WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Cataract Detection Using Deep Learning Algorithms

V. Ujwala[1], B. Sridhar[1], Ch. Rashmitha[1], Y. Divakar[1], D. Harihara Santosh[2].

[1]Student, Department of Electronics and Communication Engineering, Aditya Institute of Technology and Management, Tekkali, Srikakulam, Andhra Pradesh, India, 532201.

[2]Associate Professor, Department of Electronics and Communication Engineering, Aditya Institute of Technology and Management, Tekkali, Srikakulam, Andhra Pradesh, India, 532001.

ABSTRACT

Cataract is a leading eye disease across the world. Visual impairment caused by cataracts is a commonly observed issue and blindness worldwide. There is around 50% of overall blindness. Cataract is a medical condition characterized by the clouding of the eye's lens. Therefore, an early detection and prevention of cataract may reduce the visual impairment and the blindness. For early prediction of cataract we use deep learning algorithms like CNN, MobileNet and VGG-16. Deep learning is a subset of Machine learning, which is essentially a neural network. CNN play a major role in diverse functions like image processing, computer vision tasks like segmentation and localization, they are very popular in Deep learning. MobileNet is a stream lined architecture that uses depthwise separable convolutional neural network. VGG-16 is object detection and classification algorithm, it is one of the popular algorithms for image classification. We use these three algorithms to predict which algorithm is more efficient in cataract detection.

Keywords : *CNN , MobileNet, VGG-16, Deep learning, Image processing.*

1. INTRODUCTION

According to the World Health Organization (WHO), there are approximately 285 million visually impaired individuals worldwide, with 39 million blind people and 246 million suffering from moderate to severe blindness. The majority of cataracts form as a result of age-related changes or injury to the tissue that constitutes the eye's lens. Proteins and fibres in the lens begin to break down, causing vision to become hazy or cloudy.

While cataracts typically form in both eyes, the rate of progression may vary between them, leading to a discrepancy in vision between the eyes, with one cataract being more advanced than the other. When you're young, the lens in your eye is clear. Around age 40, the proteins in the lens of your eye start to aggregate and coalesce into clusters. This clump makes a cloudy area on your lens or a cataract. Over time, the cataract gets more severe and clouds more belonging to the lens.

2. OBJECTIVE OF THE PROJECT

A cataract is the opacification of the typically transparent lens in the human eye, resulting in visual impairment that can interfere with daily activities and work. Individuals with cataracts may experience vision similar to peering through a frosted or foggy glass window, which can hinder tasks such as reading and driving. Although cataracts generally progress gradually, early on, they may not significantly impact vision. As time passes, cataracts can progressively disrupt vision. Traditional cataract examination tools and techniques can only be handled by highly skilled ophthalmologists, making it impractical to conduct mass screenings for early-stage cataract detection due to a shortage of ophthalmologists and the time-consuming nature of these procedures. Therefore, this proposal presents a practical and cost-efficient auxiliary diagnostic system.

This approach provides physicians with insights on the subsequent course of action. Deep learning techniques are employed to analyze images of the retina, which undergo pre-processing before being fed into the trained model. Having been trained on a vast number of images, the model compares the characteristics of the input images to those of the pre-existing images, and determines whether the individual in question has cataracts or not. Here we use three Deep learning algorithms namely CNN, MobileNet, VGG-16. The proposed system of our project shows which algorithm is more accurate in showing result among these three algorithms.

3. LITERATURE SURVEY

Hans Morales-Lopez - By timely detection, it is possible to prevent cataract surgery in the initial stage of it. In this work, we compare the main characteristics of different algorithms in grading and classification, going from the classical medical methods to the actuals based on computational intelligence.[1]

Isma Shaheen - To alleviate the burden of ophthalmologist, many researchers working in the field of biomedical imaging developed a number of techniques for the automatic detection and grading of cataract. Imaging modalities used for this purpose includes slit-lamp images, retro-illumination images, digital/optical eye images, retinal images, and ultrasonic Nakagami images.[2]

Linglin Zhang - To investigate the performance and efficiency by using Depp Convolutional Neural Network (DCNN) to detect and grad cataract automatically, it also visualize some of the feature maps at pool5 layer with their high-order empirical semantic meaning, providing a explanation to the feature representation extracted by DCNN.[6]

4. DEEP LEARNING

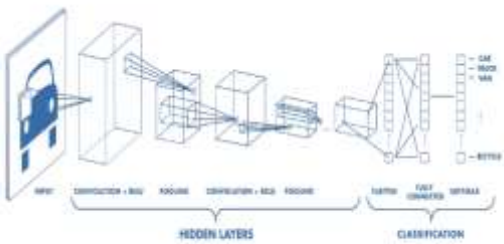
Deep learning, a subset of machine learning, relies entirely on artificial neural networks, which emulate the workings of the human brain. As a result, deep learning does not require explicit programming for every task. Although not novel, deep learning is a well-established concept. Deep learning has existed for a few years, but it is currently experiencing a surge in popularity due to the availability of substantial processing power and data, which was not the case previously. As processing power has increased exponentially over the last two decades, deep learning and machine learning have gained momentum. In essence, deep learning can be formally defined as a network of neurons. A formal definition of deep learning is- neurons.

A perceptron is a computational representation of a biological neuron that receives inputs and processes an output. Every input is assigned a weight, and the inputs are individually multiplied by their respective weights, summed, and passed through an activation function that determines if the neuron should activate and produce an output, similar to the workings of a biological neuron. Training a perceptron involves feeding it multiple training samples and calculating the output for each of them. After each sample, the weights are adjusted to minimize the output error, usually defined as the difference between the desired (target) and the actual outputs.

5. ALGORITHMS USED

A. CONVOLUTIONAL NEURAL NETWORK(CNN)

Convolutional neural networks (CNNs), a type of artificial neural network that has emerged as the leading method for various computer vision applications, is gaining attention in diverse fields, including radiology. CNNs are engineered to learn spatial hierarchies of features automatically and dynamically via backpropagation, using several building blocks such as convolution layers, pooling layers, and fully connected layers. CNN, which stands for convolutional neural network, is a type of deep learning neural network. In brief, CNN can be thought of as a machine learning algorithm that can receive an input image, assign significance (via learnable weights and biases) to different features/objects within the image, and distinguish between them. The operation of CNN involves identifying features within images. A CNN typically comprises of the input layer which is a grayscale image, the Output layer which is a binary or multi-class label, hidden layers consisting of convolutional layers, ReLU (rectified linear unit) layers, the pooling layers, and a fully connected Neural Network.



1. Convolutional layer

The convolution layer serves as the fundamental building block of CNN, carrying the majority of the network's computational burden. Its primary function is to perform a dot product between two matrices, namely the set of learnable parameters (known as a kernel) and the limited portion of the receptive field. As the kernel moves across the height and width of the image during the forward pass, it generates an image representation of the receptive region. This results in an activation map, which is a two-dimensional representation of the image showing the kernel's response at each spatial position. The size of the kernel as it slides is known as the sliding window size. It is called stride.

The formula for output operation of convolutional layer:

$$W_{out} = \frac{W - F + 2P}{S} + 1$$

2. Pooling layer

The role of the pooling layer is to summarize nearby outputs and replace certain output locations. This reduces the size of the representation and the amount of computation and weights required. The pooling operation is performed on each slice of the representation separately. Various types of pooling functions exist, including the average and L2 norm of the rectangular neighbourhood, and a weighted average based on the distance from the central pixel. Nevertheless, the most commonly used pooling method is max pooling, which reports the maximum output from the neighbourhood.

The formula for output operation of padding:

$$W_{out} = \frac{W - F}{S} + 1$$

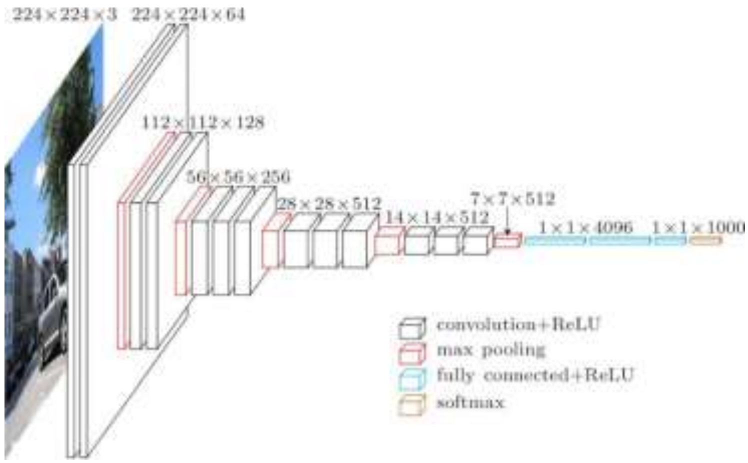
3. Fully connected layer

The fully connected layer, like a regular fully connected neural network, allows neurons in this layer to have complete connections with all neurons in the previous and following layer. As a result, it can be calculated in the same way as a matrix multiplication followed by a bias effect. The purpose of the FC layer is to facilitate the mapping of the representation between the input and output of the neural network to compute the output. The FC layer is where image classification happens in the CNN based on the features extracted in the previous layers. Here, fully connected means that all the inputs or nodes from one layer are connected to every activation unit or node of the next layer. All the layers in the CNN are not fully connected because it would result in an unnecessarily dense network. It also would increase losses and affect the output quality, and it would be computationally expensive.

B. VGG – 16

The full name of VGG is the Visual Geometry Group, which belongs to the Department of Science and Engineering of Oxford University. The input of VGG is set to an RGB image of 224x244 size. The average RGB value is calculated for all images on the training set image, and then the image is input as an input to the VGG convolution network. A fixed-size RGB image of 224 x 224 is fed as input to the cov1 layer. The image then undergoes a series of convolutions where filters are applied with a small receptive field of 3x3. This size is enough to capture the basic directions of left/right, up/down, and center. Additionally, the network configuration also includes the use of 1x1 convolution filters, which serve as a linear transformation of the input channels, followed by a non-linear activation function.

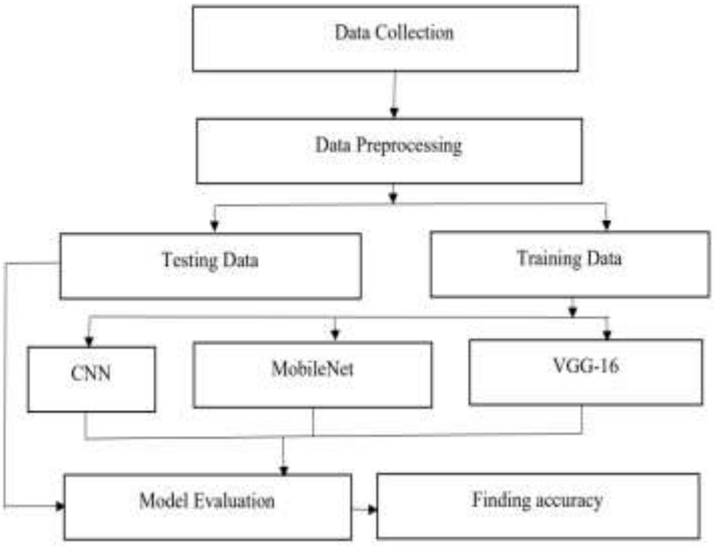
During the forward pass, the convolutional stride is fixed to 1 pixel, and the spatial padding of the convolutional layer input is chosen so that the spatial resolution is preserved after convolution, which means that for 3×3 conv. layers, the padding is 1-pixel. Spatial pooling is carried out by five max-pooling layers which follow some of the convolutional layers; not all convolutional layers are followed by max-pooling. The max-pooling is done over a 2×2 pixel window with stride 2. The ReLU non-linearity is applied to all the hidden layers. It should be noted that, except for one, the networks do not include Local Response Normalisation (LRN). The reason being that this normalization technique does not enhance performance on the ILSVRC dataset, but instead increases both memory usage and computation time.



C. MOBILENET

MobileNet is an Image Classification and Mobile Vision CNN architecture model that stands out from other models due to its low computational requirements, making it ideal for use in embedded systems, mobile devices, and computers with limited computational efficiency. Despite its efficiency, it does not significantly compromise accuracy of the results. It is also best suited for web browsers as browsers have limitation over computation, graphic processing and storage. MobileNet is a type of Convolution Neural Network architecture, this refers to the usage of depth wise separable convolution. layers instead of standard convolution layers. Depth wise separable convolutions is nothing but the depth wise convolutions followed by point wise convolutions. It is a light weight deep neural network and the complexity is also less. This architecture reduces computation cost. Therefore it is well suited for mobile and embedded applications. MobileNet is TensorFlow's first mobile computer vision model. It uses depth wise separable convolutions to significantly reduce the number of parameters compared to other networks with regular convolutions and the same depth in the nets

6. METHODOLOGY



7. WORKING OF PROJECT

In this project, firstly we took data set from ImageNet and divided images needed for our cataract detection. We installed certain modules needed for implementation of the project namely Tensor flow, OpenCV python, NumPy, Pandas, Keras and some other modules required as per code. Here, we used 'tqdm' function for graph bar improvement, Pandas for pre-processing steps, 'Matplotlib' function for visualization improvement. We also used NumPy for mathematical operations, Seaborn for confusion matrix formation, 'sequential' function to take images from data set in sequential order. RMSprop, Adam are the optimizers we used to optimize images, sklearn is used to train our dataset. We took both left and right eye images from our dataset and took random images among those for prediction. The prediction process involves utilizing a prediction function to make inferences. Initially, images are read using the imread function from OpenCV, followed by pre-processing steps that involve converting the image from BGR to RGB and resizing it. Subsequently, the pre-processed images are transformed into arrays and normalized by dividing each pixel value by 255. Finally, the predict function is applied to obtain the regression output from the model. We applied CNN, VGG-16 AND MobileNet algorithms individually and obtained a confusion matrix, training and testing accuracy graphs for all three algorithms applied.

Accuracy calculated from confusion matrix obtained for each algorithm as :

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

8. RESULTS

We obtained Confusion matrix, accuracy plot and loss plot for all three algorithms individually.

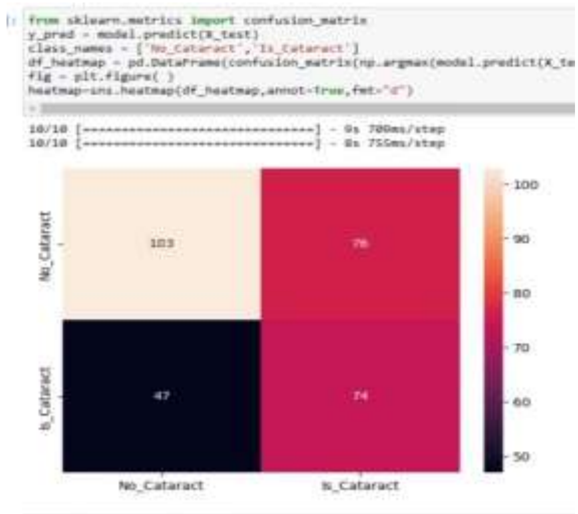


Fig : Confusion matrix of CNN

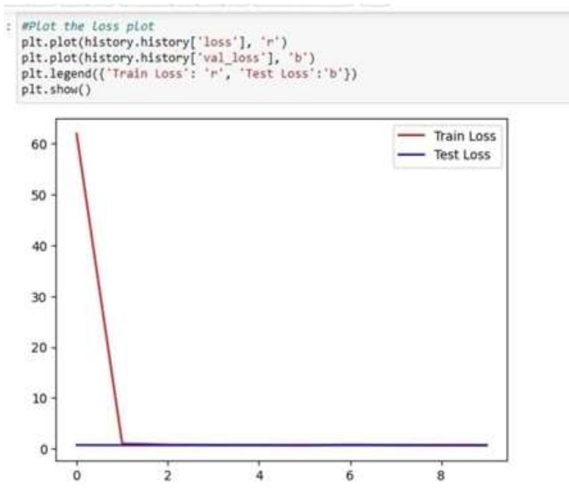


Fig : Loss plot of CNN

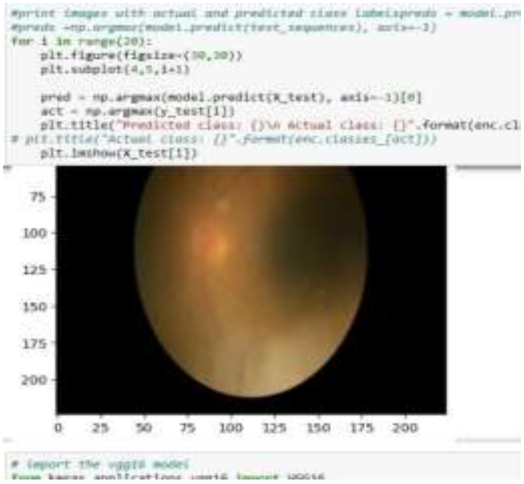


Fig : Output images of CNN

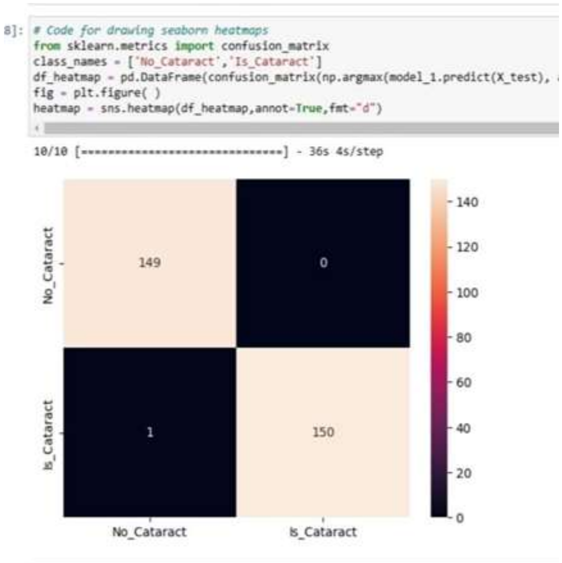


Fig : Confusion matrix of VGG-16

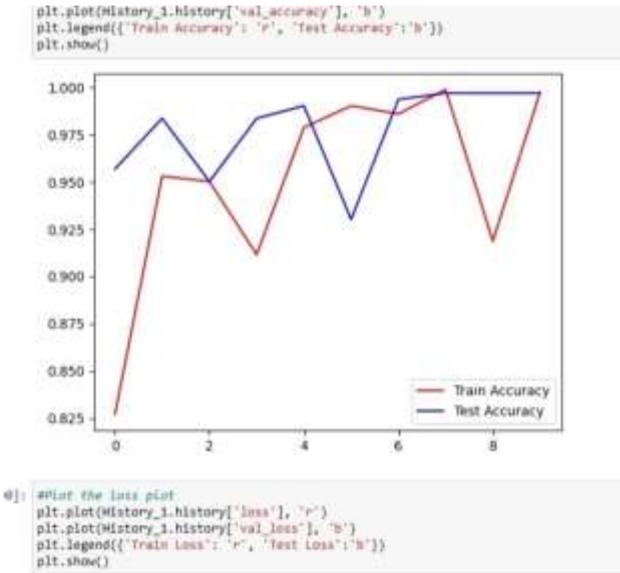


Fig : Accuracy plot of VGG-16

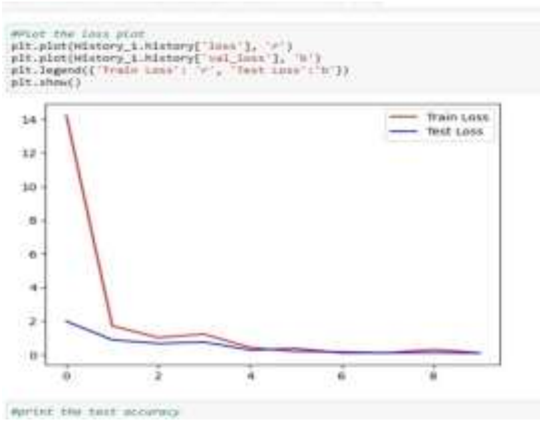


Fig : Loss plot of VGG-16

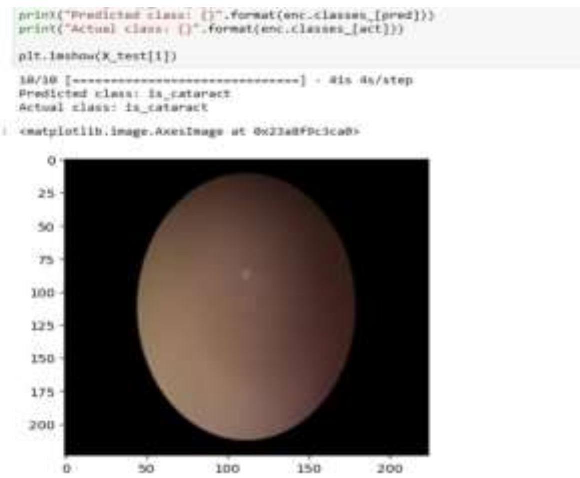


Fig : Output images of VGG-16

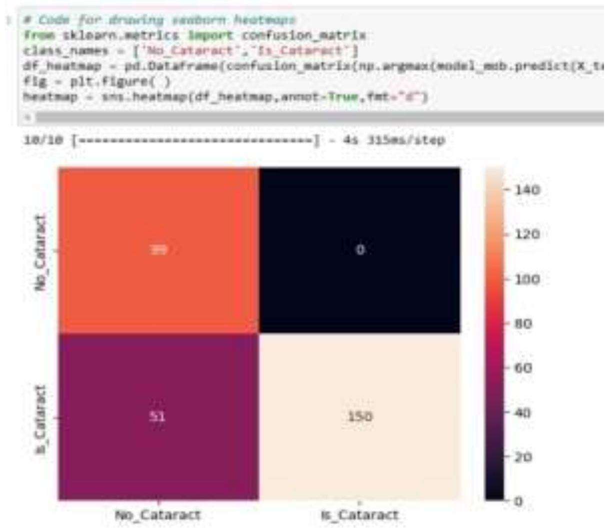


Fig : Confusion matrix of MobileNet

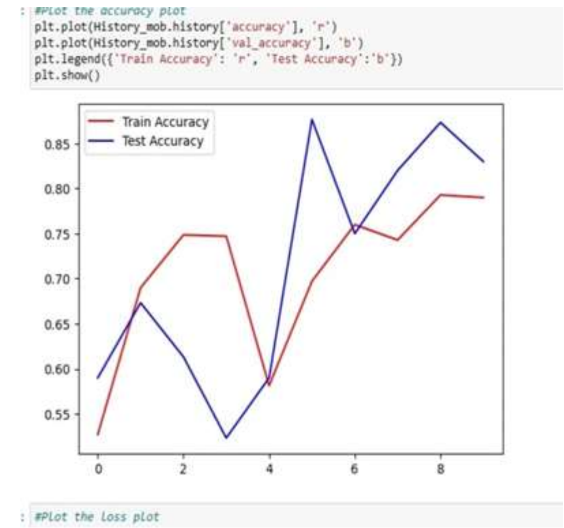


Fig : Loss plot of MobileNet

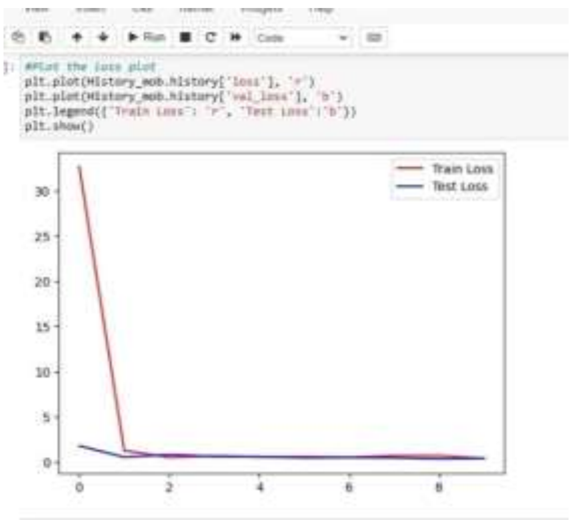


Fig : Accuracy plot of MobileNet

Fig : Output images of MobileNet

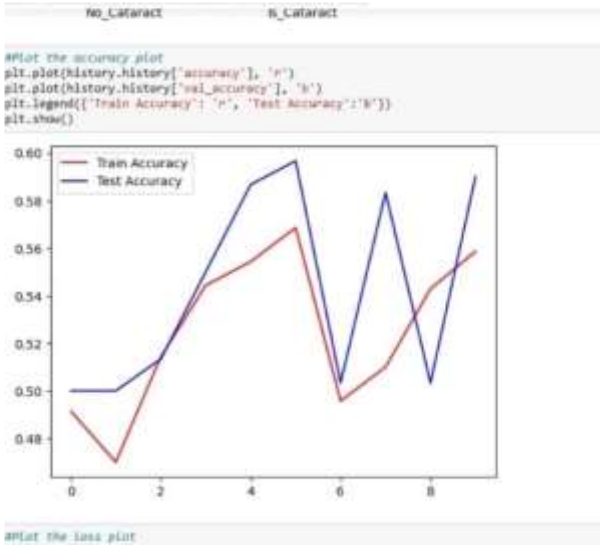


Fig : Accuracy plot of CNN

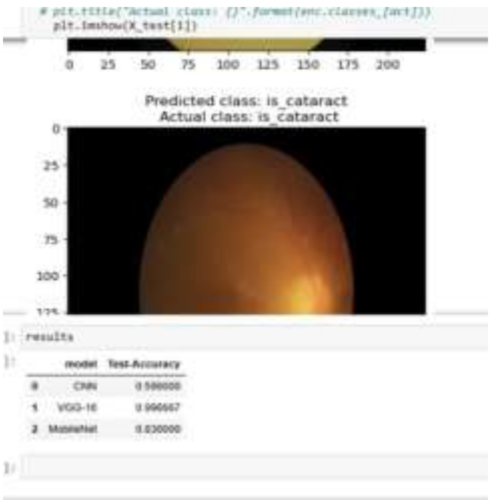


Fig : Final accuracy percentage of three algorithms

9. CONCLUSION

By using this model/system, an individual can determine if they have cataracts or not by providing their retina images. Traditional cataract testing methods can be time-consuming and expensive when compared to the usage of this model, which offers a fast and convenient solution. The model features an end-to-end structure, eliminating the need for manual feature extraction and providing a fully automated process. We applied CNN, VGG-16 AND MobileNet algorithms individually and obtained a confusion matrix, training and testing accuracy graphs for all three algorithms applied. Finally, after getting confusion matrix and training testing accuracy graphs, we got images which represents whether given actual class and derived class is same for how many images within a range of 20 images taken. Based on confusion matrix result i.e. images output we got final test accuracy of three algorithms.

10. REFERENCES

[1] Hans Morales-Lopez, Israel Cruz-Vega & Jose Rangel-Magdeleno., "Cataract Detection and Classification Systems Using Computational Intelligence" , year 2020.

[2] Isma Shaheen and Anam Tariq., "Survey Analysis of Automatic Detection and Grading of Cataract Using Different Imaging Modalities", year 2018.

[3] Md Kamrul Hasan, Tanjum Tanha, Md Ruhul Amin, Omar Faruk, Sultan Aljahdali and Mehedi Masud., "Cataract Disease Detection by Using Transfer Learning-Based Intelligent Methods Md Kamrul Hasan", year 2021.

[4] Yanyan Dong; Qinyan Zhang; Zhiqiang Qiao; Ji-Jiang Yang., "Classification of cataract fundus image based on deep learning", year 2017.

[5] Masum Shah Junayed, MD Baharul Islam, Arezoo Sadeghzadeh, and Saimunur Rahman., "An Automated Cataract Detection System Using Deep Learning for Fundus images", September 24, 2021.

[6] Linglin Zhang, Jianqiang Li, I Zhang, He Han, Bo Liu, Jijiang Yang, Qing Wang., "Automatic cataract detection and grading using deep convolutional neural network", published by IEEE, year 2017

[7] Turimerla Pratap, Priyanka Kokil, "Computer-aided diagnosis of cataract using deep transfer learning", volume 53, published in August 2019.

[8] Md. Sajjad Mahmud Khan, Mahiuddin Ahmed, Raseduz Zaman Rasel, Mohammad Monirujjaman Khan, "Cataract Detection Using Convolutional Neural Network with VGG-19 Model", published in 10 May 2021.

[9] Chi-Ju Lai, Ping-Feng Pai, Marvin Marvin, Hsiao-Han Hung, Si-Han Wang, Din-Nan Chen, "The Use of Convolutional Neural Networks and Digital Camera Images in Cataract Detection", published in 11 March 2022.

[10] Turimerla Pratap, Priyanka Kokil, "Deep neural network based robust computer-aided cataract diagnosis system using fundus retinal images", volume 70, published in September 2021.