Analogizing the Potency of Deep Learning Models for Glaucoma Detection

Yves Attry Kalin
Department of CSE
Sharda University
Greater Noida, India
2018012875.amit@ug.sharda.ac.in

Yash Bhardwaj
Department of CSE
Sharda University
Greater Noida, India
2018012564.yash@ug.sharda.ac.in

Subhash Arun Dwivedi
Department of CSE
Sharda University
Greater Noida
subhashaks95@gmail.com

Arun Prakash Agrawal
Department of CSE
Sharda University
Greater Noida, India
arun.agrawal@sharda.ac.in

Shravyam Risariya
Department of CSE
Sharda University
Greater Noida, India
2018004862.shravyam@ug.sharda.ac.in

Vivek Jha
Department of CSE
Sharda University
Greater Noida, India
2018012719.vivek@ug.sharda.ac.in

***Abstract*— The ocular disorder inhibiting vision, has been abruptly proliferating due to evolution in mankind and its tack of existence. One such vision-threatening malady is Glaucoma, which mutilates the optic nerve of the eye which can cause sudden blindness. Ophthalmologists utilize copious methodologies to detect the disease but the accuracy is constrained due to humane limitations. There has been indagation for recognizing Glaucoma incorporating Machine Learning and Deep Learning Models with varied ideations but the efficacy isn't apt for such cardinal affliction. In this paper, we have collated various Deep Learning Architectures based on diverse parameters, such as MobileNetV2, DenseNet121, InceptionV3, InceptionResNetV2, ResNet50, and VGG16, appealed over 02 Classes categorized into 'G' and 'NG'. The dataset has been aggregated from various sources and pre-processed for utmost proficiency. Profuse metrics have been integrated for validating the potency of our contrivances namely Precision, Recall, F1-Score, Cohen Kappa Score, and AUC. The modus operandi gave us phenomenal results with high veracity.**

Keywords* — *Glaucomatous (G), Non-Glaucomatous (NG), Area Under Curve (AUC), Visual Geometry Group (VGG)

I. INTRODUCTION

As claimed by Glaucoma Research Foundation [16], 65-75% of mortals possessing Glaucoma haven't been diagnosed yet. Glaucoma is the prime genesis of blindness for people over 60 years old. Woefully, only about half of the individuals who possess Glaucoma are aware of it. It is because Glaucoma doesn't have early symptoms. Ophthalmologists often term Glaucoma as "Sneak Thief of the Eye". Glaucoma is a malady occurring due to disruption of the optic nerve which is accountable for traversing visual signals from the eye to the brain. It typically transpires when the fluid in the eye does not drain as it should. This fluid progressively builds up and surges pressure in the eye which thereby mangles the Optic Nerve. More precisely, in the healthy eye, fluid called "Aqueous Humour" is made in the frontal eye and it flows through a tiny drain called "Trabecular Meshwork". The trabecular meshwork is lodged in an area known as Drainage Angle. If the fluid doesn't flow out of the drainage angle appropriately, the ocular pressure amplifies and disrupts the optic nerve. The increased pressure in the eye is known as Intraocular Hypertension. The blank spots initiate to appear in an individual's field of vision suffering from intraocular hypertension. Moreover, an individual won't notice these blank spots frequently until the optic nerve is significantly damaged and the spots become large and if the optic nerve is sternly damaged, that eye will go blind. Eye-Pressure alters regularly, due to which a single eye pressure test won't suffice to detect Glaucoma in some people who have the disease. Therefore, it's crucial to see ophthalmologists regularly and this is where we come to help them to demystify a person possessing Glaucoma.

Ophthalmologists find it difficult to scrutinize the malady and detect its presence due to complex outcomings and human curtailment. The prime rationale of the paper is to automate the fundus imagery process through malady detection and thereby aiding ophthalmologists and the host suffering from it. As a result, in this paper, we've proposed a process through which an ophthalmologist would be able to ascertain the existence of Glaucoma in an individual with utmost efficacy. For this, we subsumed and assorted varied databases such as HRF (High-Resolution Fundus) Image Database, Drishti-GS1, RIM-ONE, ACRIMA and pre-processed them contemplating it over pre-trained models selected based on diversified outlook and complexity. The outcome was remarkably compelling with high verisimilitude potency and gave us the authenticity of our modus operandi.

II. RELATED WORK

Commencing the perspicacity for the effect of Glaucoma possessed by any human, projected an outlook towards diverse methodologies for its Detection. Fractal Analysis has been incorporated with Optic Disc Segmentation and K-Means

Clustering giving an efficiency of 92%. But, due to the interlacing of the optic cup with blood vessels, the optic cup segmentation is a bit of a tedious task and increases the perpetual complexity [1]. Analysis of various image-processing techniques has been carried out over Support Vector Machine (SVM) classifiers gaining an accuracy of 90-95% over miscellaneous features lacking the number of instances [2]. Extraction of deep characteristics along with identifying Range-of-Interest and thereby applying Convolutional Neural Network (CNN) has been contemplated giving the Area Under Curve (AUC) of 0.8483, with the requirement of collating more features for better results [3]. Glaucoma Detection from a set of fundus images in an alliance with the Center of Prevention and Attention of Glaucoma in Bucaramanga, Colombia using a Cup-to-Disc Ratio (CDR) has also been induced fetching an efficacy of 88.5% with the lack of glaucomatous dataset [4].

Transfer Learning (ResNet-50 & GoogLeNet) has also experimented for early glaucoma detection giving an insight of GoogLeNet performing better than ResNet-50 with an AUC of 0.75 and 0.74 respectively, where the results aren't promising [5]. Comparative study between datasets namely HRF Image Database, Drishti-GS1, RIM-ONE, ACRIMA, etc. trained over varied pre-trained deep learning architectures has been proposed where the AUC revolves around 0.85-0.95 [6]. Furthermore, an exploratory study for Glaucoma Detection has also been proposed using the DenseNet121 model with the correctness of 95.6% and F1-Score of 0.97% over a proprietary database (RETINA) [7]. Glaucoma Detection using SVM Classifier procuring 86% accuracy persuading CDR methodology has been carried out as well. The scope of improvement pirouettes around optic cup segmentation [8]. Digital processing techniques have been used phasing RGB channels and analysing over Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR) values where it is concluded that MSE and PSNR values don't promise precise visual results [9]. Moreover, pre-trained models were anatomized over the REFUGE1 dataset utilizing CDR producing the highest AUC of 0.9785 through ResNet101V2 Architecture with the scope of multi-modal evaluation and focus upon clinical features [10].

Having said that, there hasn't been a poignant imposition for the malady disclosure and thus our modus operandi comes into play which not only focuses upon detecting a disease but also encapsulating the fact of computational complexity and the awareness of the impact of layers of a model over a problem domain.

III. Methodology Used For Analysis

The indagation through related work for Glaucoma Disorder led us to collate a 02-Class Dataset termed as 'G' (Glaucomatous) and 'NG' (Non-Glaucomatous), as the impact of revelation of any disease is much important than apprehending its intensity in the initial outlook. The representation of the dataset and the balance between the number of Instances in each class has been depicted in Fig.1.

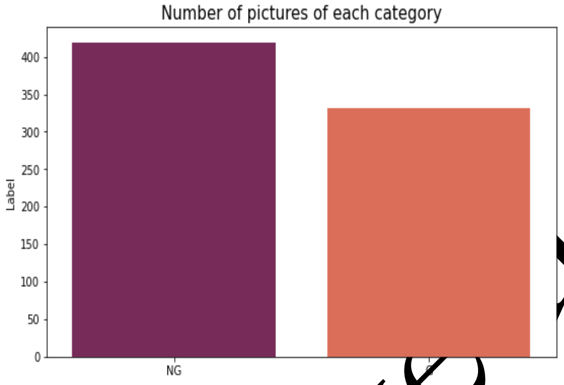

Fig.1. Demonstration of 02-Class Dataset for Glaucomatous (G) and Non-Glaucomatous (NG) Eyes.

Furthermore, we scrutinized different Deep Learning Architectures based on diverse parameters affecting the overall system's execution. When dealing with Image Data it's very prominent to encompass the phenomenon's such as computational complexity, the depth of the model utilized and its size. These parameters deliver an enormous impact over any plan of action. Thus, the influencing factors over the operation of the models are:

- High Number of Parameters which makes the execution of the model sedated,
- The increment in the number of Recurrent Units,
- The utilization of complex Activation Functions such as ReLU and,
- The employment of profound networks.

Considering the trailing elucidations, we constricted ourselves to certain pre-trained models based on the above discussed outlook and also, to demonstrate an impact of increment in the number of layers of a model over a problem domain we utilized Lightweight as well as Deeper Models for the factual analysis. Table-1 accentuates the comparison between varied prominent pre-trained models.

TABLE I. CONTRASTING AMONG THE EMPLOYED PRE-TRAINED MODELS

S.No.	Model	Size (Mega-bytes)	Parameters (approx. in Millions)	Depth
1.	MobileNetV2	14	3.5	88
2.	DenseNet121	33	8	121
3.	InceptionV3	92	23.8	159
4.	InceptionResNetV2	215	55.8	572
5.	ResNet50	99	25.6	168
6.	VGG 16	528	138	23

Let's see the structure of each model:

A. MobileNetV2 Architecture:

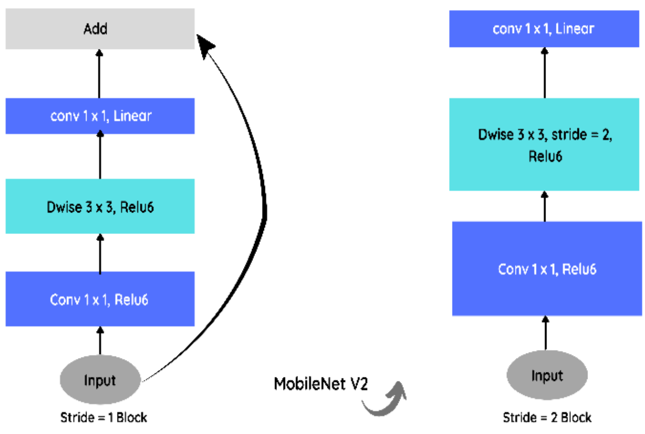

Fig.2. The Basic Structure of MobileNetV2 Model.

B. DenseNet121 Architecture:

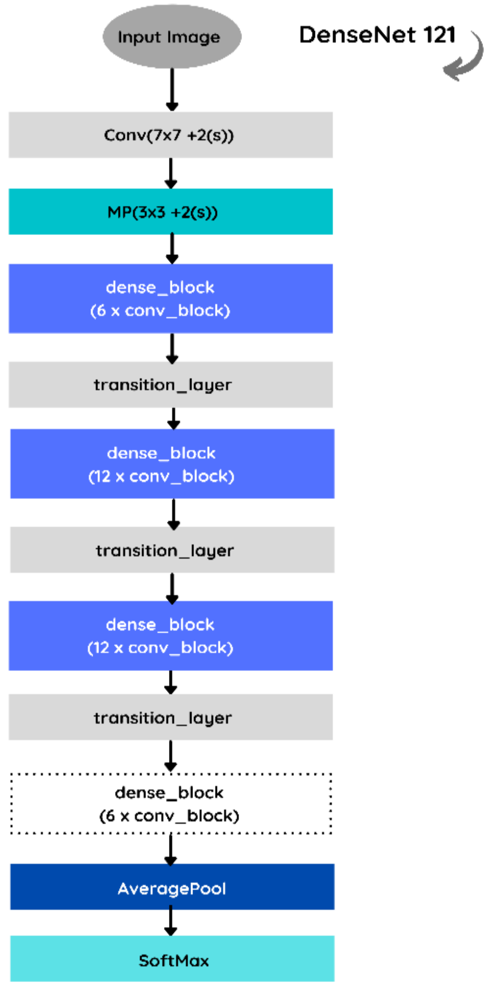

Fig.3. The Basic Structure of DenseNet121 Model.

C. InceptionV3 Architecture:

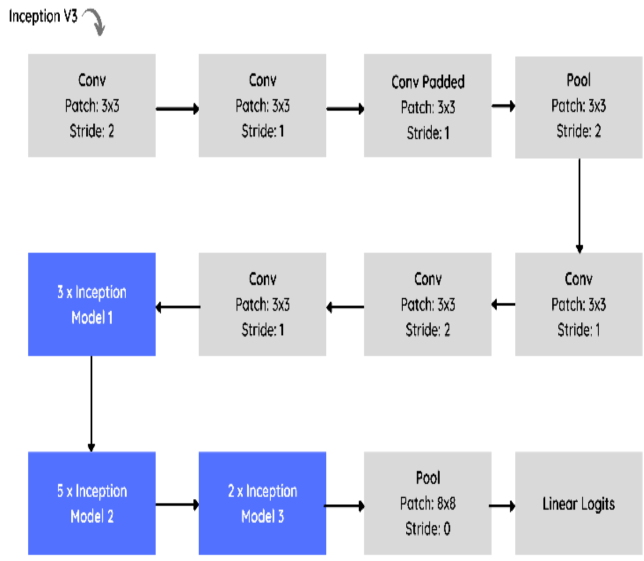

Fig.4. The Basic Structure of InceptionV3 Model.

D. InceptionResNetV2 Architecture:

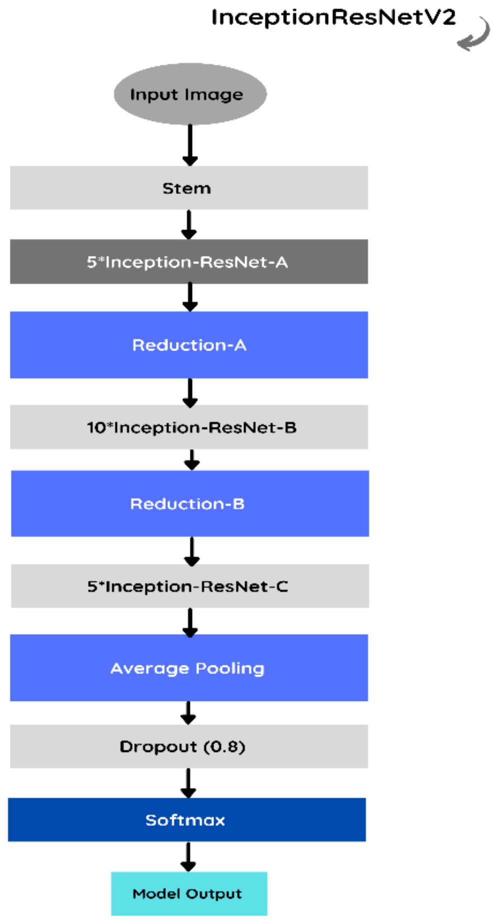

Fig.5. The Basic Structure of InceptionResNetV2.

E. ResNet50 Architecture:

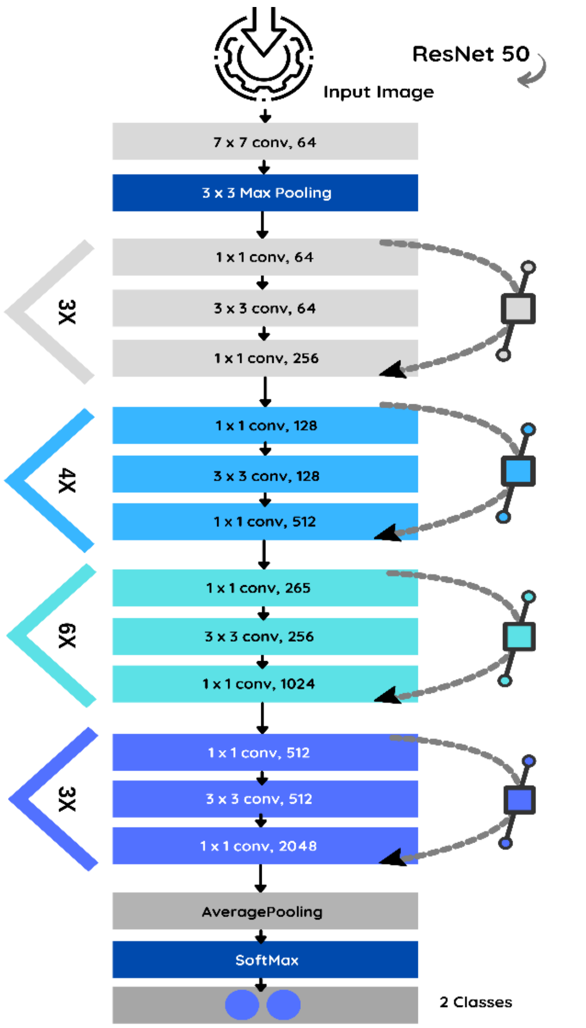

Fig .6. The Basic Structure of ResNet50 Model with Identity Mapping.

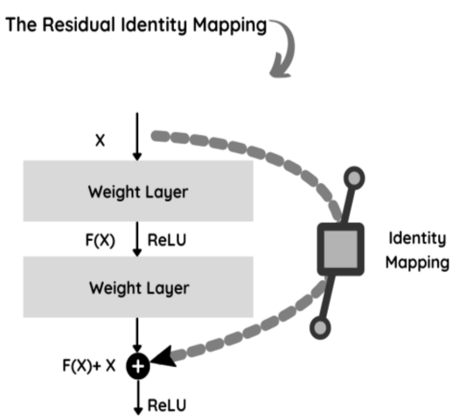

Fig .7. The Demonstration of Residual Identity Mapping.

F. VGG16 Architecture:

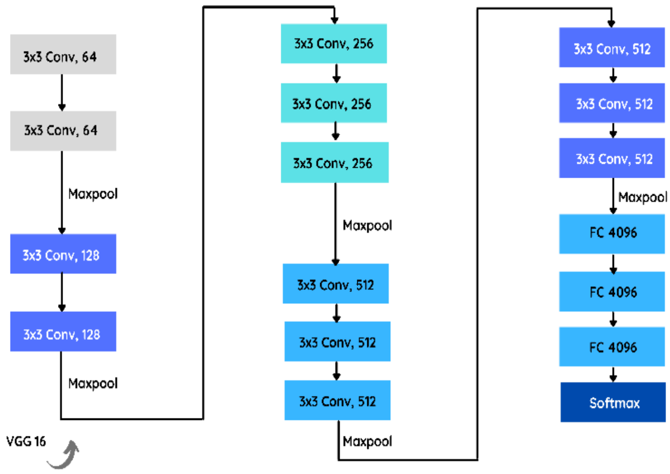

Fig .8. The Basic Structure of VGG16 Model.

The models incorporated are feasible for our problem domain and is one of the best when juxtaposed with other pre-trained models. The graphical representation has been accentuated for each model utilized commencing our modus operandi.

IV. Implementation and Tools

A. Utilized Dataset and its Source: The collated dataset has been conjectured from varied sources such as HRF Image Database, Drishti-GS1, RIM-ONE, and ACRIMA database. The images from each dataset have been aggregated, generating 751 instances which after augmentation results in 9012 samples. Dealing with the pre-trained models it's necessary to resize the instances into 224x224 for execution, so we resized our dataset and further imposed Image Erosion as a part of image processing which accentuates the dark zone in an image. The demonstration of the dataset is depicted in Fig.9.

Fig .9. The Pictorial Representation of Images collated from HRF Image Database, Drishti-GS1, RIM-ONE, and ACRIMA with Eroded Samples.

B. Metrics employed for ascertaining Accuracy: The accuracy revolving around Train, Validation and Test isn't enough to analyse and impose the efficacy of a proposed methodology. Thus, to validate our result and thereby our modus operandi we induced various metrics portraying our potency over the problem paradigm. We incorporated Precision, Recall and F1-Score calculated through the Confusion Matrix.

Also, we inculcated Area Under the Curve (AUC) detecting the Measure of Separability to gauge the efficacy of the models utilized. Furthermore, Cohen Kappa Score which is used to gauge the agreement between the evaluators has also been imposed. Fig.10. demonstrates certain metrics gauged through Confusion Matrix.

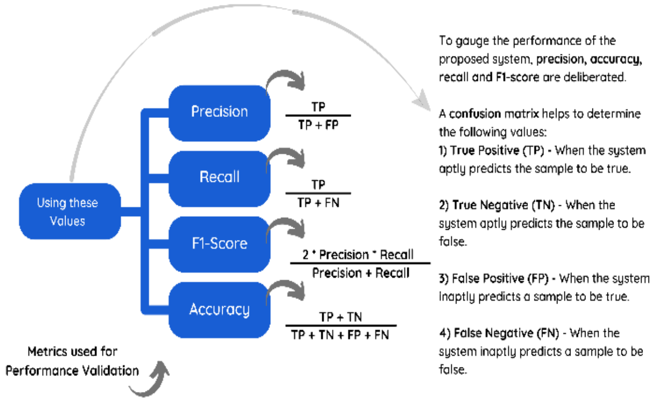

Fig .10. Metrics Evaluation Process for Precision, Recall, F1-Score, Accuracy.

C. Implementation Flow: Demonstrated in Fig.11.

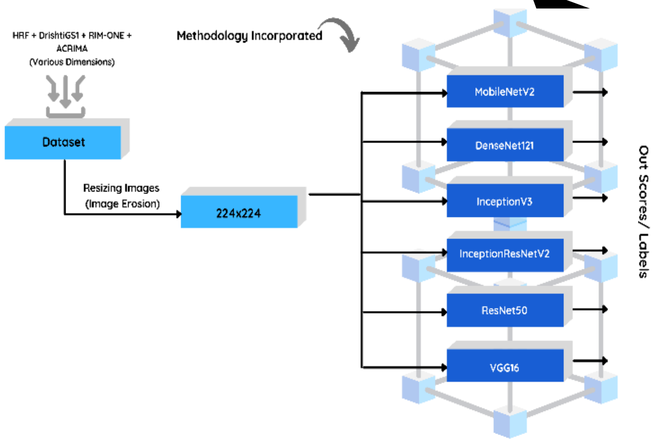

Fig .11. Implementation Structure with Full Process.

V. Experimental Results and Analysis

Post-implementation we got overwhelming and ideal results depicted in Fig.12, Fig.13 and Fig.14. The output achieved gave us an edge over the existing approaches and a formidable outlook for automating the detection of Glaucoma Disorder. Furthermore, Table-2 depicts the Train, Validation and Test Accuracies for various utilized Deep Learning Models.

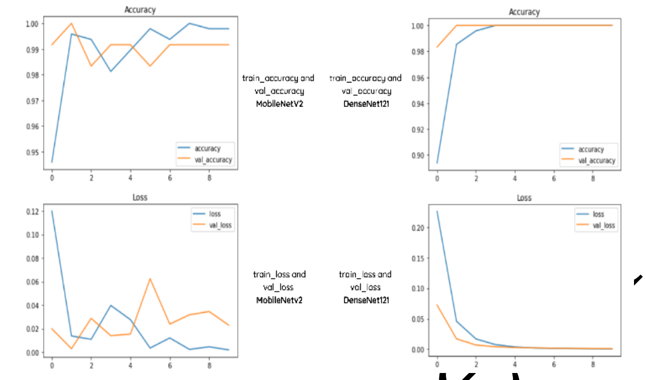

Fig .12. Graphical Results for MobileNetV2 and DenseNet121

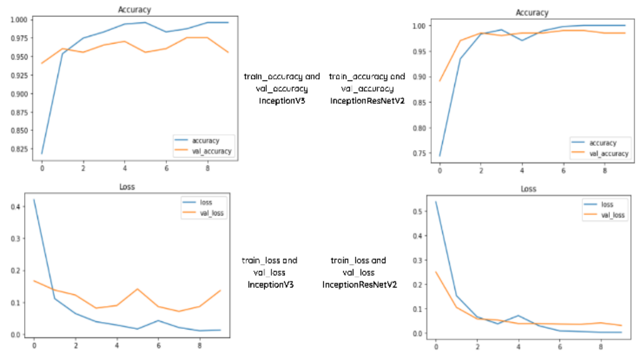

Fig .13. Graphical Results for InceptionV3 and InceptionResNetV2

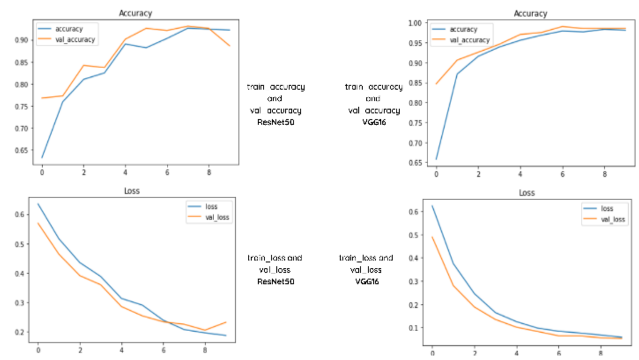

Fig .14. Graphical Results for ResNet50 and VGG16.

TABLE II. The Depiction of Train, Validation and Test Accuracy Encompassing Induced Models.

Model	**Train-Accuracy**	**Validation-Accuracy**	**Test-Accuracy**
MobileNetV2	99.79%	99.17%	99.34%
DenseNet121	99.58%	98.33%	98.68%
Inception V3	99.50%	97.52%	97.37%
InceptionResNetV2	98.94%	98.51%	96.05%
ResNet50	93.99%	91.91%	92.09%
VGG16	98.10%	97.03%	98.68%

TABLE III. VARIED METRICS ACCURACIES UTILIZED TO GAUGE POTENCY OF MODELS

Model	AUC	Precision	Recall	F1-Score	Cohen Kappa Score
MobileNetV2	0.9996	0.9935	0.9934	0.9934	0.9866
DenseNet121	0.9998	0.9871	0.9867	0.9867	0.9729
InceptionV3	1.0000	0.9737	0.9737	0.9737	0.9474
Inception-ResNetV2	0.9997	0.9639	0.9605	0.9607	0.9201
ResNet50	0.9899	0.9345	0.9209	0.9209	0.9091
VGG16	0.9984	0.9872	0.9868	0.9868	0.9737

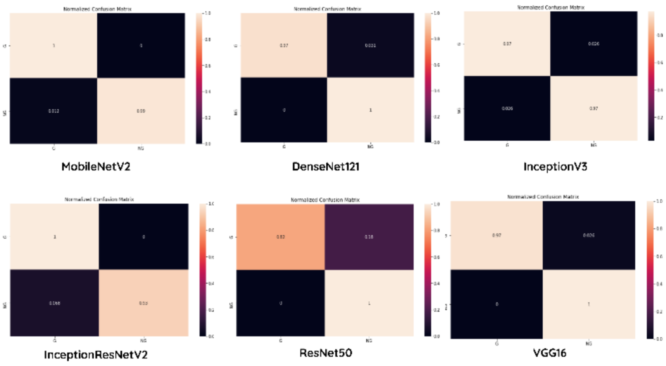

Fig .15. Confusion Matrix for MobileNetV2, DenseNet121, InceptionV3, InceptionResNetV2, ResNet50, VGG16.

VI. CONCLUSION AND FUTURE SCOPE

Inferring the accuracies gauged through varied metrics as demonstrated in Table -2 and Table -3, we come to a conclusion where MobileNetV2 performs phenomenally being the most light-weight and the least computationally complex. Moreover, Densenet121 and VGG16 also performed extremely well when juxtaposed on the basis of the size of the models. Thus, one thing to notice here is, the efficacy of the model doesn't solely depend upon the depth of the model i.e., more the number of layers in a model more the accuracy isn't true always. Moreover, the efficacy of any specific model is dependent on the problem domain and to be more precise, over the Dataset utilized. Certainly, due to which we got lower accuracy for InceptionResNetV2 and ResNet50 even though these models are huge in terms of Depth. The juxtaposition got us to a model we would incorporate for the creation of a real-time Glaucoma Disorder Detection System, MobileNetV2. The induction of automation in the field of Fundus Imagery System can assist ophthalmologists to a huge extent and thereby make the whole methodology more efficient. The future scope revolves around the inculcation of a greater number of instances which can assure and give out more potent outlook for modus operandi as the dataset remains the biggest hindrance in this sector.

REFERENCES

[1] S. C. Shetty and P. Gutte, "A Novel Approach for Glaucoma Detection Using Fractal Analysis," 2018 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2018.

[2] S. Kaushal, S. Datt Sharma, and S. Jain, "Investigation of Image Processing Techniques for Glaucoma Detection in Human Eyes," 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC), 2018.

[3] A. Li, Y. Wang, J. Cheng, and J. Liu, "Combining Multiple Deep Features for Glaucoma Classification," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018.

[4] J. Carrillo, L. Bautista, J. Villamizar, J. Rueda, M. Sanchez and D. rueda, "Glaucoma Detection Using Fundus Images of The Eye," 2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA), 2019.

[5] A. Serener and S. Serte, "Transfer Learning for Early and Advanced Glaucoma Detection with Convolutional Neural Networks," 2019 Medical Technologies Congress (TIPTEKNO), 2019.

[6] S. Serte and A. Serener, "A Generalized Deep Learning Model for Glaucoma Detection," 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2019.

[7] S. Ovreiu, I. Cristescu, F. Balta and E. Ovreiu, "An Exploratory Study for Glaucoma Detection using Densely Connected Neural Networks," 2020 International Conference on e-Health and Bioengineering (EHB), 2020.

[8] R. Krishnan, V. Sekhar, J. Sidharth, S. Gautham, and G. Gopakumar, "Glaucoma Detection from Retinal Fundus Images," 2020 International Conference on Communication and Signal Processing (ICCSP), 2020.

[9] G. Satya Nugraha, B. Amelia Riyandari, and E. Sutoyo, "RGB Channel Analysis for Glaucoma Detection in Retinal Fundus Image," 2020 International Conference on Advancement in Data Science, E-learning and Information Systems (ICADEIS), 2020.

[10] H. Gunasinghe, J. McKelvie, A. Koay and M. Mayo, "Comparison Of Pretrained Feature Extractors For Glaucoma Detection," 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), 2021.

[11] M. Aljazaeri, Y. Bazi, H. AlMubarak and N. Alajlan, "Faster R-CNN and DenseNet Regression for Glaucoma Detection in Retinal Fundus Images," 2020 2nd International Conference on Computer and Information Sciences (ICCIS), 2020.

[12] A. -M. Ştefan, E. -A. Paraschiv, S. Ovreiu and E. Ovreiu, "A Review of Glaucoma Detection from Digital Fundus Images using Machine Learning Techniques," 2020 International Conference on e-Health and Bioengineering (EHB), 2020.

[13] A. Saxena, A. Vyas, L. Parashar, and U. Singh, "A Glaucoma Detection using Convolutional Neural Network," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020.

[14] S. OVREIU, I. CRISTESCU, F. BALTA, A. SULTANA and E. OVREIU, "Early Detection of Glaucoma Using Residual Networks," 2020 13th International Conference on Communications (COMM), 2020.

[15] S. S and U. A. C, "A Novel Method for Glaucoma Detection Using Computer Vision," 2020 Third International Conference on Advances in Electronics, Computers and Communications (ICAECC), 2020.

[16] https://www.glaucoma.org/