International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

# Direct Cup-To-Disc Ratio Estimation for Glaucoma Screening Via Semi-Supervised Learning

**Sabah Samareen*1, P. Meghana Reddy *2 Saniya Naaz*3, S. Surekha*4**

***1B.Tech. Student, * 2B.Tech. Student, * 3B.Tech. Student, *4Assistant Professor**

ECE Department, Bhoj Reddy Engineering College for Women, Hyderabad, Telangana, India

**ARTICLEINFO**

**Article History:**

Accepted: 01 June 2023
Published: 12 June 2023

**Publication Issue**
Volume 10, Issue 3
May-June-2023

**Page Number**
437-449

**ABSTRACT**

Glaucoma is a chronic eye disease that leads to irreversible vision loss. The Cup-to-Disc Ratio (CDR) serves as the most important indicator for glaucoma screening and plays a significant role in clinical screening and early diagnosis of glaucoma. In general, obtaining CDR is subjected to measuring on manually or automatically segmented optic disc and cup. Despite great efforts have been devoted, obtaining CDR values automatically with high accuracy and robustness is still a great challenge due to the heavy overlap between optic cup and neuroretinal rim regions. Automation of CDR estimation using Ml is proposed using three networks i.e. CNN, MLP, ResNet. Their performance in terms of accuracy and loss is compared and graphs are obtained and analyzed. The CNN model is often used for computer vision applications hence, it is compared with MLP and ResNet models to understand and compare performance.

Keywords : CNN, MLP, ResNet, Cup-to-Disc Ratio

## I. INTRODUCTION

1. AIM

The aim of our project is to develop different techniques for detection of glaucoma using ML. Reliance on such automated algorithms reduces the probability of manual error, thus, providing greater accuracy, which is key in early detection and treatment.

## II. INTRODUCTION TO GLAUCOMA

Glaucoma can be defined as a category of related eye disorders that can damage the optic nerve. The main function of the optic nerve is that it carries information from the eye to the brain decaying of this nerve due to various reasons leads to blindness. It is crucial that glaucoma is detected as early as possible this allows proper treatment. There are multiple factors associated with this disease, however, the main cause of glaucoma is the build-up of IOP in the eye that results from blockage of intraocular fluid drainage. Although the exact cause of this blockage is unknown, it tends to be inherited and can be linked to old age, ethnicity, steroid medication, and other diseases like diabetes. The increased IOP damages the optic nerve that carries visual information of photo receptors from eye to brain. Generally, glaucoma does not show any signs or symptoms until it has progressed to advanced stage at which point the damage becomes irreversible. It has been reported that the damage to optic nerve fibres becomes noticeable and reduction in visual field is detected when about 40% of axons are already lost. However, it is possible to slow down the impairment caused by glaucoma if it is diagnosed sufficiently early. Recently, WHO recognized glaucoma as the third biggest cause of blindness after un-operated cataract and uncorrected refractive errors and the leading cause of irreversible vision loss. Glaucoma is normally diagnosed by obtaining medical history of patients, measuring IOP, performing visual field loss test, and conducting manual assessment of Optic Disc (OD) using ophthalmoscopy to examine the shape and colour of optic nerve. Optic Disc is the cross-sectional view of optic nerve connecting to the retina of each eye. It looks like a bright round spot in retinal fundus image. In case of glaucoma, the IOP Maintaining the Integrity of the Specifications damages the nerve fibres constituting optic nerve. As a result, OD begins to form a cavity and develops a crater- like depression, at the front of the nerve head, called Optic Cup (OC). The boundary of the disc also dilates and the colour changes from healthy pink to pale. The Cup-to-Disc Ratio (CDR) is one of the major structural image cues considered for glaucoma detection. Figure 1 shows healthy optic disc and its condition during various stages of glaucoma.

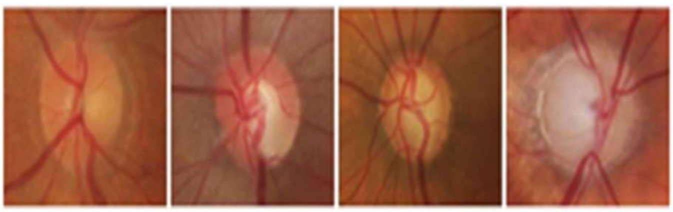

Fig 1 Stages of glaucoma. (a) Healthy Disc (b) Early Glaucoma (c) Moderate Glaucoma

In retinal images, some of the important structural indications of glaucoma are disc size, CDR, Ratio of Neuro-retinal Rim in Inferior, Superior, Nasal and Temporal quadrants (ISNT rule), and Peripapillary Atrophy (PPA) etc. These indications are usually concentrated in and around OD. Therefore, segmentation of this Region Of Interest (ROI), that is detecting the contour of OD, is not only useful for a more focused clinical assessment by the ophthalmologists but also helpful in training a computer based automated method for classification. However, automated glaucoma detection techniques based upon segmented discs are very sensitive to the accuracy of segmentation and even a small error in delineation of OD may affect the diagnosis significantly. Localization, on the other hand, gives the exact location of OD in the whole image with some surrounding context.

Automatic methods for glaucoma detection based upon this approach of ROI extraction are more resilient to localization errors. From automated classification point of view the disease pattern in retinal fundus images is inconspicuous and complex. Detecting ROI from natural scene images is comparatively easy because it has an obvious visual appearance, for example color, shape and texture etc. In contrast, the significant features of disease in medical images are hidden and not readily discernible except by highly trained and qualified field experts.

Deep Learning, however, has been shown to learn discriminative representation of data that can identify otherwise unnoticeable characteristics. Such algorithms achieve this useful and compact representation of data by applying multiple linear and non-linear transformations on training data in a cascaded fashion. Such Computer Aided Diagnosis (CAD) can be very helpful in providing standardized and cost-effective screening at a larger scale. These automated systems may reduce human error, offer timely services at remote areas, and are free from clinician's bias and fatigue.

## III. EXISTING METHOD - SEGMENTATION BASED APPROACH

In practice, obtaining reliable CDR value is subjected to measuring on segmented optic disc and cup, which usually obtained by manually contouring the borders of optic disc/cup or manual correction of contours generated by segmentation algorithms. However, manually contouring of the optic disc/cup borders is time-consuming and subjective to personal experiences. Due to the lack of sharp border information of optic disc/cup, the CDR value of the same subject often varies among different clinicians. Recently, great efforts have been devoted into automatizing the procedure. Numerous automated segmentation methods are proposed as a prerequisite to segregate disc/cup regions from the complex surroundings with clear borders, including statistical shape model, Multiview and multimodal approaches, super pixel based methods and deep learning methods. Usually, these methods require strong prior information and user interaction to increase its accuracy. Although the segmentation- based methods obtained effective performance by leveraging state-of-the-art machine learning especially deep learning techniques, accurate measurement of CDR value is still a challenging task due to the following

1) Heavy overlap and extremely weak contrast between optic cup and neuroretinal rim regions which make the automated segmentation algorithm unable to distinguish the boundaries of optic cup in fundus image;
2) The great variability of shape and inhomogeneity in appearance of optic disc which leads to critical inconsistency of the measured CDR compared with the actual one;
3) Insufficient pixel-level labels (e.g. segmentation mask) which is not enough to learn the outstanding model for optic disc/cup segmentation and CDR measurement; and
4) The error introduced by the intermediate steps
5) Even more challenges arise from the presence of various pathologies.

Although optic disc can be spotted manually as a round bright spot in a retinal image, yet performing large scale manual screening can prove to be really tiresome, time consuming, and prone to human fatigue and predisposition. CAD can provide efficient and reliable alternative solution with near human accuracy. Usually, the disc is the brightest region in the image. However, if ambient light finds its way into the image while capturing the photo it can look brighter than optic disc. Furthermore, occasionally some shiny reflective areas appear in the fundus image during image capturing. These shiny reflections can also look very bright and mislead a heuristic algorithm in considering them as candidate regions of interest.

## IV. LITERATURE SURVEY

Eswaran, et al. [1] proposed an algorithm for the extraction of OD and exudates from fundus images based on marker controlled watershed segmentation is presented. The proposed algorithm makes use of average filtering and contrast adjustment as preprocessing steps before the watershed transformation is applied. The performance of the proposed algorithm is evaluated using the test images of stare and drive databases. The results are compared with those reported earlier in the literature. It is shown that the proposed method can yield an average sensitivity value of 94%, which is higher than the value reported earlier.

Chrastek, et al. [2] proposed an automated method for the optic disc segmentation is presented. The method consists of 4 steps: localization of the optic disc, nonlinear filtering, Canny edge detector and Hough transform. The results have shown that the algorithm is very robust. The localization was 97% successful and the segmentation 82%.

Abràmoff, et al. [3] proposed that the optic disc can be selected by taking only top 5% brightest pixels and hue values in the yellow range. The surrounding pixels are then clustered to constitute a candidate region. The clusters which are below a certain threshold are discarded. Special attention is given to quantitative techniques for analysis of fundus photographs with a focus on clinically relevant assessment of retinal vasculature, identification of retinal lesions, assessment of optic nerve head (ONH) shape, building retinal atlases, and to automated methods for population screening for retinal diseases. A separate section is devoted to 3- D analysis of OCT images, describing methods for segmentation and analysis of retinal layers, retinal vasculature, and 2-D/3- D detection of symptomatic exudate associated derangements, as well as to OCT-based analysis of ONH morphology and shape. Throughout the paper, aspects of image acquisition, image analysis, and clinical relevance were treated together by considering their mutually interlinked relationships.

J. Liu, et al [4], they first divided the image into 8 × 8 pixels grid and selected the block with maximum number of top 5% brightest pixels as the centre of the disc. for the cup, two methods making use of color intensity and threshold level set were evaluated. A batch of 73 retinal images from the Singapore Eye Research Centre was used to assess the performance of the determined CDR to the clinical CDR, and it was found that the threshold and variational level set methods produced 97% accuracy in the determined CDR results, an 18% improvement over the color intensity method.

Nyúl, et al [5], employed an adaptive thresholding with a window whose size is determined to approximately match the size of the vessel thickness. A mean filter with the large kernel is then used with threshold probing for rough localization. The proposed data-driven approach requires no manual assistance and does not depend on explicit structure segmentation and measurements. First, disease independent variations, such as nonuniform illumination, size differences, and blood vessels are eliminated from the images. Then, the extracted high dimensional feature vectors are compressed via PCA and combined before classification with SVMs takes place. The technique achieves an accuracy of detecting glaucomatous retina fundus images comparable to that of human experts. Another extensively used approach is threshold-based localization. A quick look at the retinal image tells that the optic disc is mostly the brightest region in the image.

## V. BACKEND LIBRARIES AND PACKAGES

The Python programming language has been used as it best fits machine learning due to its independent platform and its popularity in the programming community. The packages and libraries used have been discussed below.

TENSORFLOW

Created by the Google Brain team and initially released to the public in 2015, TensorFlow is an open-source library that serves mainly as a backend for other libraries like keras and is used for numerical computation and large-scale machine learning. TensorFlow bundles together a slew of machine learning and deep learning models and algorithms (aka neural networks) and makes them useful by way of common programmatic metaphors. It uses Python or JavaScript to provide a convenient front-end API for building applications, while executing those applications in high- performance C++.

KERAS

Keras is a high-level, deep learning API developed by Google for implementing neural networks. It is written in Python and is used to make the implementation of neural networks easy. It also supports multiple backend neural network computation. Keras is relatively easy to learn and work with because it provides a python frontend with a high level of abstraction while having the option of multiple back-ends for computation purposes. This makes Keras slower than other deep learning frameworks, but extremely beginner-friendly.

## VI. MODULES CONV2D

Keras Conv2D is a 2D Convolution Layer, this layer creates a convolution kernel that is combined with layers input which helps produce a tensor of outputs (a tensor is an algebraic object that describes a multilinear relationship between sets of algebraic objects related to a vector space and the relationship between a quantitative dependent variable and two or more independent variables using a straight line is called a multilinear relationship).

MAXPOOLING2D

Keras MaxPooling2D is a pooling or max poolingoperation which calculates the largest or maximum value in every patch and the feature map. The results will be down sampled, or it will pool features map which was highlighting the most present feature into the patch which contains the average feature presence from the average pooling. The max pooling is found to work well in average pooling for vision tasks.

DENSE

Keras dense is one of the available layers in keras models, most oftenly added in the neural networks. This layer contains densely connected neurons. Each of the individual neurons of the layer takes the input data from all the other neurons before a currently existing one. Internally, the dense layer is where various multiplication of matrix vectors is carried out. We can train the values inside the matrix as they are nothing but the parameters. We can even update these values using a methodology called backpropagation. The dense layer produces the resultant output as the vector, which is m dimensional in size. This is why the dense layer is most often used for vector manipulation to change the dimensions of the vectors. The dense layer can also perform the vectors' translation, scaling, and rotation operations.

FLATTEN

It involves a flattening process which is mostly used as the last phase of CNN (Convolution Neural Network) as a classifier. This structure is used for creating a single feature vector for verification with keras flatten. As mentioned, it is used for to manipulate and make keras flattening happen accordingly

SEQUENTIAL

The core idea of Sequential API is simply arranging the Keras layers in a sequential order and so, it is called ways or say model which takes only one input as feed and expects one output as its name suggests. This type of model is quite capable to handle simple and layer-based problems. As its name suggests it is one of the models that is used to investigate varied types of neural networks where the model gets in one input as feedback and expects an output as desired. The Keras API and library is incorporated with a sequential model to judge the entire simple model not the complex kind of model. It passes on the data and flows in sequential order from top to bottom approach till the data reaches at end of the model.

PREPROCESSING.IMAGE

Apart from collecting some images as training data, it is necessary to employ augmentation to create variations in the image. It is especially true for more complex object recognition problemsImage, text, and sequence data can be preprocessed using keras. Keras preprocessing is the utility that was located at tf.keras preprocessing module.

It is a data augmentation and preprocessing module for the keras library of deep learning. This module provides the utility to work with text, image, and sequence data. Preprocessing layer is specifically designed for using early stages in the neural network. Image preprocessing involves resizing for rotating the image to adjust the contrast.

OPENCV-CV2

cv2 is the module import name for OpenCV module in python. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception. The library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms.

## VII. CALLBACKS

A callback is an object that can perform actions at various stages of training (e.g. at the start or end of an epoch, before or after a single batch, etc). Callbacks can be used for writing TensorBoard logs after every batch of training to monitor metrics. TensorBoard is a visualization tool provided with TensorFlow. It is a tool for providing the measurements and visualizations needed during the machine learning workflow. It enables tracking experiment metrics like loss and accuracy, visualizing the model graph.

## VIII. ACTIVATION FUNCTIONS

Activation functions are necessary to prevent linearity. Without them, the data would pass through the nodes and layers of the network only going through linear functions. The composition of these linear functions is again a linear function and so no matter how many layers the data goes through, the output is always the result of a linear function. For complex problems, the data cannot be modelled well by a linear equation. The activation functions used in the models have been discussed below.

RELU

ReLU stands for Rectified Linear Unit. Although it gives an impression of a linear function, ReLU has a derivative function and allows for backpropagation while simultaneously making it computationally efficient. The main catch here is that the ReLU function does not activate all the neurons at the same time. ReLu is generally used in the hidden layers of the model. The neurons will only be deactivated if the output of the linear transformation is less than 0. Mathematically it can be represented as:

$$f(x) = \max(0, x)$$

SIGMOID

This function takes any real value as input and outputs values in the range of 0 to 1. The larger the input (more positive), the closer the output value will be to 1.0, whereas the smaller the input (more negative), the closer the output will be to 0.0, as shown below. Mathematically, it can be given as:

$$f(x) = \frac{1}{1+e^{-x}}$$

SOFTMAX

The output of the sigmoid function was in the range of 0 to 1, which can be thought of as probability. The Softmax function is described as a combination of multiple sigmoids. It calculates the relative probabilities. The SoftMax function returns the probability of each class. It is most commonly used as an activation function for the last layer of the neural network in the case of multi-class classification. Probabilities produced by a Sigmoid are independent. Furthermore, they are not constrained to sum to one. The reason for this is because the Sigmoid looks at each raw output value separately. Mathematically, it can be represented as:

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum e^{z_i}}$$

## IX. OPTIMZER

The ultimate goal of ML model is to reach the minimum of the loss function. After the input has been given, the error is calculated and the weights are updated accordingly. This is where optimizer comes into play. It defines how to tweak the parameters to get closer to the minima. Essentially, optimization is a process of finding optimal parameters for the model, which significantly reduces the error function. The optimizer used by all the model is the Adam optimizer. Essentially Adam is a combination of Momentum and RMSProp. It has reduced oscillation, a more smoothed path, and adaptive learning rate capabilities. Combining those abilities makes it the most powerful and suitable for different problems optimizer. The equations that express the updating of weights are given below:

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{(v_t+\varepsilon)}} * m_t$$

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\, dw_t$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2)\, (dw_t)^2$$

Where, w represents the weight/ parameter, η reresentslearning rate, β represents moving average parameter such that 0<β<1, m represents aggregate of gradients, v represents sum of square of past gradients.

## X. LOSS FUNCTION

The loss function estimates how well a particular algorithm models the provided data. A loss function is a measure of how good a prediction model does in terms of being able to predict the expected outcome(or value). The learning problem is converted into an optimization problem, define a loss function and then optimize the algorithm to minimize the loss function. The loss function used is binary cross entropy. It is used in binary classification problems like two classes. example a person has covid or not or my article gets popular or not. Binary cross entropy compares each of the predicted probabilities to the actual class output which can be either 0 or 1. It then calculates the score that penalizes the probabilities based on the distance from the expected value. That means how close or far from the actual value.

## XI. MODEL 1- CONVOLUION NEURAL NETWORK

A Convolutional Neural Network, also known as CNN or ConvNet, is a class of neural networks that specializes in processing data that has a grid-like topology, such as an image. A digital image is a binary representation of visual data. It contains a series of pixels arranged in a grid-like fashion that contains pixel values to denote how bright and what color each pixel should be. Once the filters have passed over the image, a feature map is generated for each filter. These are then taken through an activation function, which decides whether a certain feature is present at a given location in the image.

In Pooling, it basically takes a filter and a stride of the same length. It then applies it to the input volume and outputs the maximum number in every sub-region that the filter convolves around. The intuitive reasoning behind this Pooling layer is that once it is known that a specific feature is in the original input volume (there will be a high activation value), its exact location is not as important as its relative location to the other features. Thus, this layer drastically reduces the spatial dimension (the length and the width change but not the depth) of the input volume.

The network has 3 layers:

• The kernel size used throughout is 3, 3

• The first layer/ input layer has 2 parts i.e. convolution and max pooling, (62, 62) represents the convolved input size (64-3+1, 64-3+1) and 32 represents the dimensions of output, i.e. number of convolution operations performed or number of filters (each filter represents a feature or a set of features), max pooling is performed on the output of convolution, MaxPooling kernel has a shape of (2, 2) and strides of (2, 2). Applying that to a (62, 62) image results in an image of shape (((62 - 2)//2) + 1, ((62 - 2)//2) + 1)) = (31, 31).

• The second layer/ hidden layer is similar to the input layer and also comprises of convolution and max pooling operations. (29, 29) represents size of the convolved input from first layer, this is followed by pooling which has taken up only 14, 14 features from convolution operation as a result of pooling. Convolving a (31, 31) image with a (3, 3) filter, with strides results in an output of size (31 - 3 + 1, 31 - 3 + 1) = (29, 29). Since there are 32 such filters, the output shape becomes (29, 29, 32). The default MaxPooling kernel has a shape of (2, 2) and strides of (2, 2). Applying that to a (29, 29) image results in an image of shape (((29 - 2)//2) + 1, ((29 - 2)//2) + 1)) = (14, 14).

• The final layer is a densely connected network that finally leads to a single output neuron. The Flatten layer takes all pixels along all channels and creates a 1D vector. Therefore, an input of (14, 14, 32) is flattened to (14 * 14 * 32) = 6272 values. 6272 represents number of elements in input tensor

• The number of parameters represent number of weights and biases trained

## XII. MODEL 2- MULTI-LAYER PERCEPTRON

A multilayer perceptron (MLP) is a type of artificial neural network that generates a set of outputs from a set of inputs. An MLP is characterized by several layers of input nodes connected as a directed graph between the input nodes connected as a directed graph between the input and output layers. MLP uses backpropagation for training the network. MLP is a deep learning method. Multi-Layer perceptron defines the most complex architecture of artificial neural networks. It is substantially formed from multiple layers of the perceptron. A perceptron can be thought of as a standalone model, while the artificial neuron is the smallest computational unit of a neural network, so it's an abstraction for a relatively simple function (e.g. sigmoid) that will be composed with other simple functions to produce a more complicated function, which is typically non-linear.

The network has 3 layers:

• The network is dense i.e. every neuron of one layer is connected to every other neuron of the next layer

• The structure and dimension of each layer has been defined duringthe development

• There are no convolution or max pooling operations being performed here

## XIII. MODEL 3- RESNET-50

Residual Network (ResNet) is a deep learning model used for computer vision applications. It is a Convolutional Neural Network (CNN) architecture designed to support hundreds or thousands of convolutional layers. Previous CNN architectures were not able to scale to a large number of layers, which resulted in limited performance. However, when adding more layers, researchers faced the “vanishing gradient” problem.

Neural networks are trained through a backpropagation process that relies on gradient descent, shifting down the loss function and finding the weights that minimize it. If there are too many layers, repeated multiplications will eventually reduce the gradient until it “disappears”, and performance saturates or deteriorates with each layer added. ResNet provides an innovative solution to the vanishing gradient problem, known as “skip connections”. ResNet stacks multiple identity mappings (convolutional layers that do nothing at first), skips those layers, and reuses the activations of the previous layer. Skipping speeds up initial training by compressing the network into fewer layers.

Most ResNet models skip two or three layers at a time with nonlinearity and batch normalization in between. More advanced ResNet architectures, known as HighwayNets, can learn “skip weights”, which dynamically determine the number of layers to skip. The main motivation of the ResNet original work was to address the degradation problem in a deep network. Adding more layers to a sufficiently deep neural network would first see saturation in accuracy and then the accuracy degrades.

the ResNet 50 architecture contains the following elements:

- A convoultion with a kernel size of 7 * 7 and 64 different kernels all with a stride of size 2 giving 1 layer.
- The next layer is max pooling with a stride size of 2.
- In the next convolution there is a 1 * 1,64 kernel following this a 3 * 3,64 kernel and at last a 1 * 1,256 kernel, these three layers are repeated in total 3 times thus giving 9 layers in this step.
- Next layer is the kernel of 1 * 1,128 after that a kernel of 3 * 3,128 and at last a kernel of 1 * 1,512 this step was repeated 4 times thus giving 12 layers in this step.
- After that there is a kernel of 1 * 1,256 and two more kernels with 3 * 3,256 and 1 * 1,1024 and this is repeated 6 time thus giving a total of 18 layers.
- And then again a 1 * 1,512 kernel with two more of 3 * 3,512 and 1 * 1,2048 and this was repeated 3 times thus giving a total of 9 layers.
- Finally, an average pool and end it with a fully connected layer containing 1000 nodes and at the end a softmax function so this gives us 1 layer.

The activation functions and the max/ average pooling layers are not counted. Thus totalling this it gives 1 + 9 + 12 + 18 + 9 + 1

= 50 layers Deep Convolutional network.

## XIV. ACCURACY AND LOSS COMPARISON

The orange graph indicates the accuracy/loss performance of the model on training dataset and the blue graph indicates the accuracy/loss performance on test data.

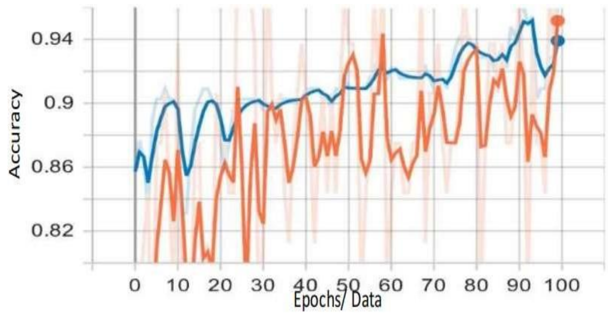


Fig 2 Accuracy graph of CNN

The training graph indicates an impressive accuracy approximately 94% by just 60 epochs. Throughout the epochs, the accuracy mainly lies between 85% to 95%. During the testing, the accuracy never falls below 85% and approached a maximum of 95%.

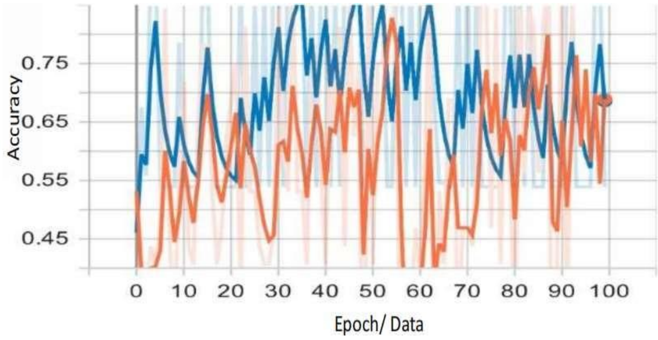


Fig 3 Accuracy graph of MLP

The accuracy of the training data seems to oscillate between 40% to 90% throughout. This shows that the CNN model performed well very quickly and remained stabilized compared to the MLP. The test data set accuracy also seems to fluctuate between 45% to 90%.

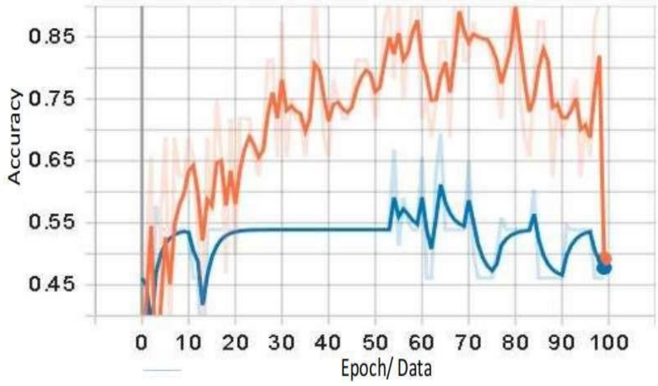


Fig 4 Accuracy graph of ResNet-50

The ResNet-50 model seems to have a stable accuracy between 75% to 90%. However, the model performs very badly on the test dataset.

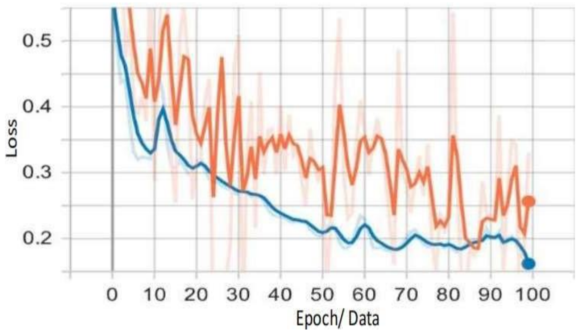


Fig 5 Loss graph of CNN

The loss of the CNN model for the training and testing data remains within a low range of 0.2 to 0.5 approximately and the loss was significantly reduced within the first 20 epochs.

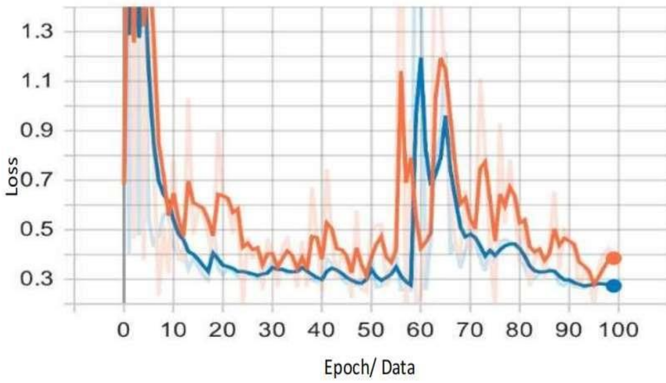


Fig 6 Loss Graph of MLP

The loss of the MLP model for the training and testing data remains quite stable and low throughout the epochs, however, a sudden spike can be observed around epoch 50 to 70 which is due to image processing and spatial variance of the model.

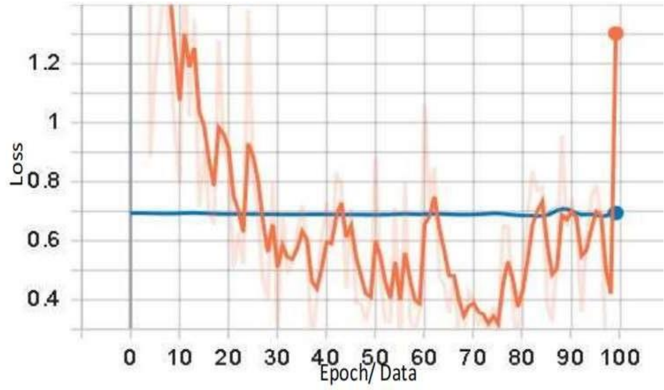


Fig 7 Loss Graph of ResNet-50

The loss of the ResNet-50 model for the testing data remains quite stable at 0.7. Even though the loss at training data is quite high it remains quite stable. At lower epochs, it is observed that the loss was well above 1.0.

Table 1. Average accuracy and loss values

| Network | Accuracy(%) | Loss |
|---|---|---|
| CNN | 87.79 | 0.2618 |

| | | |
|---|---|---|
| MLP | 87.02 | 0.2607 |
| ResNet-50 | 45.04 | 0.9493 |

## XV. OUTPUTS

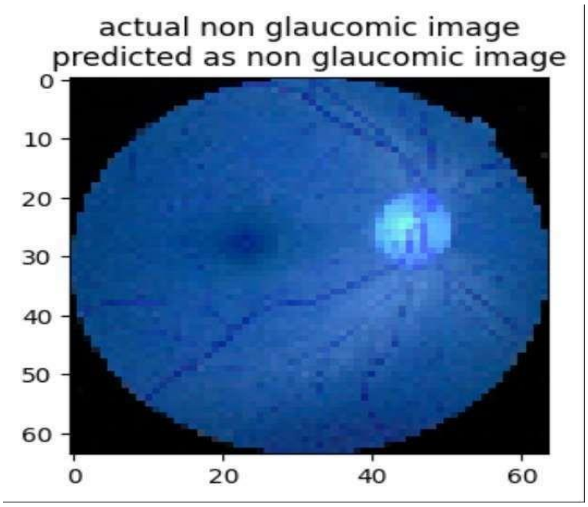


Fig 8 Output of CNN

As the image indicates this image was rightly predicted by the CNN model, which shown an accuracy of 87.79% and loss of 0.2618.

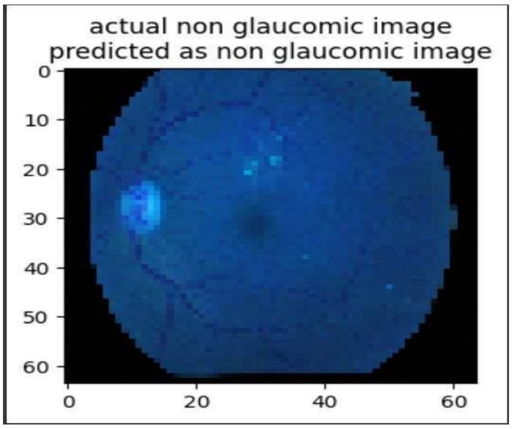


Fig 9 Output of MLP

As the image indicates this image was rightly predicted by the MLP model, which shown an accuracy of 87.79% and loss of 0.2607.

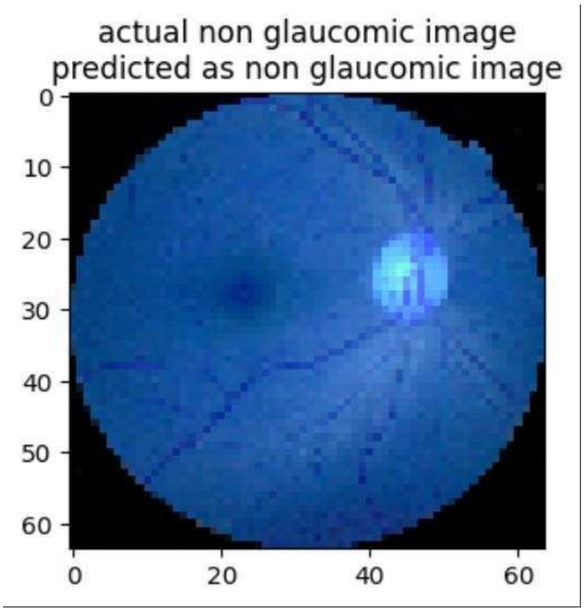


Fig 10. Output of ResNet-50

As the image indicates this image was rightly predicted by the ResNet-50 model, which shown an accuracy of 45.04% and loss of 0.9493.

## XVI. CONCLUSION

CNNs are most commonly employed in computer vision. Given a series of images or videos from the real world, with the utilization of CNN, the AI system learns to automatically extract the features of these inputs to complete a specific task, e.g., image classification, face authentication, and image semantic segmentation. Different from fully connected layers in MLPs, in CNN models, one or multiple convolution layers extract the simple features from input by executing convolution operations. Each layer is a set of nonlinear functions of weighted sums at different coordinates of spatially nearby subsets of outputs from the prior layer, which allows the weights to be reused. Applying various convolutional filters, CNN machine learning models can capture the high-level representation of the input data, making CNN techniques widely popular in computer vision tasks. Convolutional neural network example applications include image classification (e.g., AlexNet, VGG network, ResNet, MobileNet) and object detection (e.g., Fast R-CNN, Mask R-CNN, YOLO, SSD).

MLPs models are the most basic deep neural network, which is composed of a series of fully connected layers. Today, MLP machine learning methods can be used to overcome the requirement of high computing power required by modern deep learning architectures. Deep residual networks like the popular ResNet-50 model is a convolutional neural network (CNN) that is 50 layers deep. A Residual Neural Network (ResNet) is an Artificial Neural Network (ANN) of a kind that stacks residual blocks on top of each other to form a network. In recent years, the field of computer vision has undergone farreaching transformations due to the introduction of new technologies. As a direct result of these advancements, it has become possible for computer vision models to surpass humans in efficiently solving different problems related to image recognition, object detection, face recognition, image classification, etc. In this regard, the introduction of deep convolutional neural networks or CNNs deserves special mention. These networks have been extensively used for analyzing visual imagery with remarkable accuracy. But, while it gives us the option of adding more layers to the CNNs to solve more complicated tasks in computer vision, it comes with its own set of issues. It has been observed that training the neural networks becomes more difficult with the increase in the number of added layers, and in some cases, the accuracy dwindles as well. It is here that the use of ResNet assumes importance. Deeper neural networks are more difficult to train. With Resnet, it becomes possible to surpass the difficulties of training very deep neural networks.

Both MLP and CNN can be used for Image classification however, MLP is good for simple image classification, CNN is good for complicated image classification. MLPs use one perceptron for each input (pixel in an image) and the amount of weights rapidly becomes unmanageable for large images. It includes too many parameters because it is fully connected. Each node is connected to every other node in next and the previous layer, forming a very dense web, resulting in redundancy and inefficiency. As a result, difficulties arise whilst training and overfitting can occur which makes it lose the ability to generalize. Another common problem is that MLPs react differently to an input (images) and its shifted version they are not translation invariant. For example, if a picture of a cat appears in the top left of the image in one picture and the bottom right of another picture, the MLP will try to correct itself and assume that a cat will always appear in this section of the image. Hence, MLPs are not the best idea to use for image processing.

CNN’s leverage the fact that nearby pixels are more strongly related than distant ones. They analyze the influence of nearby pixels by using something called a filter /Kernel and we move this across the image from top left to bottom right. For each point on the image, a value is calculated based on the filter using a convolution operation. A filter could be related to anything, for pictures of humans, one filter could be associated with seeing noses, and our nose filter would give us an indication of how strongly a nose seems to appear in our image, and how many times and in what locations they occur. This reduces the number of weights that the neural network must learn compared to an MLP, and also means that when the location of these features changes it does not throw the neural network off. Apart from this, the weights are smaller and shared less wasteful, easier to train than MLP and more effective too. They can also go deeper. Layers are sparsely connected rather than fully connected. Additionally, it can also find the similar pattern even if the object is somewhat rotated/tilted using a concept called Pooling, which makes CNN more robust to changes in the position of the feature in the image. After the filters have passed over the image, a feature map is generated for each filter. These are then taken through an activation function, which decides whether a certain feature is present at a given location in the image. In Pooling, it basically takes a filter and a stride of the same length. It then applies it to the input volume and outputs the maximum number in every sub-region that the filter convolves around. The intuitive reasoning behind this Pooling layer is that once it is known that a specific feature is in the original input volume (there will be a high activation value), its exact location is not as important as its relative location to the other features. Thus, this layer drastically reduces the spatial dimension (the length and the width change but not the depth) of the input volume.

The neural network (in MLP) will learn different interpretations for something that is possibly the same. But in CNN, the number of weights is dependent on the kernel size (see Weight sharing) instead of the input size which is really important for images. So, by forcing the shared weights among spatial dimensions which drastically reduces the number of parameters, the convolution kernel acts as a learning framework. That's how convolutional layers reduce memory usage and compute faster.

## XVII. REFERENCES

[1]. Eswaran C, Reza AW, Hati S. Extraction of the contours of optic disc and exudates based on marker-controlled watershed segmentation. In:Computer science and information technology, 2008. ICCSIT'08. International conference on; 2008. p. 719–23. IEEE.

[2]. Chràstek R, Wolf M, Donath K, Michelson G, Niemann H. Optic disc segmentation in ssretinal images. In: Bildverarbeitung Für die Medizin 2002: Springer; 2002. p. 263–6.

[3]. Abramoff MD, Garvin MK, Sonka M. Retinal imaging and image analysis. IEEE Rev Biomed Eng. 2010;3:169–208.

[4]. Liu, J., Wong, D., Lim, J., Jia, X., Yin, F., Li, H., Xiong, W., Wong, T.: Optic cup and disk extraction from retinal fundus images for determination of cup-todisc ratio. In: Industrial electronics and applications, 2008. ICIEA 2008. 3rd IEEE Conference On, pp.1828–1832 (2008). IEEE

[5]. Nyul LG. Retinal image analysis for automated glaucoma risk evaluation. Proc of SPIE Vol. 2009;7497:74971.

**Cite this article as :**

Sabah Samareen, P. Meghana Reddy Saniya Naaz, S. Surekha, " Direct Cup-To-Disc Ratio Estimation for Glaucoma Screening Via Semi-Supervised Learning, International Journal of Scientific Research in Science, Engineering and Technology(IJSRSET), Print ISSN : 2395-1990, Online ISSN : 2394-4099, Volume 10, Issue 3, pp.437-449, May-June-2023.