IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Automated OCT-Based Retinal Disease Classification Using Lightweight MobileNetV3 Deep Neural Network

1st Er. Balwinder Singh
Dept. of Computer Science and Engineering
Punjabi University, Patiala
Email: singhbalwinder2514@gmail.com

2nd Dr. Navneet Kaur
Dept. of Computer Science and Engineering
Punjabi University, Patiala
Email: mavi_navneet@yahoo.co.in

3rd Dr. Sikander Singh Cheema
Dept. of Computer Science and Engineering
Punjabi University, Patiala
Email:cheemasikander8@gmail.com

***Abstract*—In today's digital world , where screens are inevitable part of lifestyle, eye-related problems are rapidely increasing. Many Retinal diseases like Diabetic Macular Edema (DME), Choroidal Neovascularization (CNV), Age-related Macular degeneration (AMD) , Diabetic Retinopathy (DR) , Macular Hole (MH) and drusen, are the major causes of permanent vision loss worldwide. To prevent the further vision loss, timely and precise diagnosis is very important. For that,a standard, noninvasive imaging technique like Optical Coherence Tomography (OCT) are popular nowadays which provides cross-sectional images of the retina in high-resolution, enabling detailed morphological assessment. Just Because, the manual understanding of huge number of OCT scans is subjective, time-extensive and requires high expertise, that's why delays in diagnosis and variability in results of diagnosis is common , which create a big problem. To overcome this problem , we present an study on deep learining bases automated diagnostic system for retinal diseases to classify various retinal diseases by using retinal OCT scans. so, We introduced a model based on Convolutional Neural Network to find out the presence of common retinal problems efficiently. This model was trained and validated using a dataset of retinal OCT images named as "OCT-8", which is publicly available on kaggle. This dataset consists retinal OCT images classified into 8 classes named as: AMD, CNV, CSR, DME, DR, DRUSEN, MH and NORMAL. This automated identification system has a great efficieny to work as a decision support tool, make easy to identification task and to provide quick and accurate outputs to help the ophthalmologists.**

***Index Terms*—Deep Learning, Convolutional Neural Networks,Retinal Disease Classification, Optical Coherence Tomography (OCT), MobileNetV3, EfficientNet,Age-Related Macular Degeneration (AMD) , Central Sereous Retinopathy (CSR) , Diabetic Macular Edema (DME), Diabetic Retinopathy (DR) , Choroidal Neovascularization (CNV), Macular Hole (MH), Drusen, Automated Diagnosis.**

I. Introduction

Visual problems and blindness is a biggest nightmare for global health, which affects many people in terms of their quality of life and financial aspects. A substantial portion of vision loss is attributable to retinal diseases that damage the photoreceptive tissue present on the retina of the eye. Among the most widespread vision-destroying retinal conditions are Age-Related Macular Degeneration (AMD), by the presence of drusen and potentially progressing to Choroidal Neovascularization (CNV), Diabetic Retinopathy (DR) and Diabetic Macular Edema (DME), a major complication of diabetes. Additionally, conditons such as Central Serous Retinopathy (CSR), marked by fluid accumulation under the retina, and Macular Holes (MH) , which involve a physical tear in a fovea, present severe risk to central vision.

The early detection is the main strategy for reducing the serious impact of these retinal problem so that treatment can start without delay. A Postponed diagnosis can leads to irreversible damage to retinal structures and permanent vision loss. That's why, Optical Coherence Tomography (OCT) comes in play as a major technological reform in ophthalmic imaging, and became a standard method for retinal assessment. OCT is a non-invasive imaging technique which works as an "optical biopsy," and provide fragmentary images of the retinal layers in high resolution. It helps to opthalmologist to identify and visualize important features like sub-retinal and intra-retinal fluid, and deposits of drusen patches, which is important for the identificatin and treatment of conditions such as DME and AMD. To tackle this clinical demand, Deep Learning (DL) is introduced in this field to make task easy and effiecient. Primary work done by Pekala et al. (2018) indicates that segmentation of retinal layers with an accuracate results can be acheived using Fully Convolutional Networks (FCNs) Without human experts. Later, Sun et al. (2019), Sahin (2022), and Kim Tran (2021)'s studies show that various retinal diseases can be classified efficiently using the advanced Convolutional Neural Network (CNN) models, and according to some studies their accuracy rates can be reached above 94 percent (Jin, 2024).

II. Related Work

The automatic recognition of retinal diseases using Optical Coherence Tomography (OCT) images become very advanced with the use of Deep learning. Initial study by Kermany et al. [19] shows that retinal diseases from OCT images can be precisely classified using the convolutional neural networks (CNNs). Also, De Fauw et al. [18] showed that for diagnosis and decision making point of view performance of deep learning systems is quite impressive.

Some studies are also performed which have focused on improving classification accuracy using deep learning models.Like, Sahin [1] introduced a CNN-based framework for retinal disease detection, where as Talebzadeh et al. [2] showed the limited datasets are one of the major limitation for by deep learning techniques, which affect their accuracy. Kim and Tran [4] and Berrimi and Moussaoui [6] also showed that multi-class retinal disease classification can be effieciently achieved by CNN- based models.

To enhance the diagnosis accuracy, some studies also focused on segmentation based approaches, instead of just relaying on classification. Like, Pekala et al. [3] shows that for better retinal structure visualization and identification, deep learning based segmentation of OCT image, is a better choice. Similarly, Lin et al. [10] and Guo et al. [13] highlights the significance of detecting the regions of retinal fluid, which works as a critical indicators of disease worsening of disesase.

With the time, efficient and lightweight deep learning models comes into play for real-time applications. Like, Sunija et al. [7] intoduced OCTNet, a lightweight model based on CNN for classification of retinal disease efficiently . Later, for multi-scale approaches, the multi-scale CNN model like ROCT-Net[11] introduced by Sotoudeh-Paima et al. [12], further by represent features at different resolutions, performance can be enhanced.

Later, due to their advanced feature extraction capabilities and scalability, basic architectures like ResNet [15], EfficientNet [16], and MobileNetV3 [17] have been pre-dominantly used as foundation models for OCT image classification. To improve the performance on limited medical datasets, these models are generally applied through transfer learning.

Recently,. Noor et al. [14] introduced an AI-based model for retinal disease classification, which improve explainability of task. Elkholy and Marzouk [8] and Wang [9] also showed that modern deep learning architectures, enhance classification performance. Also, survey studies by Rasti et al. [22] provide a in depth analysis of deep learning techniques and their challenges for OCT image analysis.

Even with these advancements, many existing approaches depends on computationally intensive models, which limits their deployment in real-time environments. That's why, lightweight and efficient architectures like MobileNetV3, which can achieve high accuracy while less computational complexity is needed.

III. Material and Methods

A. Dataset

For this study , we used a publicly avaliavle datased on Kaggle named as 0CT-8 [23], which provides a high-quality retinal OCT images labeled with different categories of retinal diseases, which forms the base for training, validating, and testing the model. The dataset was divided into three folders (train, test, and val) and contained subfolders for each image category (AMD, CNV, CSR, DME, DR, DRUSEN, MH and NORMAL). There are 24,000 OCT images (JPEG) and 8 categories (AMD, CNV, CSR, DME, DR, DRUSEN, MH and NORMAL)

Class	Train	Validaion	Test	Total
AMD	2,300	350	350	3,000
CNV	2,300	350	350	3,000
CSR	2,300	350	350	3,000
DME	2,300	350	350	3,000
DR	2,300	350	350	3,000
DRUSEN	2,300	350	350	3,000
MH	2,300	350	350	3,000
NORMAL	2,300	350	350	3,000
Total	18,400	2,800	2,800	24,000

TABLE I
TRAINING,TESTING AND VALIDATION DATASETS USED IN EXPERIMENT

Images are labeled as (disease)-(randomized patient ID)-(image number by this patient) for e.g., drusen_test_1001.jpg

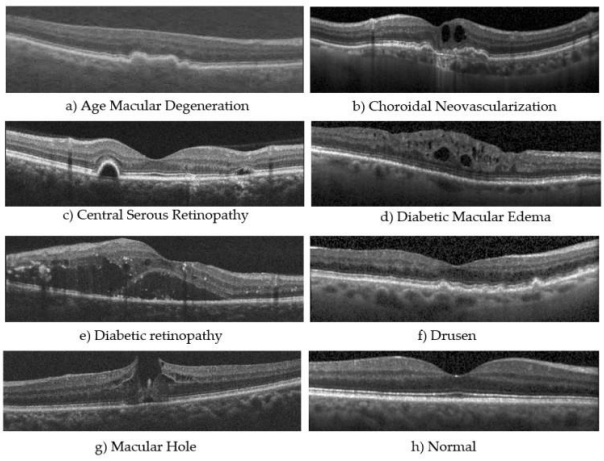

Fig. 1. Sample OCT images showing AMD, CNV, CSR, DME, DR, DRUSEN, MH and NORMAL

B. Data Preprocessing and Augmentation

To prepare the images for the model and improve its generalization capabilities,the following pre-processing procedures were carried out:

- Image Resizing: All OCT images will be resized to a similar dimension of 224 × 224 pixels to match the input size requirements of the pre-trained model.
- Normalization: Pixel values of the images, originally in the range [0, 255], will be normalized to a range of [0, 1] by dividing each pixel value by 255. This ensured that the input data had a consistent scale.
- Data Augmentation (On-the-fly): To prevent overfitting and expose the model to a different variety of image modifications, randomly applied to the training data during the training process as following mentioned:
  Random Rotation: ±10 degrees
  Random Horizontal Flip: 50% probability
  Random Zoom: 10-20%
  Width and Height Shift: ±10%

C. Training Method

The introduced deep learning-based system for identification of various retinal diseases using retinal images use Convolutional Neural Networks (CNNs) with a pretrained MobileNet model which is fine-tuned for this specific task. MobileNetV3 was created by Google's AI research team, Google Brain.The model is designed to run efficiently on mobile and embedded hardware. Its architecture consists advanced techniques, such as network architecture search (NAS), to make a balance between strong performance and low computational cost. Because of their lightweight structure, MobileNetV3 models use less memory, which makes them well-suited for resource-constrained hardware and results in faster loading times.

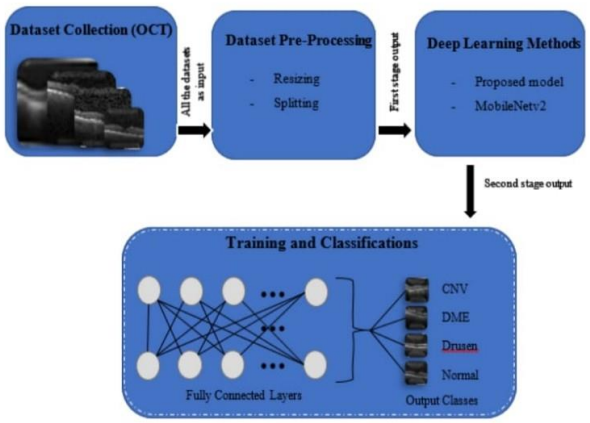

Fig. 2. Block Diagram of Study

An overview of MobileNetV3 Architecture

The MobileNEtV3 architecture was specially designed for lightweight mobile devices.To fine-tune layers, it generally use Neural Architecture Search and the NetAdapt algorithm. The main building blocks of the architecture are as follow:

- Input Layer : Images of fixed size (224 × 224 pixels) are used as input.
- Convolutional Layers:MobileNetV3 mainly use depthwise separable convolutions to reduce computational complexity as well as to achieve good accuracy , it's two major consecutive layers are as follow:

  Depthwise Convolution: In step, specific kernel is used for each input channel for convolved them separately. For each input channel, a small filter is used to minimize the computation as compare to standard convolutions.

  Pointwise Convolution : It helps to combine information from different channels by combining output channel sequentially from output of depthwise convolution.
- Downsampling and Bottlenecks: Downsampling is used to minimize spatial dimensions. Also, to minimize the number of input channels Bottleneck layers (1 × 1 followed by 3 × 3 depthwise convolution) is used.
- Activation function: Rectified Linear Unit (ReLU) as activation function is applied to each layer so that variability can be increased. Also, at the end, Global Average Pooling (GAP) used to minimize spatial dimensions for each value per channel.
- Output Layer : This layer generate final predicted output by using above steps. For classification tasks, this output layer consists several units, where each unit indicates to particularly one class from all classes present in the dataset and shows result by converting the raw outputs into the probabilities for the classes.

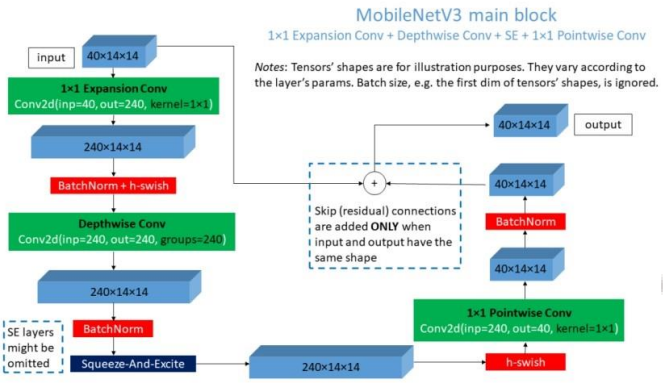

Fig. 3. MobileNetV3 architecture block diagram

Require: Training Data
Ensure: Evaluated Model
1: Initialize MobileNetV3 model
2: $model \leftarrow$ MobileNetV3(pretrained = True)
3: Modify the final classification layer
4: Replace final layer with Dense layer for disease classes
5: Define optimizer and loss function
6: $optimizer \leftarrow \text{Adam}(model.parameters, \text{lr} = 0.0003)$
7: $loss\_function \leftarrow$ sparse_CrossEntropyLoss()
8: Load OCT retinal image dataset
9: $dataset \leftarrow$ Load_OCT_Retinal_Dataset()

Perform preprocessing
Resize images to 224×224
Normalize pixel values
Apply data augmentation (rotation, flip, zoom)
Split dataset
$train\_set, test\_set \leftarrow$ Train_Test_Split$(dataset)$
Train the model
for $epoch = 1$ to num_epochs **do**
  for each batch in $train\_set$ **do**
    $predictions \leftarrow model(batch\_images)$
    $loss \leftarrow loss\_function(predictions, labels)$
    $optimizer.zero\_grad()$
    $loss.backward()$
    $optimizer.step()$
  end for
end for
Evaluate model performance
$metrics \leftarrow$ Evaluate$(model, test_set)$
Calculate evaluation metrics:
Accuracy, Precision, Recall, F1-score, Confusion Matrix
Return trained model and evaluation results

Parameter	Value
Proposed Model	MobileNetV3-based classifier
Transfer Learning Type	Feature Extraction / Frozen Base Model
Loss Function	Sparse Categorical Crossentropy
Batch Size	64
Trainable Layers	Head is trainable, Base model is frozen
Optimization Algorithm	Adam Optimizer
Learning Rate	0.0003 (3e-4)
Activation Functions	Swish (Dense layers), Softmax (output layer)
Number of Epochs	20

TABLE II
TRAINING PARAMETERS OF THE PROPOSED MOBILENETV3 MODEL

D. Statistical Analysis Method

The following Performance metrics, were used to evaluate the model's efficiency and performance. Detailed description of metrics are described along with Definitions and equations, where, TP represents to true positives, TN represents to true negatives, FN signifies to false negatives, and FP signifies the false positives.

- Accuracy: It is described as the ratio of the overall correct predictions to the total predictions

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

- Precision : This indicates the how much data instates predicted positive by the model is actually correct. It calculated by the ratio of true positives to the total positives. It is crucial to find when the computational cost of False Positive is high.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

- Recall (Or Senstivity) : It measures how many of actual positive cases our model managed to find. It calculated by the ratio of True Positive to the sum of True Positive and False Negative. It is important to find when missing a positive case is dangerous , for example in our case if model predict diseased eye to the normal eye.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

- F1-score : It calculates the average value of Precision and Recall, by combining them into single metric.

$$F1\text{-score} = \frac{2 \cdot (Precision \cdot Recall)}{Precision+Recall} \quad (4)$$

- Specificity: It measures how well the model identify actual negative cases. It calculated as the ratio of True Negative to the sum of True Negative and False Positive.

$$Specificity = \frac{TN}{TN+FP}$$

IV. EXPERIMENTS AND RESULTS

A. Experiment Environment

The selected models was implemented on linux system with a version of ununtu16.04. The NVIDIA 2080ti Gpu with 251.5 GiB memory,GPU is used for model training. For implementation we use python 3.6 version based on pytorch 1.0.1, keras2.2.4 and tensorflow 1.12.0. The whole network was trained for 30 epoches for both architectures.

For experimentation, we used publicly available dataset named as OCT-8, which consists different sizes of images of retinal OCT images of 8 different classes. For trining purpose we used 70% of total available data and remaing data is used for validation and testing, 15% for each. Before using these images for trainig, normalization is applied on images to scale their dimensions.

B. Performance of Proposed Model

The introduced model based on the MobileNetV3 architecture was evaluated using performance metrics like accuracy, precision, recall, F1-score, sensitivity, and specificity. As per experiment's result ,the training and validation loss curves shows a consistent downward curve across epochs, indicating good efficiency of the model parameters. The training loss decreases sharply from an initial value slightly above 0.40 to nearly 0.21 by the final epoch, which shows step- by-step reduction of error of training dataset over time.

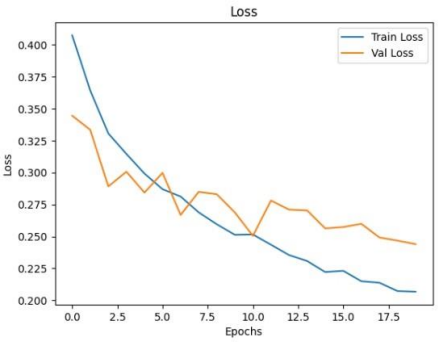

Fig. 4. Visualization of Loss Result

The validation loss graph also show a similar downward pattern, whearas with some slight changes during intermediate epochs. Also, the gap between training and validation loss remains less, which indicates that the model is not overfit. But in next epochs, both curves starts to allign and become stable, which shows a stable and balanced learning state.

The accuracy graph shows that the model"s efficiency gently boost with training. In early epochs where accuracy is comparatively low , but later on it's keeps increasing gradually and touches almost 92% accuracy till final epoch, which shows model can learn and understand valuable parameters and find out useful pattern from the data.

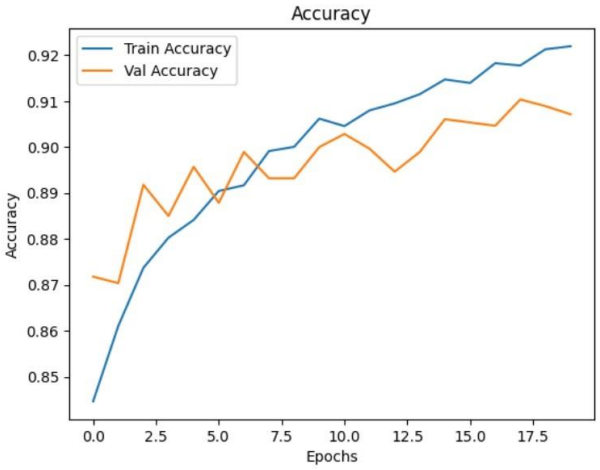

Fig. 5. Visualization of Accuracy Result

The validation accuracy also show similar trend as training accuracy and acheive around 91% of accuarcy till final epoch, which is quiet near to training accuracy.

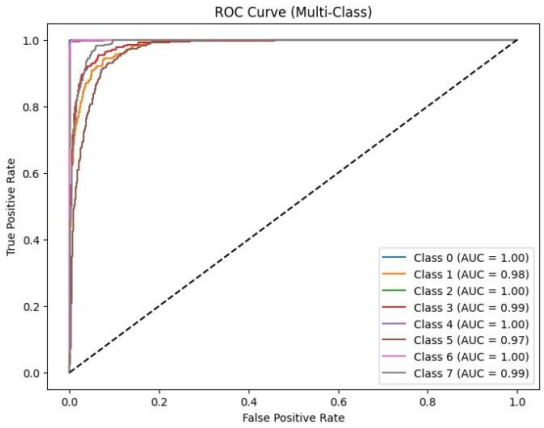

Fig. 6. Visualization of ROC

The class-wise ROC curves also strengthen the reliability of model. As most classes attain highly exact AUC values, with some reaching 1.00, while the other classes maintain values close to 0.97–0.99. This shows, the classifier is highly efficient across different classes, with only slight difference in performance between them.

The micro-average ROC curve shows a sharp rise toward the top-left corner of the graph, which indicate strong classification ability. The shown Area Under the Curve (AUC) of approximately 0.99 highlights the model's ability to classify different classes with high efficiency. Such a high AUC value show a good balance between true positive rate and false positive rate across different decision boundary for a model.

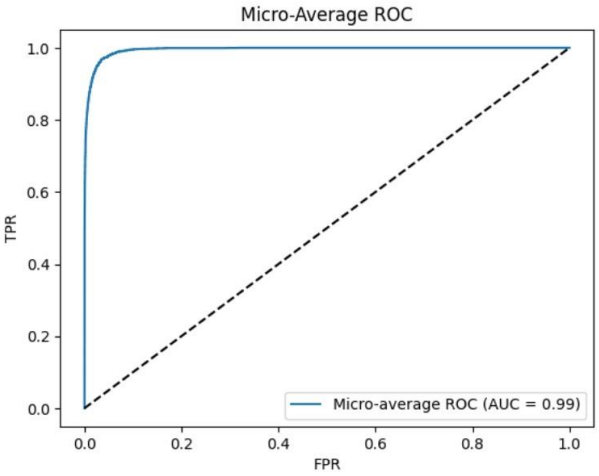

Fig. 7. Visualization of Micro-average ROC curve

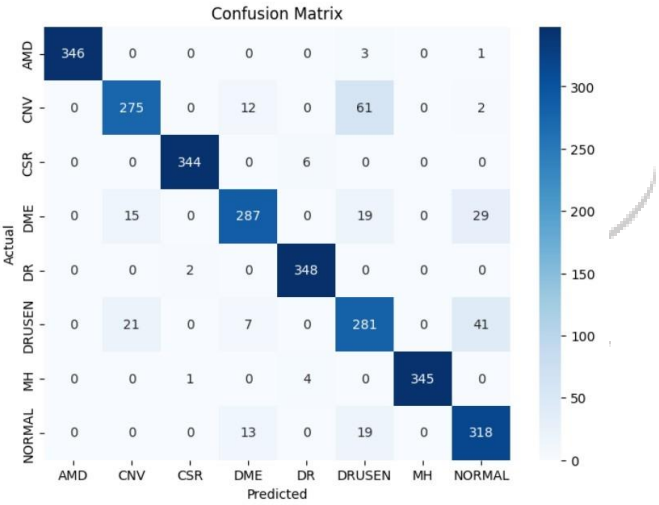

Fig. 8. Visualization of Confussion Matrix

The confusion matrix shows that most of the samples are correctly classified, as values are aligned in diagonal. Although AMD, CSR, DR, and MH show very high correct predictions

with less error,whereas CNV, DME ad DRUSEN lacks ,but have decent acceptable accuracy.

Some sort of confusion is seen between some class pairs, which involves CNV, DME, and DRUSEN, where a small number of samples are falsely identified by the model. Also, some false predictions observed for NORMAL samples, some being predicted as DME or DRUSEN. Regardless these some errors, the overall prediction results shows high classification precision and good performance over different categories.

for evaluate the trained models, We visualize the decision-making process of predictions using Grad-CAM. Grad-CAM is a gradient-based deep network visualization technique which shows categorization using the pixels of image to explain the classification on the basis of deep neural network models in termsf of heat maps.

Because we use VIRIDIS heatmap for Grad-CAM visualization, so Green–yellow regions in images shows higher important retinal feature areas, whereas blue regions indicates less focus area and the darker center patches and purple region are typically due to pre-processing and not indicate any actual features.

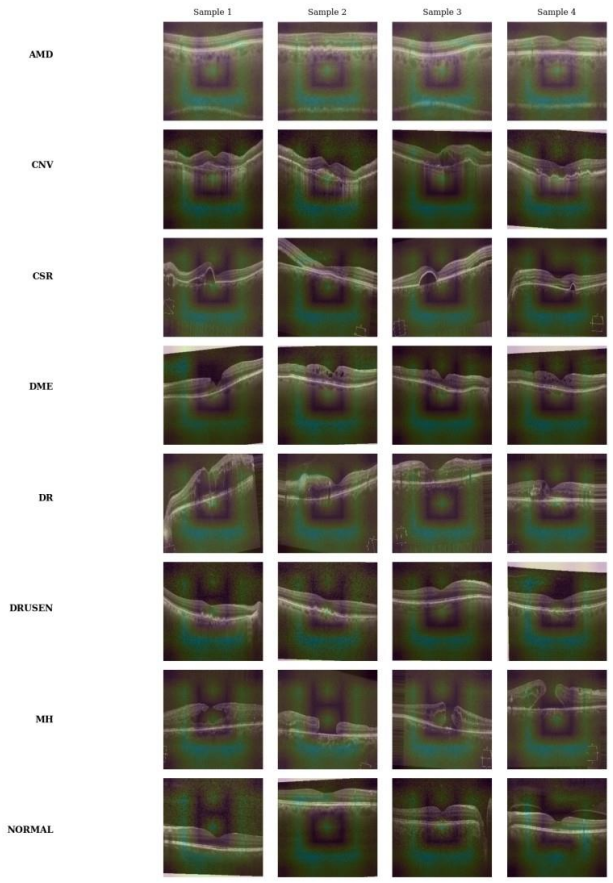

Fig. 9. Gradient-weight class activation mapping on test images

According to Grad-CAM results the model is mostly focusing on the middle retinal area, because most of the important features lies there. But in some classes like AMD, DR, and MH, the highlighted regions look clear and consistent, indicating their strong performance.

Although, in classes like CNV, DME, and DRUSEN, the consideration is little bit shifted or distracted which results to more more false classification for these classes. Generally, the model identify the important region correctly, but not perfectly in every case, which is responsible for errors for some classes during prediction.

C. Results

Overall we can say that the model achieves an accuracy of around 90% on test data, which is clearly supported by classification report, which also shows similar results. The F1-score for both macro and weighted, is near about 0.90, which indicates that the model sustain a good balance between precision and recall.

Class	Accuracy	Precision	Recall	F1-Score
AMD	0.99	1.00	0.99	1.00
CNV	0.73	0.88	0.73	0.80
CSR	0.99	0.99	0.99	0.99
DME	0.79	0.90	0.79	0.84
DR	0.99	0.98	0.99	0.99
DRUSEN	0.83	0.69	0.83	0.75
MH	0.98	1.00	0.98	0.99
NORMAL	0.89	0.80	0.89	0.84

TABLE III
CLASSIFICATION REPORT

As per confusion matrix, most of the predictions are correctly alligned along the diagonal, which shows a large portion of samples are correctly classified. However, some classes show sort of overlapping, where misclassification occur more frequently than other classes. The ROC curves also shows strong performance, with high AUC values for almost all classes.

V. CONCLUSION

Based on overall results, the MobileNetV3 model performs ver well for this classification task. During training it learn useful insights very well and shows a very minimal performance gap between training and validation performance, which clearly indicates that the model is not suffering from overfitting. Also, it achieves around 90% accuracy on test data and the F1-score is also lie in the similar range, which indicates that the performance of model is balanced and reliable.

The confusion matrix and ROC curves strengthen the above fact by showing the correctl prediction for most of the classes, although some categories still get confused with each other, but this minimal confusion is expected in multi-class problems where some similar features are shared by the classes.

Overall, MobileNetV3 shows itself as a reliable and efficient model here, by showing strong results and being lightweight , but still a scope of performance improvement for some classes with better tuning and less confusion is pending, which may achieve in future.

REFERENCES

[1] M. E. Sahin, "A deep learning-based technique for the diagnosis of retinal diseases using OCT images," Turkish Journal of Science and Technology, 2022. doi: 10.55525/tjst.1128395.

[2] M. Talebzadeh, A. Sodagartojgi, Z. Moslemi, S. Sedighi, B. Kazemi, and F. Akbari, "Deep learning-based retinal abnormality detection from OCT images with limited data," World Journal of Advanced Research and Reviews, vol. 21, no. 3, 2024. doi: 10.30574/wjarr.2024.21.3.0716.

[3] M. Pekala, N. Joshi, D. E. Freund, N. M. Bressler, D. Cabrera DeBuc, and P. Burlina, "Deep learning based retinal OCT segmentation," arXiv preprint arXiv:1801.09749, 2018.

[4] J. Kim and L. Tran, "Retinal disease classification from OCT images using deep learning," in IEEE CIBCB, 2021.

[5] M. Miranda and F. J. Romero, "Antioxidants and retinal diseases," Antioxidants, vol. 8, no. 12, p. 604, 2019.

[6] M. Berrimi and A. Moussaoui, "Deep learning for identifying and classifying retinal diseases," in Proc. ICCIS, IEEE, 2020, pp. 1–6.

[7] A. P. Sunija, S. Kar, S. Gayathri, V. P. Gopi, and P. Palanisamy, "OCTNet: A lightweight CNN for retinal disease classification from OCT images," Computer Methods and Programs in Biomedicine, vol. 200, p. 105877, 2021.

[8] M. Elkholy and M. A. Marzouk, "Deep learning-based classification of eye diseases using convolutional neural network for OCT images," Frontiers in Computer Science, vol. 5, 2023. doi: 10.3389/fcomp.2023.1252295.

[9] M. Wang, "Research on retinal OCT image classification based on deep learning," Journal of Computing and Electronic Information Management, 2023. doi: 10.54097/wnr87znj.

[10] M. Lin, G. Bao, X. Sang, and Y. Wu, "Recent advanced deep learning architectures for retinal fluid segmentation on optical coherence tomography images," Sensors, vol. 22, no. 8, 2022. doi: 10.3390/s22083055.

[11] M. Rahimzadeh and M. R. Mohammadi, "ROCT-Net: A new ensemble deep convolutional model for detecting retinal diseases from OCT images," arXiv:2203.01883, 2022.

[12] S. Sotoudeh-Paima et al., "Multi-scale convolutional neural network for automated AMD classification using retinal OCT images," arXiv:2110.03002, 2021.

[13] Y. Guo et al., "Automated segmentation of retinal fluid volumes from OCT using deep learning," arXiv:2006.02569, 2020.

[14] M. Noor et al., "RetinaVision: XAI-driven deep learning framework for retinal disease classification," arXiv:2602.19324, 2026.

[15] K. He et al., "Deep residual learning for image recognition," in Proc. CVPR, 2016. (Used widely as backbone in OCT classification models)

[16] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. ICML, 2019.

[17] A. Howard et al., "Searching for MobileNetV3," in Proc. ICCV, 2019.

[18] J. De Fauw et al., "Clinically applicable deep learning for diagnosis and referral in retinal disease," Nature Medicine, vol. 24, no. 9, 2018.

[19] D. S. Kermany et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," Cell, vol. 172, no. 5, 2018.

[20] H. Fu et al., "Deep learning-based automated detection of retinal diseases from OCT images," IEEE Transactions on Medical Imaging, 2019.

[21] Z. Li et al., "Deep learning for detecting retinal diseases from OCT images," IEEE Access, 2020.

[22] A. Rasti et al., "Deep learning-based classification of retinal OCT images: A survey," Neural Computing and Applications, 2022.

[23] Obuli Sai Naren., Retinal OCT Image Classification - C8 [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/2736749 (2021)

[24] Z. Li, Y. Han, and X. Yang, "Multi-Fundus Diseases Classification Using Retinal Optical Coherence Tomography Images with Swin Transformer V2," *Journal of Imaging*, vol. 9, no. 10, p. 203, Sep. 2023, doi: 10.3390/jimaging9100203.