1109/LGRS.2022.3142279

[48] P. Lv, W. Wu, Y. Zhong, F. Du and L. Zhang, "SCViT: A Spatial-Channel Feature Preserving Vision Transformer for Remote Sensing Image Scene Classification," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-12, 2022, Art no. 4409512, doi: 10.1109/TGRS.2022.3157671.

[49] A. Shafique, S. T. Seydi, T. Alipour-Fard, G. Cao and D. Yang, "SSViT-HCD: A Spatial–Spectral Convolutional Vision Transformer for Hyperspectral Change Detection," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 6487-6504, 2023, doi: 10.1109/JSTARS.2023.3251646.

[50] Z. Liu et al., Swin Transformer: Hierarchical vision transformer using shifted windows, https://arxiv.org/abs/2103.14030 (accessed Jan. 5, 2024).

[51] Iyer, P. G., Ashkenazy, N., Liu, J., Laura, D., Marquez, M. A., Albini, T. A. (2023, August 16). Monitoring Delayed Toxoplasmosis-Related Branch Retinal Artery Occlusion Using Widefield en face Optical Coherence Tomography and Multimodal Imaging. Case Reports in Ophthalmology. https://doi.org/10.1159/000528787

[52] Btd. (2023, November 23). Explainable AI (XAI): Generating counterfactual explanations for interpretable machine learning. Medium. https://baotramduong.medium.com/explainable-ai-generating-counterfactual-explanations-for-interpretable-machine-learning-735ee56a3a00

[53] Vermeire, T., Brughmans, D., Goethals, S., De Oliveira, R. M. B., Martens, D. (2022). Explainable image classification with evidence counterfactual. Pattern Analysis and Applications, 25(2), 315–335. https://doi.org/10.1007/s10044-021-01055

[54] Inuwa,M.(2023,August 18). *S*WIN Transformers — Modern Computer Vision Tasks. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2023/08/swin-transformers-modern-computer-vision-tasks/h-understanding-swin-transformers

[55] Inuwa,M.(2023,June 14). Introduction to Vision Transformers (VIT). Analytics Vidhya. https://www.analyticsvidhya.com/blog/2023/05/introduction-to-vision-transformers-vit/h-how-do-vision-transformers-work

[56] H.Zhu B. Chen, and C. Yang, “Understanding why VIT trains badly on small datasets: An intuitive perspective,” arXiv.org, https://arxiv.org/abs/2302.03751 (accessed Jan. 7, 2024).

[57] R. Parra et al., “Ocular toxoplasmosis fundus images dataset,” Zenodo, https://zenodo.org/record/4479724 (accessed Jan. 9, 2024).

[58] C.Marie and J. William A. Petri, “Toxoplasmosis - infectious diseases,” MSD Manual Professional Edition, https://www.msdmanuals.com/en-jp/professional/infectious-diseases/extraintestinal-protozoa/toxoplasmosis (accessed Jan. 9, 2024).

[59] M.Hollemans Mobilenet version 2, https://machinethink.net/blog/mobilenet-v2/ (accessed Jan. 9, 2024).

[60] S.Mukheerje“The annotated resnet-50,” Medium, https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758 (accessed Jan. 9, 2024).

[61] G. Learning, “Everything you need to know about VGG16,” Medium, https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918 (accessed Jan. 9, 2024).

[62] VGG-19 has 16 convolution layers grouped into..., https://www.researchgate.net/figure/VGG-19-Architecture-39-VGG-19-has-16-convolution-layers-grouped-into-5-blocks-After$_fig5_359771670(accessedJan.9, 2024)$.

[63] D. Kalita “Basics of CNN in Deep Learning,” Analytics Vidhya, https://www.analyticsvidhya.com/blog/2022/03/basics-of-cnn-in-deep-learning/ (accessed Jan. 9, 2024).

[64] J. brownlee “How to choose an activation function for deep learning,” MachineLearningMastery.com, 21-Jan-2021. [Online]. Available: https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/. [Accessed: 09-Jan-2024]