ISSN No:-2456-2165

Comparative Study of Image Classification Algorithms for Eyes Diseases Diagnostic

Mahmoud Smaida
The National University of Water and Environmental Engineering
Revine, Ukraine

Dr. Yaroshchak Serhii
Applied Mathematics
The National University of Water and Environmental Engineering
Revine, Ukraine

Abstract:- Deep learning is the most technology in 21century, it gives more information about how computers can understand data and learning from. In deep learning, networks of artificial neurons analyse large dataset to automatically discover patterns. In this paper, we will introduce the part of these techniques to know how we can use deep learning to create our own model to diagnosis eye diseases. The most idea will be addressed is the evaluation performance model using confusion matrix. In this study, we will compare three models of neural network, CNN, Vgg16 and Inceptionv3 in order to evaluate performance of the models.

In 0ur work, a deep learning convolutional network based on keras and tensorflow is deployed using python for image classification. a number of different images, which contains four types of eye diseases, namely Diabetic retinopathy, Glaucoma, Myopia and Normal are used for image classification. Three different structures of neural network, CNN, VGG16 and Inception V3 are compared on GPU system in Google Colab, with three different combinations of classifiers. It is shown that, the results for each combination and observed that for multi-image classification, Inception V3 combination gives better classification accuracy (81.00 %) than any other models. Using of confusion matrix showing us where our classifier is confused when it makes prediction.

***Keywords**:- Inception V3, CNN, Vgg16, Eye Diseases, Confusing Matrix, Deep Learning, Diabetic Retinopathy, Glaucoma, Myopia.*

I. INTRODUCTION:

One of the most important features of the use of neural network are confusion matrix. It is not specific to a neural network but it is applicable in general to any classification algorithm.

Confusion matrix basically gives us an idea about how well our classifier has performed with respect to performance on individual classes. It is also identified as a performance measurement technique for Machine learning classification. typically, a confusion matrix is filled up based on the test set whose true labels is known. The test data is passed through the classifier and predictions are noted. A table of predicted labels vs true labels is then filled out.

Confusion matrixes are important because it tells how accurate a model's outcomes, evaluate the performance of a classification model and allowing developers to determine which data their model may be unable to classify correctly.

In this study, we need to know the performance of the classification model on a test dataset to see the actual values. Classification accuracy alone can be misleading if you have an unequal number of observations in each class or if you have more than two classes in your dataset. Computing the confusion matrix can give you a better idea of how well your classification model works and what mistakes it makes.

The name of our project is Development of Neural Network Algorithms for Automation of Early Diagnostics of Eye Diseases (GMDsystem), it is a neural network expert system to assist ophthalmologist in medical frontlines to diagnose early eye diseases (Glaucoma, Myopia and Diabetic retinopathy). In disease detection, for example, in our work we have four diseases for classification, the accuracy result could be high or low, it depends on the model. So, we need to determine which data our model may be unable to classify correctly.

The first step in evaluating a classification is to check confusion matrix. Indeed, a number of statistical models and accuracy measures are based on confusion matrix. digital colour fundus photographs of the retina as it shown in Fig 1, the focus will be on the features of the images to detect eye diseases [1]. using convolution neural network CNN, VGG16 architecture and InceptionV3.

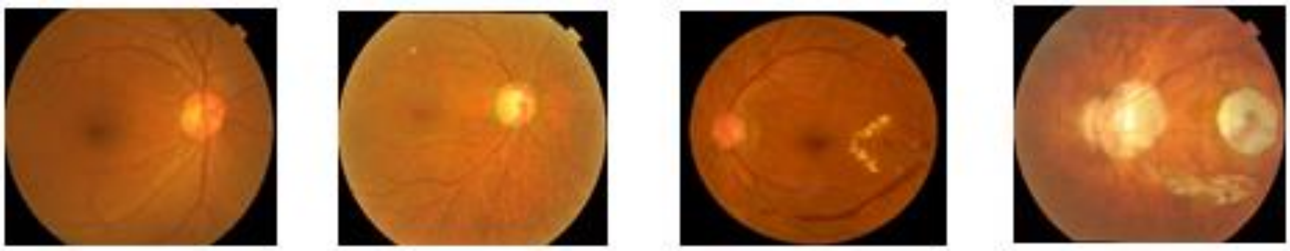


Fig. 1:- Normal Fundus, Glaucoma, Myopia and Diabetic retinopathy

II. PROBLEM FORMULATION:

Eye diseases have a large variety of forms, the textures are sometimes hard to be identify and recognized by optometrist. Therefore, using information technology to improve current system must be made to provide maximum comfort to the patient/optometrist and improve health care. In this paper we will using the confusion matrix to evaluate three different CNN architecture to determine in which disease our model unable to classify correctly.

III. RELATED WORK

Many researchers have suggested their work using neural network Most of these studies have been conducted recently, focusing on modern research and facts. Review of few is as below:

Labatut, Vincent, and Hocine Cherifi. [2] the authors in this work, reviewed the main measures used to assess accuracy from different classification. They consider the case where a person wants to compare different classification algorithms, checking them on a specific sample of data to determine which will be best for the sample population. The authors considered three factors: changes in error level, proportions and number of classes. The authors then compared the scale from a more theoretical point of view. In the case studied here, it turns out that some characteristics of the measurements are not related to their differentiation. First, all monotonically related measures are similar to their work, Second, their range is a little importance, Finally, the complex measures are difficult or impossible to interpret correctly. In these conditions, the authors advise the user to choose the simplest measures, the interpretation of which is simple. They recommend using both TPR and PPV, or a combination of content such as the F-measure.

Visa, Sofia, et al. [3] Authors introduce new technologies for selecting functions and demonstrate them in a real data set. The proposed system creates subsets of attributes based on two criteria: (1) individual traits are characterized by a high difference (classification); (2) the features of the subgroup complement each other, that is, they mix different categories. The method uses information from the confusion matrix and evaluates one attribute at a time.

Nezami, Omid Mohamad, et al. [4] This paper presents a deep learning model to improve engagement recognition from images using pre-training on available basic facial expression data, before training on specialised engagement data. In the first of two steps, a facial expression recognition model is trained to provide a rich face representation using deep learning. In the second step, the authors used the model's weights to initialize their deep learning-based model to recognize engagement. The model was trained on new engagement recognition dataset with 4627 engaged and disengaged samples. The results were the engagement model outperforms effective deep learning architectures that the authors applied for the first time to engagement recognition.

Loussaief, Sehla, and Afef Abdelkrim. [5] The authors used different techniques and algorithms in machine learning framework for image classification. They introduced the Bag of Features paradigm used for input image encoding and highlighted the SURF as its technique for image features extraction. Confusion matrix was applied to evaluate the works. Through experimentations they proofed that using SURF local feature extractor method for image vector representation and SVM (cubic SVM) training classifier performs best prediction average accuracy.

Bizios, Dimitrios, et al. [6] In this work they compared the performance between two methods of machine learning classifiers, support vector machine (SVM) and artificial neural networks (ANN) based on measurements of the thickness of the layer of the retinal nerve fiber (RNFLT) using optical coherence tomography (OCT), for the diagnosis of glaucoma.

The result was similar between ANN and SVM in this study. Both machine learning classifiers worked very well, with similar diagnostic performance. Input parameters have a greater impact on diagnostic performance than the type of machine classifier. the results show that parameters based on A-scan thickness measurements converted using RNFL processed by machine classifiers can improve the diagnosis of OCT-based glaucoma.

IV. DATA DESCRIPTION

Kaggle: A data science site that contains many interesting data sets from the outside. On his main list, you can find all kinds of niche datasets, from ramen ratings to basketball data and pet licenses in Seattle. [11].

I Challenge-GON Comprehension: Large dataset of 1200 annotated retinal fundus images of subjects without glaucoma (90%) and patients with glaucoma (10%).

The dataset comprises more than 35 breeds of eye diseases. To make it simpler, we'll reduce the dataset with the 4 main breeds. The dataset is comprised of photos of Glaucoma, Myopia, Diabetic retinopathy and Normal eye provided as a subset of photos from a large dataset of 955 Retinal Image. all the images were collected in total from Kaggle dataset, In high resolution images.

The images will be the input of our CNN. We are provided a training set and a test set of images of eye diseases. Each type of images has individual folder and each image has a filename that is its unique id.

Python language will be used to achieve our goal in google colab environment.

Glaucoma	Myopia	Diabetic retinopathy	
161	54	180	560

Table 1:- Number of images according to eye diseases

V. RESEARCH METHODOLOGY

The block diagram of the three proposed methodologies is shown in fig.2 and fig.3. Each block of proposed flow diagram is clearly labelled and represents processing steps. Using these methodologies, we compare three different structure of CNN, VGG 16 and inception V3 in order to evaluation using confusion matrix.

Firstly, step image dataset is prepared, there are 4 files in dataset, which contains 955 images of Diabetic retinopathy, Glaucoma, Myopia and Normal, where 955 images used for training and 190 images used for testing purpose. In the next steps, we fit the CNN created to the image data set and train, test the system with training and test data sets, respectively. Finally, we get the accuracy for different CNN structures and compare these accuracies to measure performance, and then we obtain the resulting CNN structures.

Three methods are studied in this paper in order to evaluate our classifier using confusion matrix:

- The CNN consists of three hidden layers and pooling layers occurring in an alternating fashion.
- Pre-trained CNNs based VGG 16 algorithms using the last block layer training (Block 5).
- Pre-trained CNNs based Inception v3 algorithms using the last block layer training ('mixed6).

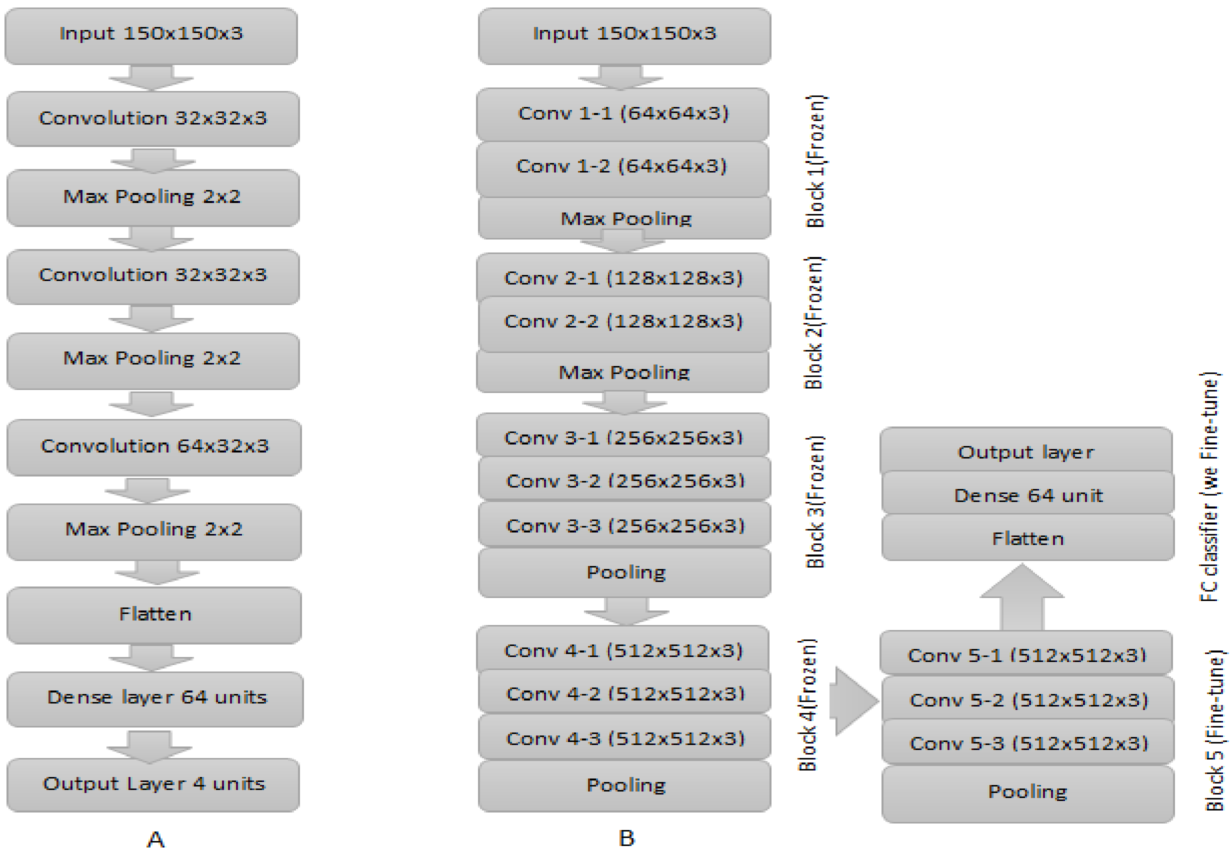


Fig.2:- The block diagram of CNN and VGG 16

A. Convolution Neural Network:

As the fig.2 shown for convolutional layer, the size of input image is set to 150*150 pixels with 3 channels (RGB). To extract the features from the image we use 32 filters of size 3*3 pixels. For pooling layer, we use a window of size 2*2 pixels, which used to compress the original image size for further processing. After that we use another convolution layer used 32 filters with size 3*3 and max pooling size 2*2. The last convolution layers are used 64 filter size 3*3 with max pooling size 2*2. And then we use fully connection (Dense 64 units) and output layer (4 unit) for predict the eye diseases. CNNs adjust their filter weights through backpropagation, which means that after the forward pass, the network is able to look at the loss function and make a backward pass to update the weights.

In experiment, we use confusion matrix to evaluate our work, and analyze that which combination gives better classification accuracy for eye disease classification.

B. VGG 16:

It is a convolutional neural network structure developed by Visual Geometry Group from oxford university in 2014. This model loads a set of weights pre-trained on ImageNet used 16-layer network.

The size of the input images in VGG16 network are 224x224 RGB, Images are passed through 5 blocks of convolutional layers, where each block consists of an increasing number of 3x3 filters. The stride is fixed to 1 while the convolutional layer inputs are padded. Blocks are separated by maximum pooling layers. Maximum pooling is done over 2*2 windows with stride 2. the five blocks of convolutional layers are followed by three fully connected layers (FC). The last layer is a soft-max layer that represented the output layer. The full form is shown in Fig.2 [12].

C. Inception V3 (GoogleNet):

Inception-v3 is a convolutional neural network (CNN) which has 48 deep layers that trained on more than a million images from the ImageNet database. It can classify images into 1,000 categories of objects [13].

Inception-v3 is one of the most famous models can be used for transfer learning, it is allowing to retrain the final layers of existing model, resulting in a significant decrease time training and the time the size of the dataset required. As mentioned above inception-v3 trained on more than million images from the ImageNet database, which means you can maintain the knowledge that the model had learned during its original training and apply it to smaller dataset, the resulting in highly accurate classifications without the need of training all the model and computational power.

Inception Layer as the Fig.3 show is a combination of set of layers (namely, 1×1 Convolutional layer, 3×3 Convolutional layer, 5×5 Convolutional layer) with their output filter banks concatenated into a single output vector forming the input of the next stage [14].

In addition to the layers mentioned above, there are some important points in the original inception layer:

- 1×1 Convolutional layer before applying another layer, which is mainly used for dimensionality reduction
- Parallel Max-Pooling layer, which provides another option to the inception layer.

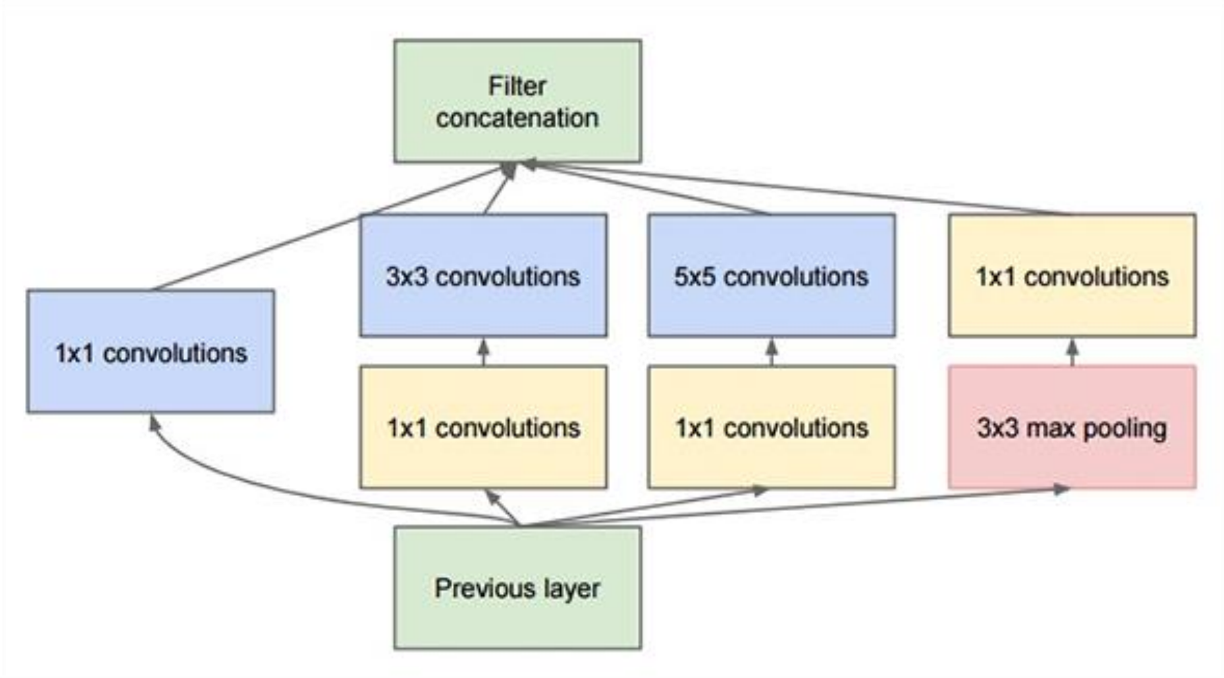
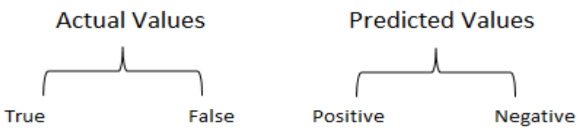


Fig. 3:- The block diagram of Inception V3

VI. SELECTED MEASURES

In this section, we formally describe the most common measures used to compare classifiers. various measures based on the result of the confusion matrix. In this article, the comparison will be done using the confusion matrix to measure the model's Recall, Precision, Accuracy and F-measure.

Let us understand TP, FP, FN, TN for two class classification:
True Positive (TP): Your predicted positive and it is true.
True Negative (TN): Your predicted negative and it is true.
False Positive (FP): your predicted positive but it is false.
False Negative (FN): You predicted negative but it is false.
We describe predicted values as Positive and Negative. And True and False as actual values.



How to Calculate Confusion Matrix for a 2-class classification problem:

Recall or Sensitivity: Is a measure of completeness or quantity. simply, high recall means that an algorithm returned most of the relevant results.

$$Recall = \frac{TP}{TP + FN}$$

Precision: Is a measure of exactness or quality, high precision means that an algorithm returned substantially more relevant results than irrelevant ones.

$$Precision = \frac{TP}{TP + FP}$$

Specificity: corresponds to the true negative rate of the considered class.

Specificity=TN/TN+FP

Accuracy: Is a measure of how much we predicted the classes correctly.

	Class1 Predicted	Class2 Predicted
Class1 Actual	TP	FP
Class2 Actual	FN	TN

Accuracy = TP / TP+FP+FN+TN

F-measure: F-score helps to measure Recall and Precision at the same time to make them comparable.

$$F\text{-}measure = \frac{2*Recall*Precision}{Recall + Precision}$$

VII. EXPERIMENTS AND RESULTS

According to the models that we explained above, all these models are implemented using python language in google Colab environments, and applied eye diseases as a dataset (Diabetic Retinopathy, Glaucoma, Myopia and Normal) for classification.
In our experiments, we use four categories of classification (eye diseases) according to the models above in order to compare them in accuracy using the confusion matrix to obtain the best model for the detection of eye diseases, and determine in any class the models confuse.

A. Results on CNN:
In this model, The CNN structure has been applied as it shown in Fig.2 A, with eye diseases dataset and the results reported as show in Table.2 below;

	Predicted					
Actual		Diabetic Retinopathy	Glaucoma	Myopia	Normal	Total
	Diabetic Retinopathy	10	0	19	7	36
	Glaucoma	0	0	18	14	32
	Myopia	0	0	10	0	10
	Normal	0	0	31	81	112
	Total	10	0	78	102	

Table.2:- CNN confusion matrix

Fig. 4:- CNN confusion matrix

Here, how to read this matrix:

- The total number of test example of any class would be the sum of corresponding row (**i.e. the TP+FN for that class**).
- The total number of FN for a class is sum of value of the corresponding row (**excluding the TP**).
- The total number of FP for a class is sum of value of the corresponding column (**excluding the TP**).
- The total number of TN for a certain class will be the sum of all columns and rows **(excluding that class's column and row).**

All results shown in Table.5.

B. Results on VGG16:

The structure of VGG16 as it shown in Feg.2 B has been applied with fine-tune the final layers (Block 5+Fully connected). The results reported as the following in Table.3:

		Predicted				
Actual		Diabetic Retinopathy	Glaucoma	Myopia	Normal	Total
	Diabetic Retinopathy	35	0	1	0	36
	Glaucoma	12	10	1	9	32
	Myopia	2	0	8	0	10
	Normal	8	6	0	98	112
	Total	57	16	10	107	190

Table 3:- VGG confusion matrix

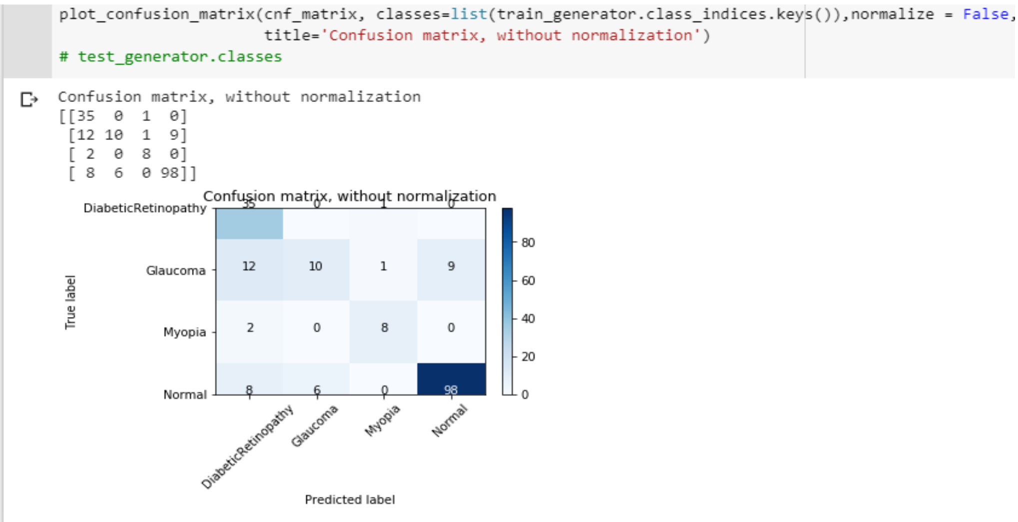


Fig. 5:- VGG confusion matrix

The result shown in Table.5.

C. Results on Inception v3:

As shown in Fig.3, we applied the structure of inception v3 with pre-training and the results documented as the following in table.4:

		Predicted				
Actual		Diabetic Retinopathy	Glaucoma	Myopia	Normal	Total
	Diabetic Retinopathy	33	0	2	1	36
	Glaucoma	8	9	4	11	32
	Myopia	0	0	10	0	10
	Normal	10	0	0	102	112
	Total	51	9	16	114	190

Table.4:- InceptionV3 confusion matrix

Fig.6:- Inception v3 confusion matrix

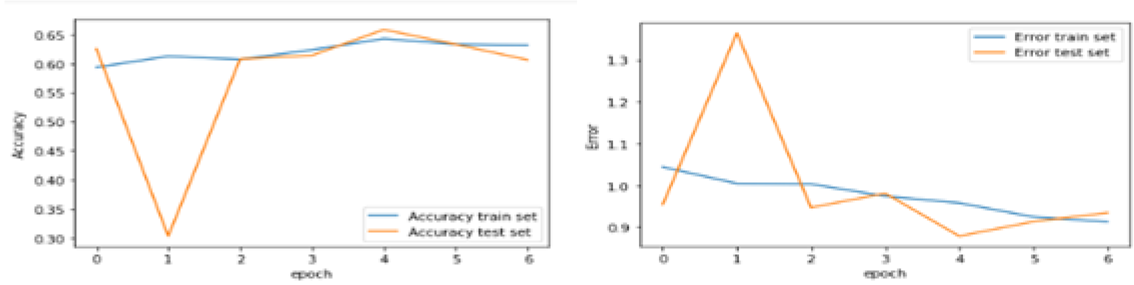
The result shown in Table.5.

VIII. RESULTS

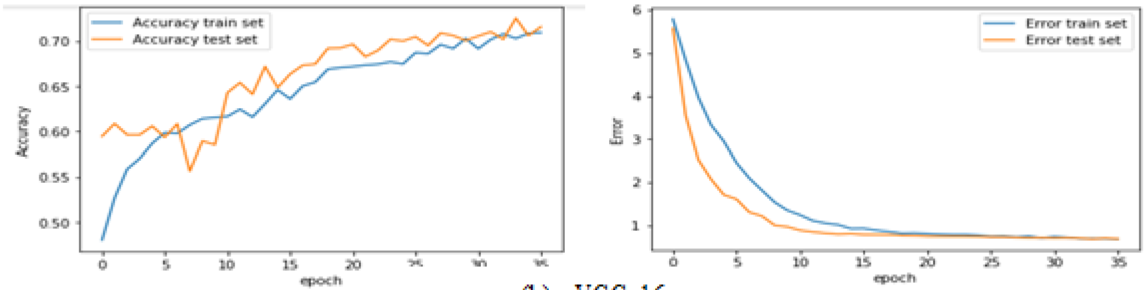
There are three classification accuracies obtained (as shown in table 5) from above models, and these accuracies are graphically represented in below graphs (a, b, c), where each model structure shown with epochs and accuracies.

	Recall				Precision				Specificity				F-Score				Accuracy
Models	Diabetic Retinopathy	Glaucoma	Myopia	Normal	Diabetic Retinopathy	Glaucoma	Myopia	Normal	Diabetic Retinopathy	Glaucoma	Myopia	Normal	Diabetic Retinopathy	Glaucoma	Myopia	Normal	Overall
CNN	0.277	0.000	1.000	0.723	1.000	0.000	0.128	0.794	1.000	1.000	0.622	0.764	0.433	0.000	0.226	0.756	53.1 %
VGG16	0.972	0.625	0.800	0.875	0.614	0.312	0,800	0.915	0.857	0.962	0.980	0.884	0.753	0.416	0.800	0.895	79.4 %
InceptionV3	0.916	0.281	1.000	0.910	0.647	1.000	0.625	0.894	0.883	1.000	0.966	0.846	0.758	0.438	0.769	0.901	**81.0 %**

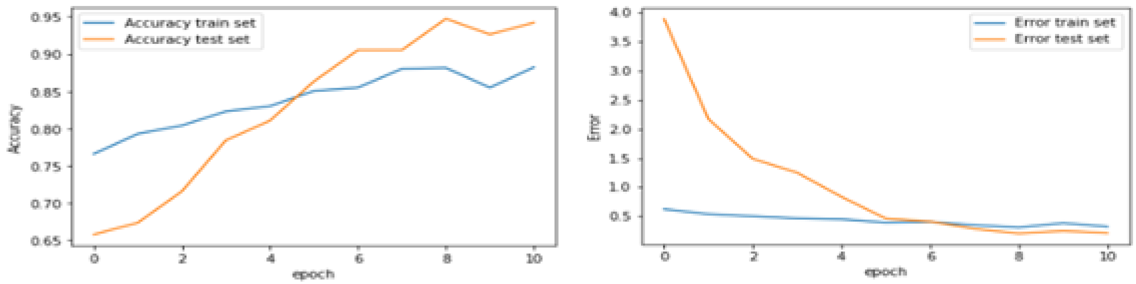
Table.5:- Obtained accuracies with different combinations of confusion matrix



(a) CNN



(b) VGG 16



(c) InceptionV3

Fig. 7

We compare accuracies of graph a, b and c, and we find out the following:

- Inception v3 with Fine-tune (graph c) gives better accuracy 81.00 %, which is far better than accuracies of graph a (53.1 %) and graph b (79.4 %).
- The confusion matrix shows that all classification model is confused with Glaucoma when it makes prediction. Therefore, this problem must be addressed to optimize the classification.

IX. CONCLUSION

Deep learning is a learning method for data analysis and predictions, now days it also become very popular for image classification problems, in this paper we have presented three methods for multi-class classification and we found that the deep neural network models can outperform traditional methods that rely on image classification.

We have compared between three models of multiclass classification CNN, VGG16 and Inception V3 in order to measure the accuracy using confusion matrix, to know where exactly the classifier confuse. Due to the small number of the training datasets (eye diseases), we implemented the Fine-tuning and data augmentation to increase the accuracy of experiments in the test set. All the models mentioned above are deployed using python for multiclass image classification. In this study, we compared these three different structures of CNN on GPU system using google Colab. With experiments, we obtained results for each combination and observed that for multi-image classification, Inception V3 combination gives better classification accuracy (81.00 %) than any other models as it shown in table.5. So, the using of confusion matrix shows all classification models in varying proportions are confused with Glaucoma when it makes prediction.

REFERENCES

[1]. American Macular Degeneration Foundation www.macular.org, 2019.

[2]. Labatut, Vincent, and Hocine Cherifi. "Evaluation of performance measures for classifiers comparison." arXiv preprint arXiv:1112.4133 (2011).

[3]. Visa, Sofia, et al. "Confusion Matrix-based Feature Selection." *MAICS* 710 (2011): 120-127.

[4]. Nezami, Omid Mohamad, et al. "Automatic Recognition of Student Engagement using Deep Learning and Facial Expression." *arXiv preprint arXiv:1808.02324* (2018).

[5]. Loussaief, Sehla, and Afef Abdelkrim. "Machine learning framework for image classification." *2016 7th International Conference on Sciences of*

Electronics, Technologies of Information and Telecommunications (SETIT). IEEE, 2016.

[6]. Bizios, Dimitrios, et al. "Machine learning classifiers for glaucoma diagnosis based n classification of retinal nerve fibre layer thickness parameters measured by Stratus OCT." Acta ophthalmologica 88.1 (2010): 44-52.

[7]. LeCun, Y., Bengio, Y., et al. (1995). Convolutional networksforimages,speech,andtimeseries. Thehandbook of brain theory and neural networks, 3361(10):1995.

[8]. Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105.

[9]. Graves, A., Mohamed, A.-r., and Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In Acoustics, speech and signal processing (icassp), 2013 ieee international conference on, pages 6645– 6649. IEEE.

[10]. Kim, Y. (2014). Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.

[11]. Open Dataset Finders, https://lionbridge.ai/datasets/the-50-best-free-datasets-for-machine-learning/,2019

[12]. Tindall, Lucas, Cuong Luong, and Andrew Saad. "Plankton classification using vgg16 network." (2015).

[13]. Szegedy, Christian, et al. "Going deeper with convolutions." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

[14]. *ImageNet*. http://www.image-net.org/,11/2019

[15]. **Understanding Confusion Matrix.** https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62. 11/2019

[16]. Priss, Uta. *Proceedings of the Tenth Midwest Artificial Intelligence and Cognitive Science Conference*. AAAI Press, 1999.

[17]. Chaudhari, Ms Pooja, and Girish A. Kulkarni. "Using Artificial Neural Network to Detect Glaucoma with the Help of Cup to Disk Ratio."

[18]. Somfai, Gábor Márk, et al. "Automated classifiers for early detection and diagnosis of retinopathy in diabetic eyes." BMC bioinformatics 15.1 (2014): 106.

[19]. Silberman, Nathan, et al. "Case for Automated Detection of Diabetic Retinopathy." AAAI Spring Symposium: Artificial Intelligence for Development. 2010.

[20]. Priya, R., and P. Aruna. "SVM and neural network based diagnosis of diabetic retinopathy." International Journal of Computer Applications 41.1 (2012).

[21]. Luculescu, Marius Cristian, and Simona Lache. "Computer-aided diagnosis system for retinal diseases in medical imaging." WSE AS Trans Syst 7 (2008): 264-276.

[22]. Köhler, Thomas, et al. "Computer-aided diagnostics and pattern recognition: Automated glaucoma detection." Teleophthalmology in Preventive Medicine. Springer, Berlin, Heidelberg, 2015. 93-104.

[23]. Xu, Kele, Dawei Feng, and Haibo Mi. "Deep Convolutional Neural Network-Based Early Automated Detection of Diabetic Retinopathy Using Fundus Image." Molecules 22.12 (2017): 2054.

[24]. Adrian Rosebrock, "Keras Conv2D and Convolutional Layers", https://www.pyimagesearch.com/2018/12/31/keras-conv2d-and- convolutional-layers/, on December 31, 2018.

[25]. Maher, Raju, et al. "Automated detection of diabetic retinopathy in fundus images." International Journal of Emerging Research in Management and Technology 4.11 (2015): 137-145.

[26]. B.J. Erickson, B. Bartholmai, Computer-Aided Detection and Diagnosis at the Start of the Third Millennium, Journal of Digital Imaging, nr. 15, pag. 59–68, 2002.

[27]. M. Luculescu, S. Lache, MaculaTEST – Computer-Aided Diagnosis System for Macular Diseases, The 6th WSEAS International Conference on COMPUTATIONAL INTELLIGENCE, MANMACHINE SYSTEMS and CYBERNETICS (CIMMACS '07), Puerto De La Cruz, Tenerife, Canary Islands, Spain, December 14-16, 2007,pp.77-82.

[28]. Attia, Mohammed, et al. "Multilingual multi-class sentiment classification using convolutional neural networks." *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

[29]. Pavithra, G., and T. C. Manjunath. "Design of algorithms for diagnosis of primary glaucoma through estimation of CDR in different types of fundus images using IP techniques." Int. J. Innov. Res. Inf. Secur.(IJIRIS) 4.5 (2017): 12-19.

[30]. Kumar, P. S. J., and Sukanya Banerjee. "A survey on image processing techniques for glaucoma detection." International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) 3.12 (2014).

[31]. González, Javier, Michael Osborne, and Neil D. Lawrence. GLASSES: Relieving the myopia of Bayesian optimisation." (2016).

[32]. Fandango, Armando. Mastering TensorFlow 1. x: Advanced machine learning and deep learning concepts using TensorFlow 1. x and Keras. Packt Publishing Ltd, 2018.

[33]. Ganguly, Kuntal. Learning Generative Adversarial Networks: Next-generation Deep Learning Simplified. Packt Publishing, 2017.

[34]. Pujari, Pradeep, Md Rezaul Karim, and Mohit Sewak. "Practical Convolutional Neural Networks." (2018).

[35]. Manaswi, Navin Kumar, Navin Kumar Manaswi, and Suresh John. Deep Learning with Applications Using Python. Apress, 2018.

[36]. Kelly, Sloan. Python, PyGame and Raspberry Pi Game Development. Apress, 2016.

[37]. Ketkar, Nikhil. *Deep Learning with Python*. Apress, 2017.