Optical Coherence Tomography (OCT) Images for Eye Diseases Diagnosis

A Project Report Submitted to Faculty of Engineering in Partial Fulfillment of the Requirements for the Degree of M.Sc.(Engg.) in

Information and Communication Technology(ICT)

Submitted By: 22309017

Comilla University, Cumilla.

Date of Submission: 21 July 2025

Letter of Transmittal

21th July, 2025

To

Pintu Chandra Paul

Assistant Professor

Dept. of ICT

Comilla University

Subject: Submission of Project Report *on* “**Optical Coherence Tomography (OCT) Images for Eye Diseases Diagnosis”.**

Dear Sir,

It’s my pleasure to submit my project report on **“Optical Coherence Tomography (OCT) Images for Eye Diseases Diagnosis”** which was assigned to me as a project fulfillment of ICT Department of Comilla University.

While making this paper I came across much pleasant experience. Which will undoubtedly benefit of my in the years ahead. This paper attempts to describe my learning's and experienced gained about the use of many Electronics. Despite the several constraints, I gave my all efforts to make this paper meaning one.

I have tried sincerely to comprehend and my knowledge in writing this paper. I will be grateful and oblige if you accept my project report and evaluate it with your sagacious judgment.

Yours Sincerely,

Arshadul Islam Sourav

ID:22309017

Session: 2022-23

Dept. Of ICT, Comilla University.

DECLARATION

"I thus certify that this submission is entirely original work of mine, in accordance with the standards and instructions provided by my supervisor, sir, and that, to the best of my knowledge and belief, it does not contain any previously written material that has been approved for the award of any other degree or diploma from the university or other higher education institution, with the exception of instances in which appropriate credit has been provided within the text."

....

Arshadul Islam Sourav

ID-22309017, Session-(2022-23)

Dept of ICT, Comilla University

Date: 21th july 2025

Certificate of Acceptance

This certifies that the undertaking report entitled **"Optical Coherence Tomography (OCT) Images for Eye Diseases Diagnosis"** by me has been carried out as a partial fulfillment of requirement for the Final Year project. The dissertation has been carried out under my guidance and is a document of the authentic work completed successfully. Their performance has been satisfactory during this project period.

I wish every success in life.

Supervisor:

………………………………

Pintu Chandra Paul

Assistant Professor

Dept. of Information and Communication Technology

Comilla University, Cumilla.

ABSTRACT

As a component of the nervous system, the human eye is an organ that responds to visible light.One of the most crucial aspects of human existence is eye visibility, which enables us to observe our immediate surroundings and gather data thanks to the retina's capacity to visualize. But these days, Retinal damage is one of the most prevalent eye conditions, along with CNV, DRUSEN, AMD, and DME. These conditions can seriously harm our eyes and lead to vision loss and blindness. Optical Coherence Internal features of biological tissues may be imaged non-invasively using a type of contemporary scanning instrument called tomography. By monitoring the optical reflections of biological tissues, optical coherence tomography is a type of contemporary scanning technology that can see interior structures non-invasively in cross-section. OCT retinal pictures will assist the medical field or ophthalmologists in gaining a clear understanding of the condition or damage to the rear portion of the retina, macula, and optic nerve. My study's primary goal is to suggest a new deep learning-based classification model that uses sequential CNN architecture and transfer learning to automatically categories various retinal disorders from retinal pictures captured by an optical coherence tomography (OCT) instrument. I suggest a deep CNN architecture, and I compare the results with pre-trained models like VGG-16 and sequential CNN architectures like VGG-16 and VGG-19. where the results of my suggested architecture are 85%.

ACKNOWLEDGEMENT

First and foremost, we would like to sincerely thank Allah Ta'ala for His glorious grace, which enabled us to successfully finish the final year project.

We sincerely thank and express our sincere gratitude to Pintu Chandra Paul, an assistant professor in the ICT department of Comilla University in Cumilla. My supervisor's extensive expertise and intense interest in this area are beneficial to the project's completion. It has been made possible by his unending tolerance, academic direction, unceasing support, active and continuous supervision, helpful critique, insightful counsel, and his reading of several subpar manuscripts and their revision at every turn.

We would like to sincerely thank Professor Dr. Md. Tofael Ahmed, Head of the Department of ICT, as well as the other academic members and personnel of Comilla University's ICT department, for their helpful assistance in completing our project. We would like to express our gratitude to my Comilla University classmates who participated in this conversation as part of their coursework. Additionally, I don't want to pass up the chance to thank Google and Wiki for their contributions. These resources were really helpful to me as I worked on this project. Lastly, we must respectfully thank our parents for their unwavering patience and support.

.

Thanking you,

Arshadul Islam Sourav

Session: 2022-23

Dept. Of ICT, Comilla University

Acronyms & Abbreviations

OCT	Optical Coherence Tomography
CNN	Convolutional Neural Network
CNV	Choroidal Neovasculization
DME	Diabetic Macular Edema
VGG	Visual Geometry Group
AMD	Advanced Micro Devices
GPU	Graphics Processing Unit
ReLU	Rectified Linear Unit
LRN	Local Response Normalization
SGD	Stochastic Gradient Descent
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
ADAM	Adaptive Moment Estimation
ResNET	Resedual Neural Network
ILSVRC	ImageNet Large Scal Visual Recognition Challenge
KNN	K-nearest Neighbor
GBM	Glioblastoma Multiforme
LGBM	Light Gradient Boosting Machine

List of Figures

Table of Contents

CHAPTER 1

INTRODUCTION

The human eye is an organ of the body, part of the nervous system that interacts with visible light and allows us to use new information for a different of purposes, including seeing environments of our surroundings area, maintaining balance, and maintaining our circadian tone [1]. Our vision is shaped and made possible by several components of the brain and eye. The key components of our eye that enable us to transform light and electricity into pictures are the lens, retina, and optic nerve.[2] Retinal damage is one of the most common eye diseases. The most common cause of retinal detachment is age, minor injury or injury. Retinal detachment can damage vital tissue. They can affect our vision and some of them can be so severe that we may blind [3]. For retinal damages one could notice various visual effects such as –

- Blurry spots in one's central vision.
- Defects in side vision.
- Also can result in vision loss and eventual blindness.
- General distortion of images such as straight lines appearing wavy To protect it, we must take care of it. The primary reasons of retinal detachment are the lives of those who suffer from eye conditions such CNV, DRUSEN, AMD, and DME. Optical Coherence Tomography (OCT) retinal pictures were classified using deep learning based on comorbidities since the retina was injured. When discovered in the early stages, the likelihood of recovery and vision translation is nearly none. In other words, the patient's eyesight is either gone entirely or partially [4]. The acronym for optical coherence is OCT. By detecting the optical reflections of biological tissues, tomography, a sophisticated scanning technology, may see interior structures noninvasively in cross-section. OCT is a potent imaging technique for medical diagnostics that functions similarly to an optical biopsy. In the process of helping ophthalmologists assess the retina, macula, and optic nerve for damage early on and take a clear look at the bottom back of the attention [4]. The likelihood of

recovery can be improved by early illness detection or prediction. In addition to affecting the retina and a patient's vision, eye illnesses are often associated with heart disease and hypertension, which makes them predictive and diagnostic. Even illnesses in their early stages can save a patient's life.

Four main types of eye diseases are treated in this study:

CNV: A frequent condition that can cause vision loss is called choroidal neovascularisation. It is caused by the formation of new blood vessels that emerge from the choroid through a rupture in the Bruch membrane and enter the sub-retinal pigment epithelium or sub-retinal space. CNV can worsen any injury to the Bruch membrane. [4] usually affects those under 50, and early detection is crucial for timely therapy to be initiated, perhaps preventing fibrosis and the resulting irreversible loss of central visual acuity in this economically active population (Roy et al., 2017).

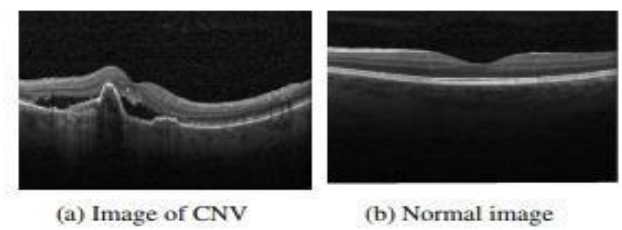
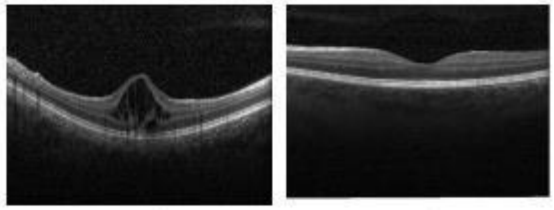


Figure 1.1: Image of CNV VS NORMAL

DME: A kind of eye illness called diabetic macular edema is brought on by damage to the retina's blood vessels. If DME is not treated, it will result in a build-up of fluid in the macula, which will enlarge the retinal layer and cause permanent blindness.

In DME, the macula area experiences a buildup of fluids and proteins. Vision is severely compromised when the retina swells. This buildup of fluid and proteins starts as a result of chronically high blood sugar, which harms the blood vessels.

The most frequent reason why diabetic individuals lose their eyesight is DME, which is characterised as a retinal thickness that involves or approaches the macula's centre. Figure 2 shows the OCT picture separated into Normal and DME.

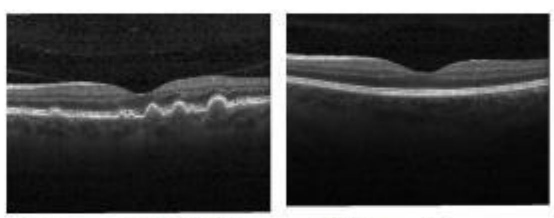


(a) Image of DME (b) Normal image

Figure 1.2: Image of DME VS Normal

DRUSEN: DRUSEN is one of the initial indications of AMD pathogenesis. Although drusen, which are yellow deposits beneath the retina, are not signs of eye disorders, their abundance can cause AMD and visual loss. These days, ophthalmologists employ OCT imaging scanning to identify DRUSEN and determine if they are serious enough to cause AMD or whether they can go away, allowing them to make decisions in advance.

Optical disc Drusen are calcified deposits of extruded mitochondria that appear in the upper part of the optic nerve in approximately 2% of the population (Gaier et al., 2017). In the figure 3 you can view the OCT image, divided into normal and DRUSEN.



(a) Image of DRUSEN (b) Normal image

Figure 1.3: Image of DRUSEN VS NORMAL [17]

1.1 Background

Classical convolution neural networks (CNN), a popular deep learning network for computer-aided diagnosis, can classify optical coherence tomography (OCT) pictures with excellent accuracy. A common criticism of classical CNN is that it suppresses positional relations in a pooling layer. In order to get around that limitation, we tried applying a capsule network to OCT pictures as it can learn positional information from images. Our goal in this study is to increase classification accuracy by using a capsule network in place of CNN.

1.2 Motivation

Labeled OCT images for classification is to increase the precision and effectiveness of diagnosing and managing ocular diseases using non-invasive imaging technology. Optical Coherence Tomography (OCT) provides high-resolution, three-dimensional images of the eye that can help identify subtle changes in tissue morphology associated with different ocular diseases. However, analyzing OCT images manually can be time-consuming, and interpretation can be subjective, leading to potential errors in diagnosis and treatment. Machine learning algorithms, trained on labeled OCT images, have the potential to overcome these challenges by providing automated, accurate, and consistent analysis. Such algorithms can classify OCT images into different categories, such as healthy, diseased, or specific diseases, with a high degree of accuracy.

1.3 Objectives

- Patients have a better chance of recovery and visual restoration if retinal injury is identified or predicted early on. Which will help the ophthalmologists to take a clear look on the back of the eye and determine at early stages the damage caused to the retina, macula, and optic nerve.

- In addition to affecting the retina and the patient's vision, eye disorders are linked to heart conditions and hypertension. This implies that early detection and diagnosis of eye diseases can help protect the patient's heart health.

- Most of the diseases are microscopic, so wanted to develop a system that automatically classify diseases

- The purpose of this research is to provide a new transfer learning-based categorisation model.
- Using retinal pictures from an optical coherence tomography (OCT) instrument, CNN architecture and a suggested model are used to automatically categorise the various retinal disorders.

1.4 Organization of Report

The current chapter gives a brief description of the leaf diseases detection technique. Figure 1.1 shows project report structure.

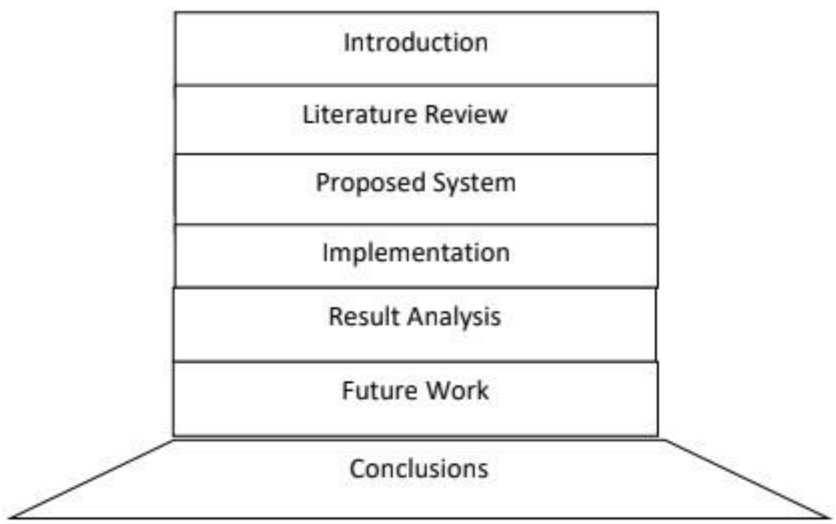


Figure 1.4: Project report structure.

1.5 Contributions

This system is designed to classify people who are able to birth a normal baby or not using some machine learning techniques such Decision tree, KNN, Random forest classifier, XGBoost, GBM, LGBM. This model is able to detect which behavior of any person is responsible for the birth of an abnormal child. A dataset is created by the survey among the people who have one or more babies. This dataset is prepared in matrix format for further use of anybody.

CHAPTER 2

REVIEW OF THE LITERATURE

Numerous studies have been conducted using a variety of machine learning and deep learning techniques to automatically classify optical coherence tomography (OCT) pictures. Some related works are:

❑ "Automated Diagnosis of Glaucoma Using Deep Learning Methods" by Fard et al. This study proposed a deep learning-based method for automated diagnosis of glaucoma using labeled OCT images

❑ "Automated Diagnosis of Retinal Diseases Using Convolutional Neural Networks" by Liu et al. This study proposed a convolutional neural network (CNN) based method for automated diagnosis of retinal diseases using labeled OCT images

❑ "Retinal Layer Segmentation in Optical Coherence Tomography Images Using Deep Learning" by Chen et al. This study proposed a deep learning-based method for automated segmentation of retinal layers in OCT images, which can be used to aid in the diagnosis and monitoring of retinal diseases

CHAPTER 3

Classification of Optical Coherence Tomography retinal image using Convolutional Neural Networks.

3.1 Optical Coherence Tomography retinal images

In this manner, an OCT image classifier can quickly and automatically categories the four types of diseases that are apparent in OCT: CNV, DME, DRUSEN, and NORMAL. The technique selected and used helps classify OCT pictures of real patients' retinas in order to determine whether or not they have any of the aforementioned conditions. A total of 84,495 pictures from the Labelled Optical Coherence Tomography dataset were utilized for the categorization (Kermany et al., 2018).

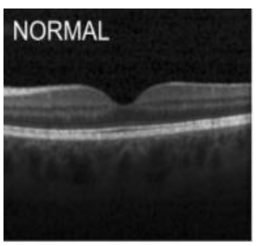


Figure 3.1.1: Normal

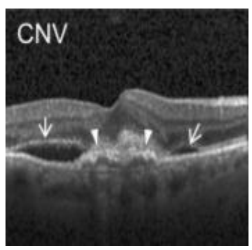


Figure 3.1.2: CNV

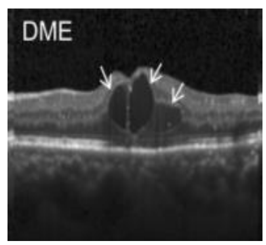


Figure 3.1.3: DME

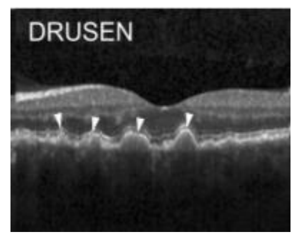


Figure 3.1.4: DRUSEN

3.2 Features Extraction Techniques

An image's features are its distinguishing elements or patterns. An picture, for instance, is regarded as a matrix. The components of the matrix, which we humans might refer to as features of the square, show the intensity of the items in that image. Properties such as ridges, corners, edges, and areas of interest are examples of features.

Convolutional neural networks (CNNs) can replace traditional feature extractors because they are more efficient, have a strong capacity to extract complex features that represent the picture in greater detail, and can learn task-specific features.

A convolutional neural network (CNN) is a type of neural network that was created to analyse multi-dimensional input, such as time series and picture data. Weight calculation and feature extraction are part of the training process. Applying a convolution operator, which is helpful for resolving complicated processes, yields the name of such networks. As seen in Fig. 2.2.1, the given input data is first sent to a feature extraction network, and the collected features are then sent to a classifier network. There are several pairs of convolutional and pooling layers in the feature extraction network. A group of digital filters make up the convolutional layer, which applies the convolution process on the input data. The threshold is determined by the pooling layer, which also serves as a dimensionality reduction layer. Several parameters must be changed during back-propagation, which reduces the number of connections in the neural network design.

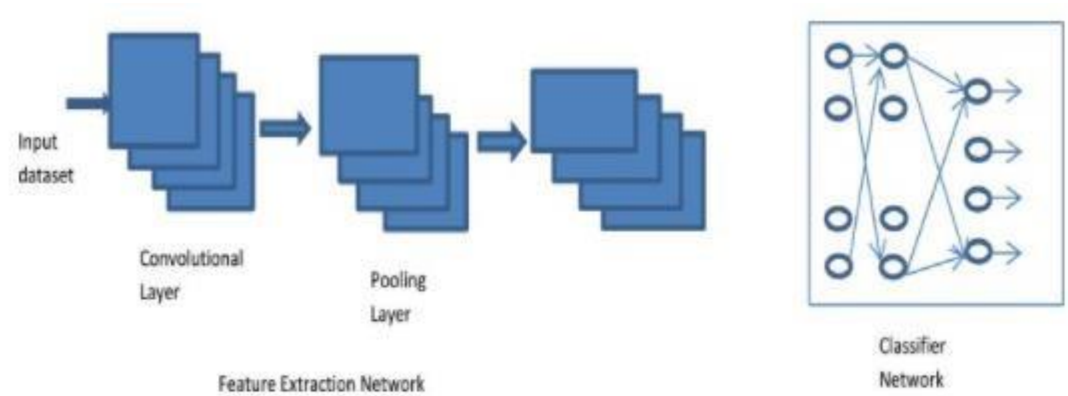


Figure 3.2.1: Features Extraction Techniques

3.3 Convolutional Neural Network

Neural networks discussed in the preceding sections are comparable to convolutional neural networks (CNN). CNNs use nonlinear activation to determine their weights, biases, and outputs. Neurons in neural networks are fully linked to the subsequent levels after receiving inputs. No connections are shared by neurons in the same layer. Regular neural networks will be exceedingly vast because of their enormous number of neurons, which will cause over-fitting when used for pictures. Due to their vast size, photos cannot be used in this way. Expand the model's size since it needs a lot of neurons. With its height, breadth, and depth, a picture may be thought of as a volume. An image's depth is its red, blue, green, or grey channels. To capitalize on the volume, a

CNN's neurons are placed volumetrically. Each of the layers transforms the input volume to an output volume as shown in the following image:

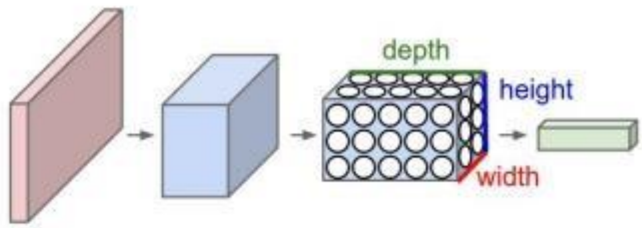


Figure 3.3.1: Convolutional neural network

3.3.1 Kernel

Kernel is the parameter convolution layer used to convolve the image. The convolution operation is shown in the following figure: **3.3.1.1**

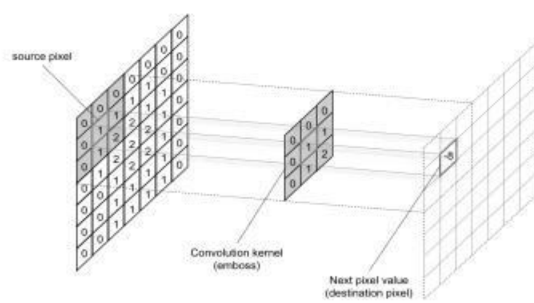


Figure 3.3.1.1: Convolutional operation

Size and stride are two of the kernel's parameters. Any rectangle's dimensions can be used for the size. The number of pixels that are shifted each time is called stride. An image of roughly the same size is produced by a stride of length 1, and half of that size is produced by a stride of length 2. Achieving the same input size can be facilitated by padding the picture.

3.3.2 Max pooling

Convolution layers are sandwiched by pooling layers. By sampling, pooling layers minimize the picture size across layers. By choosing the highest value inside a window, the sample is carried out. Over the timeframe, average pooling averages are calculated. Additionally, pooling serves as a regularization method to prevent over-fitting. All of the feature channels are used for pooling. Different strides can also be used for pooling. CNN's receptive field is indicated by the window's size.

The following figure shows an example of max pooling:

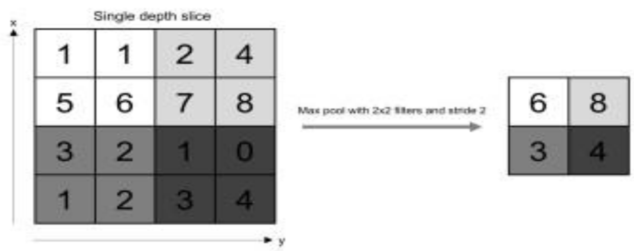


Figure 3.3.2.1: Max pooling operation

3.3.3 Activation functions

Neural nets are nonlinear because of the activation functions. Whether or whether a perceptron fires is determined by an activation function. Functions are crucial in modifying the gradients during training activation. Higher magnitude values are attenuated by an activation function. Deep networks are able to learn complicated functions because of the activation function's nonlinear behavior. With the exception of the rectified unit at 0, the most of the activation functions are differential and continuous. For every minor change in input, the output of a continuous function varies somewhat. Every point in the domain of a differential function has a derivative.

3.4 Batch normalization

Training neural networks is more stable and effective when batch normalization, also known as batch-norm, is used. The output of a layer with a zero mean and a one standard deviation is normalized. The network trains more quickly as a result of less over-fitting. When training intricate neural networks, it is quite helpful.

CHAPTER 4

Model Architectures

A particular type of multi-layer neural network called a convolutional neural network (CNN, or ConvNet) is made to identify visual patterns straight from pixel pictures with little to no preprocessing. Some model architectures are discussed below:

4.1 Brief description of CNN Transfer Learning model architectures

Training deep convolutional neural network models on very big datasets might take days or even weeks.

Reusing the model weights from previously trained models created for common computer vision benchmark datasets, such the Image Net image recognition tasks, is a method to expedite this process. You may either download and utilize the best-performing models straight away or incorporate them into a new model to solve your own computer vision issues.

This article will teach you how to create convolutional neural networks for computer vision applications using transfer learning. You will understand after reading this post: Using models that have been trained on one problem as a foundation for a related problem is known as transfer learning. Because transfer learning is adaptable, pre-trained models can be used directly, as preprocessing for feature extraction, or as part of completely new models. Many of the best-performing models on the Image Net image recognition tests, including VGG, Inception, and ResNet-50, are easily accessible using Keras.

4.2 Convolutional neural network (CNN) architecture

4.2.1 VGGNet

VGG is a traditional convolutional neural network (CNN) architecture that is sometimes referred to as VGGNet. In order to improve the model's performance, VGG was created to deepen these CNNs.

Visual Geometry Group, or VGG for short, is a common multi-layer deep Convolutional Neural Network (CNN) architecture. The term "deep" describes how many layers there are in VGG-16 or VGG-19, which have 16 and 19 convolutional layers, respectively.

Innovative object recognition models are based on the VGG architecture. The VGGNet was created as a deep neural network that outperforms Image Net on a variety of tasks and datasets. Furthermore, it remains one of the most widely used image recognition architectures today.[23]

4.2.1 VGG16

In 2014, the VGG16 convolution neural net (CNN) architecture won the ILSVR (Imagenet) competition. It is regarded as one of the best vision model architectures available today. The most distinctive feature of VGG16 is that, rather than having several hyper-parameters, they concentrated on having 3x3 convolution layers with stride 1 and consistently employed the same padding and maxpool layer of 2x2 stride 2 filters. Throughout the whole design, it continuously uses this configuration of convolution and max pool layers. Its final output consists of two FC (fully connected layers) and a soft max. The 16 in VGG16 stands for its 16 weighted layers. With over 138 million parameters, this network is rather vast.

4.3 PROPOSED ARCHITECTURES

We have proposed one convolutional neural networks CNN architectures which have 4 parts .Construction of my proposed architecture is Factorization into Smaller Convolutions with a skip connection and a with some convolutional layers, all have been trained on google Colab GPU ,which is Tesla K8 GPU, the architectures details are as follows :

4.3.4 The Layers

First, I set the model's initialization by indicating that it is a sequential model. Once the model is initialized.

```
def build_model():
    input_img = Input(shape=(224,224,3), name='ImageInput')
    x = Conv2D(64, (3,3), activation='relu', padding='same', name='Conv1_1')(input_img)
    x = Conv2D(64, (3,3), activation='relu', padding='same', name='Conv1_2')(x)
    x = MaxPooling2D((2,2), name='pool1')(x)

    x = SeparableConv2D(128, (3,3), activation='relu', padding='same', name='Conv2_1')(x)
    x = SeparableConv2D(128, (3,3), activation='relu', padding='same', name='Conv2_2')(x)
    x = MaxPooling2D((2,2), name='pool2')(x)

    x = SeparableConv2D(256, (3,3), activation='relu', padding='same', name='Conv3_1')(x)
    x = BatchNormalization(name='bn1')(x)
    x = SeparableConv2D(256, (3,3), activation='relu', padding='same', name='Conv3_2')(x)
    x = BatchNormalization(name='bn2')(x)
    x = SeparableConv2D(256, (3,3), activation='relu', padding='same', name='Conv3_3')(x)
    x = MaxPooling2D((2,2), name='pool3')(x)

    x = SeparableConv2D(512, (3,3), activation='relu', padding='same', name='Conv4_1')(x)
    x = BatchNormalization(name='bn3')(x)
    x = SeparableConv2D(512, (3,3), activation='relu', padding='same', name='Conv4_2')(x)
    x = BatchNormalization(name='bn4')(x)
    x = SeparableConv2D(512, (3,3), activation='relu', padding='same', name='Conv4_3')(x)
    x = MaxPooling2D((2,2), name='pool4')(x)

    x = Flatten(name='flatten')(x)
    x = Dense(1024, activation='relu', name='fc1')(x)
    x = Dropout(0.7, name='dropout1')(x)
    x = Dense(512, activation='relu', name='fc2')(x)
    x = Dropout(0.5, name='dropout2')(x)
    x = Dense(2, activation='softmax', name='fc3')(x)

    model = Model(inputs=input_img, outputs=x)
    return model
```

In order to prevent any negative values from being passed on to the following layer, I additionally apply relu (Rectified Linear Unit) activation to each layer.

CHAPTER 5

Methodology

One major issue with computer-aided diagnosis (CAD) for medical applications is the categorization of retinal diseases. In order to automatically identify choroidal neovascularisation (CNV), diabetic macular edema (DME), DRUSEN, and NORMAL in optical coherence tomography (OCT) pictures, this research focusses on a 4-class classification issue.

However, the viability of traditional manual OCT assessment in clinical practice has essentially become impossible due to the growing volume of image data generated by modern OCT technologies [6–8].

In this study, we prepared and processed a relatively large dataset of retinal OCT images taken in real-world settings in order to address the aforementioned problems and maximise the clinical utility of automated detection. We also aimed to investigate the use of an ensemble of four improved ResNet50 to automatically classify CNV, DME, DRUSEN, and NORMAL in order to provide an accurate and timely detection of key pathology. In order to better understand the model's decision-making process, we also conducted occlusion testing and qualitative evaluation of the model predictions. Additionally, we provide a critical analysis of the pathology and misclassification areas found by the occlusion tests. Lastly, we looked at the possible advantages of combining retinal OCT scans with information about the patients' medical histories that was gathered in the past. With its high sensitivity, high specificity, and high accuracy, the suggested method may help ophthalmologists establish a diagnosis.

In a proper manner. Here, it is agreed that each image's width and height shall be 60 pixels. There are many different methods for both upcaling and downscaling. As part of the data pretreatment approach for this thesis, the intercubic interpolation method was used for picture scaling.

In order to make sure that the photos have the same size and aspect ratio, we then preprocessed the data using the uniform aspect ratio. Since the majority of neural network models require that the input images are square, each image must be verified to be square and cropped accordingly. To choose a square portion of the image, crop it.

Normalizing picture inputs is my third method of preprocessing the data. In order to guarantee that each input parameter—in this example, a pixel—has a comparable data distribution, data normalization is a crucial step. This speeds up convergence during network training. Each pixel's mean is subtracted, and the result is then divided by the standard deviation to normalize the data. Such data would have a distribution that is similar to a Gaussian curve with zero at its center. We may decide to scale the normalized data in the interval [0, 1] or [0, 255] since we require the pixel numbers for picture inputs to be positive. The normalized data for our data-set example is shown in the montage that follows.

We employ data augmentation since there are fewer photos for two groups. Adding altered copies of existing photos to the dataset is another popular preprocessing method. Typical affine transformations include rotations, scaling, and others. To expose the neural network to a broad range of variables, this is done. As a result, the neural network is less likely to identify undesirable features in the dataset. from 11,000 pictures. As a result, we use 20% of the data for model testing and 80% for training.

For learning model I apply three types of approaches .They are Transfer Learning, Sequential CNN architecture, proposed model. A short description about them is giving below:

5.1 Transfer Learning:

Reusing components of a previously trained model in a new machine learning model is known as transfer learning. The two models can exchange generalised information if they are designed to carry out comparable tasks. The resources and quantity of labelled data needed to train new models are decreased using this method of machine learning advancement.

It is being utilized as a technique in the development process more and more, and it is becoming a significant aspect of the progress of machine learning. The resnet50, insecption50, vgg16, and vgg19 are examples of several kinds of transfer learning models.[30]

5.2 Sequential CNN architecture

We call this traditional CNN a "sequential CNN model" since each layer only links to the layers that come before and after it. The spine photos are used to train this model from scratch. Three fully-connected and five convolutional layers make up the suggested model. The input layer receives input photos with dimensions of 227 × 227 × 3. The first convolutional layer has 96 filters, each with a size of 7 by 7 and stride of 2. To improve learning, each convolutional layer is followed by an activation of a Rectified Linear unit (ReLU) [21]. The filters in each of the other convolutional layers are 3x3 in size. A Gaussian distribution with a zero mean is used to initialize the weights. Inspired by the lateral inhibition mechanism of biological neural networks, a local response normalization (LRN) layer is incorporated after the first and second convolutional layers to facilitate generalization [12]. The LRN layers and the fifth convolutional layer are followed by max-pooling layers with a stride of two and a pooling window of 3×3. The Softmax classifier receives input from the third fully connected layer, which contains two neurons. The first two fully connected layers each have 4096 neurons. In order to mitigate over-fitting problems, dropout regularization is accomplished by removing 50% of the neurons in the first and second fully-connected layers during training. Stochastic gradient descent (SGD) with momentum is used to optimize the multinomial logistic regression goal in order to train the suggested model. A randomized grid search strategy is used to optimize the model for its hyper parameters [22]. A 5×10-4 weight penalty is applied when using L2-regularization. Before convergence, the learning rate is lowered three times from its original value of 0.001. After 60 epochs, the training is terminated, and the mini-batch size is 10. Because of the implicit regularization enforced by the larger depth, lower convolutional filter dimensions, use of the L2-regularization parameter, and dropouts in the fully connected layers, the suggested model converges more quickly.[31]CNN's sequential architecture ResNe-50, VGG16, and VGG19 .

5.3 My proposed model:

I proposed a CNN architecture which has 4 parts .they are

1. Factorization into Smaller Convolutions
2. Skip connection
3. Utility of Auxiliary classifiers
4. Convolution, max-polling and fully connected layer

5.4 Confusion matrices

Confusion matrices show the number of actual and expected values. True Negative, or "TN," is the output that indicates the proportion of correctly identified negative cases. Likewise, "TP" stands for True Positive, which denotes the quantity of correctly identified positive cases. The number of real negative instances that are categorized as positive is known as the False Positive value (FP), while the number of real positive examples that are labelled as negative is known as the False Negative value (FN). Accuracy is one of the most often utilized measures while doing categorization. The accuracy of a model (through a confusion matrix) is calculated using the given formula below.[4]

$$Accuracy=TN+TP/TN+FP+FN+TP$$

Other measures based on the confusion matrix might be helpful for assessing performance because accuracy can be deceptive when applied to unbalanced datasets. The "confusion matrix()" function in Python, which is a component of the "sklearn" package, may be used to retrieve the confusion matrix [17]. You may use "from sklearn.metrics" import confusion matrix to import this function into Python. Users must supply the function with both actual and expected values in order to retrieve the confusion matrix.

The confusion matrix consists of four basic characteristics (numbers) that are used to define the measurement metrics of the classifier. These four numbers are :

- **TP (True Positive):** The number of patients who have been correctly identified as having malignant nodes, indicating that they have the illness, is represented by TP.

- **TN (True Negative):** The number of accurately diagnosed healthy patients is denoted by TN.
- **FP (False Positive):** The number of patients who are misclassified as having an illness but are actually healthy is represented by FP. Another name for FP is a Type I mistake.
- **FN (False Negative):** The number of patients who are misclassified as healthy but are actually afflicted with the illness is represented by FN. Another name for FN is a Type II mistake.

The ratio of properly diagnosed patients with the disease (TP) to all patients expected to have the disease (TP+FP) is a measure of an algorithm's precision.

Precision=TPTP+FP

The ratio of correctly diagnosed diseased patients (TP) to the total number of patients who really have the disease is known as the recall measure.

Recall=TPTP+FN

The number of patients who have been diagnosed with the illness is the perspective that underlies remembering. Another name for recall is sensitivity.

F1 score is also known as the F Measure. The F1 score states the equilibrium between the precision and the recall.[32]

F1Score=2*precision* recallprecision + recall

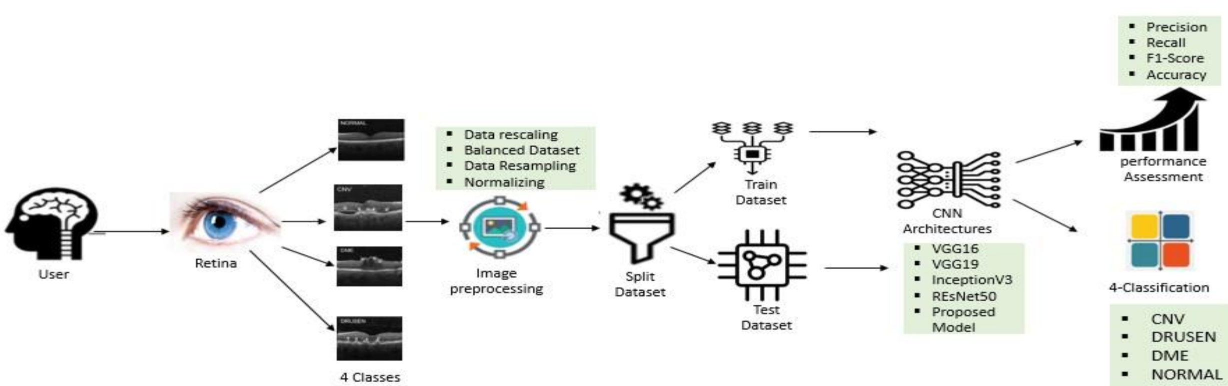


Figure 5.1: Proposed Method

CHAPTER 6

RESULTS AND DISCUSSIONS

Our result part is divided into three section .First section’s name is result with Transfer learning w/VGG-16, Transfer learning w/VGG-19, Transfer learning w/InceptionV3. Second section’s name is result with Sequential CNN Architecture VGG-16, Third section’s name is result with my proposed model.

6.1 Result Section

6.1.1 Result with Transfer learning

I talk about the outcome of transfer learning with VGG-16 here. The models' performance in classifying retinal damages will be visualized by their train accuracy, test accuracy, validation accuracy, train loss, test loss, confusion matrices, and time result.

6.1.1.1 Result in Transfer learning w/VGG-16

When we tested the Transfer learning with VGG-16 architecture on our test set data, we achieved an accuracy of 85%. The input shape in my experiment is 224X224x3, and it is fed into the CNN layers with pre-trained weights. Next, identify the key elements in the pictures. We utilize 10 epochs with a batch size of 128 to learn the model. Using the ADAM optimizer, the model's learning rate was.0001. Moreover, RELU serves as an activation function.

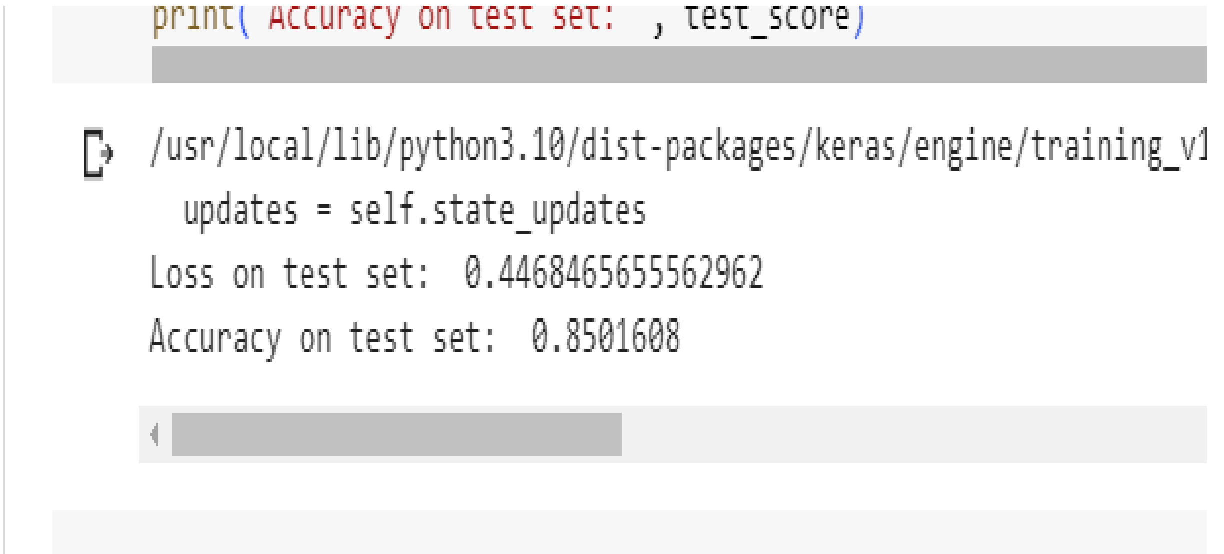


Figure6.1.1: Transfer learning w/VGG-16's Accuracy

In order to determine the model's performance, I additionally take into account the model's test and train losses. In other words, loss is a figure that represents how poorly the model predicted a particular case. The loss is zero if the model's forecast is accurate; if not, the loss is larger. Finding a collection of weights and biases that, on average, have minimal loss across all cases is the aim of model training [33].

The confusion matrix transfer learning with VGG-16 architecture is also evaluated by us. Due to a confusion matrix, the model's performance may be visualized by comparing the actual and expected outputs. The confusion matrix of the transfer learning w/VGG-16 architecture indicates that, out of the 68 pictures in the data, CNV is 0 as properly, 1 as DME, and 1 as DRUSEN, which is incorrect. Currently, the model counts 65 photos as CNV, and it properly predicts 0 images as NORMAL. One picture shows DRUSEN as incorrect, whereas four photos show DME. 57 pictures are shown here as DME, which is accurate without it. Each of the following predictions is incorrect: 3 as DRUSEN, 3 as CNV, and 7 as NORMAL. Here, 65 pictures are labelled as DRUSEN, which is accurate in its absence. One as NORMAL, two as DRUSEN, and two as CNV are all incorrect predictions.

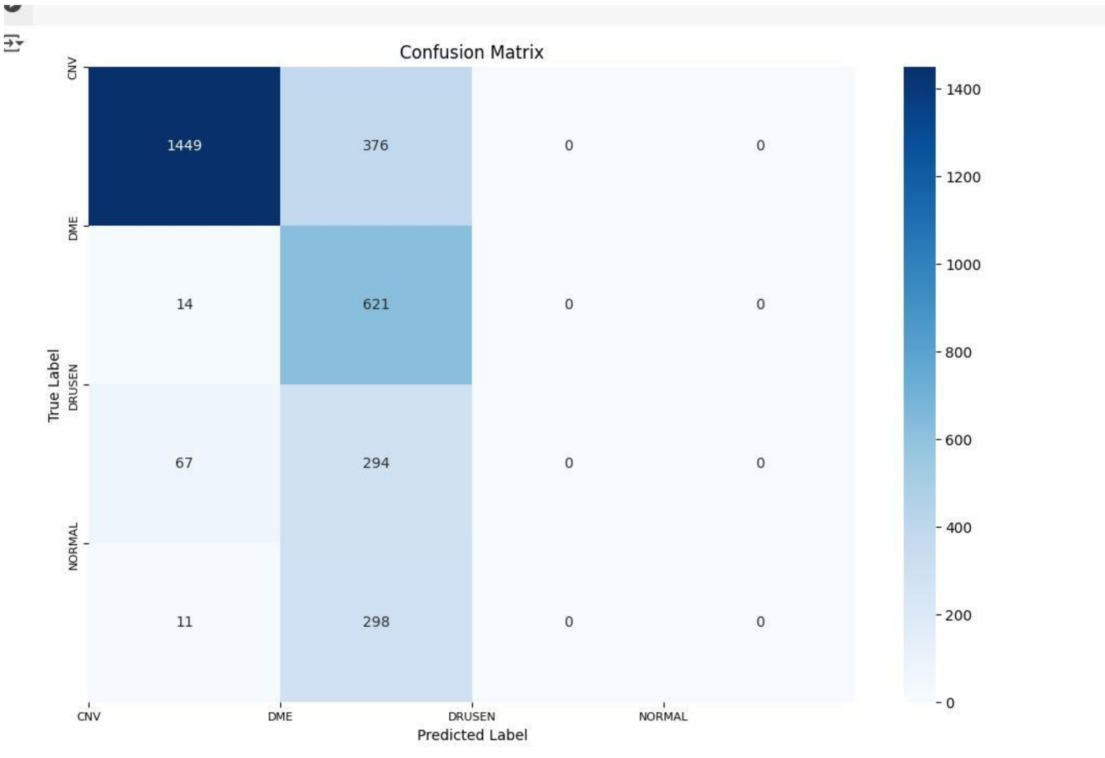


Figure6.1.2: Transfer learning w/VGG-16's Confusion matrix

For checking the performance of the model also count the precision, recall, f1-score because a model's classification can also be define by those.

Predicted Label

```
# Calculate Precision and Recall
tn, fp, fn, tp = cm.ravel()

precision = tp/(tp+fp)
recall = tp/(tp+fn)

print("Recall of the model is {:.2f}".format(recall))
print("Precision of the model is {:.2f}".format(precision))
```

```
Recall of the model is 0.93
Precision of the model is 0.76
```

Figure6.1.2: Transfer learning w/VGG-16's Precision and Recall

CHAPTER 7

Conclusion & Future Work

7.1 Conclusion

In order to categories retinal pictures acquired from an OCT scanning equipment, we suggested a novel classification model based on deep neural networks. The results indicate that the application of convolutional neural networks might provide very intriguing outcomes. 85% in image classification, offers a novel, easy, and efficient way to identify CNV, DME, and DRUSEN early, and may be used to evaluate physicians in medical diagnosis. Additionally, we have shown that using sequential models and some pre-trained models can improve model efficacy and time outcomes.

7.2 Future work

Other datasets can be used in this experiment. The model cannot be trained at 448 × 448 × 3 resolution due to compute resource limitations. Therefore, 448 x 44 8 × 3 resolution can be used to train the suggested techniques. Even though my suggested design improves retrieval accuracy, the ResNet50 architecture model can be used since its network architecture weights are somewhat high and slow to train. In further research, we will confirm the outcome using a wide range of datasets.

References

[1] https://en.wikipedia.org/wiki/Human_eye

[2] https://my.clevelandclinic.org/health/articles/21204-vision

[3]https://associatedretinaconsultants.com/symptomsofdamagedretina/#:~:text=Retinal%20damage%20is%20one%20of,serious%20enough%20to%2 0cause%20blindness.

[4] Deep learning for identifying and classifying retinal diseases

[5] Classification of Optical Coherence Tomography using Convolutional Neural Networks

[6] Ali, Liaqat & Khelil, Khaled & Kanwal et al. (2017). Machine learning based computer-aided diagnosis of liver tumours. 139-145. 10.1109/ICCI-CC.2017.8109742.

[7] Hoover, A., and Goldbaum, M. (2003). Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels. IEEE Trans. Med. Imaging 22, 951–958.

[8] Hoover, A., Kouznetsova, V., and Goldbaum, M. (2000). Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Trans. Med. Imaging 19, 203–210

[9] Kaiser, P.K., Brown, D.M., Zhang, K., Hudson, H.L., Holz, F.G., Shapiro, H., Schneider, S., and Acharya, N.R. (2007). Ranibizumab for predominantly classic neovascular age-related macular degeneration: subgroup analysis of first-year ANCHOR results. Am. J. Ophthalmol. 144, 850–857

[10] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. In 2016 IWWW Conference on Computer Vision and Pattern Recognition, pp. 2818–2826.