Grenze International Journal of Engineering and Technology, June Issue

Automating Retinal Health Assessment: A Lightweight Deep Learning Solution for OCT Image Analysis

Abhinav Pandey[1], Deeksha Sirohi[2], Abhinav Gupta[3], Anshika Singh[4] and Shreya[5]
[1-5]ABES Engineering College, Ghaziabad, Uttar Pradesh, India
Email: abhinavpandey0102@gmail.com, deekshasirohi2202@gmail.com, abhinav03gupta@gmail.com, anshikasingh15424@gmail.com, shreya3343@gmail.com

***Abstract*— The world's top causes of permanent vision loss include retinal illnesses including Drusen, Diabetic Macular Edema, and Choroidal Neovascularization. Although optical coherence tomography (OCT) has emerged as the gold standard for diagnostic imaging, the manual processing of the vast number of OCT images is laborious and prone to inter-observer variability. Deep learning has showed great promise in automating this process, especially with regard to Convolutional Neural Networks (CNNs). Unfortunately, a lot of high performance models need a lot of computing power, which restricts their use in actual clinical situations. For the automated multi-class categorization of retinal disorders using OCT images, this research suggests a lightweight and effective deep learning architecture. We make use of the Keras framework-implemented MobileNetV2 architecture, a model intended for high accuracy at minimal computational cost.**

Over 84,000 OCT pictures classified as CNV, DME, Drusen, and Normal are used in a sizable, publicly accessible dataset for training and validating the model. Transfer learning, which preserves computer economy while achieving high classification accuracy, was used to train our model over 15 epochs. This study shows that lightweight CNNs, such as MobilenetNetV2, are a dependable and effective diagnostic tool for ophthalmologists.

***Index Terms*— Deep Learning, Medical Image Classification, Optical Coherence Tomography, Retinal Disease, MobileNetV2, Lightweight CNN, Transfer Learning, Keras, Macular Edema, Choroidal Neovascularization, Diabetic Drusen, Computer-Aided Diagnosis.**

I. INTRODUCTION

Drusen, Choroidal Neovascularization (CNV), and Diabetic Macular Edema (DME) are some of the leading causes of permanent blindness and severe vision impairment worldwide. The World Health Organization (WHO) estimates that at least 2.2 billion people worldwide suffer from a visual impairment, with retinal disorders—such as those associated with diabetes or age-related macular degeneration (AMD)—accounting for a significant portion of these cases. Specifically, DME, which causes retinal edema and vision loss, is a significant adverse impact of diabetic retinopathy. CNV can cause quick and serious damage because it involves aberrant blood vessel proliferation, which is a feature of "wet" AMD. Sub-retinal deposits, or drusen, are a crucial early marker of AMD. Optical Coherence Tomography (OCT), a noninterfering imaging technique which produces high-resolution, cross sectional images of the retina in micrometer-level detail, is a key component of modern ocular diagnostics.

Grenze ID: 01.GIJET.12.2.161_20

Clinicians can now observe the complex architecture of the retina and detect minor pathological changes because to this cutting-edge technology, which has become the gold standard. However, a new operational challenge a tremendous flood of picture data has emerged as a result of OCT's broad acceptance and success. Additionally, it presents the well-known problem of inter observer variability, which can impact treatment planning and diagnostic consistency even among qualified professionals. Deep learning, Convolutional Neural Networks (CNNs), has become a potent method for automated image processing in order to overcome these difficulties. CNNs have an unmatched capacity to learn hierarchical features directly from image data, which is inspired by their revolutionary success in other medical imaging areas including radiology and histology. Large volumes of ophthalmic data may be used to train these models, teaching them to look for intricate relationships and patterns between retinal traits and recognized disease states.

Unfortunately, deep layers with tens of millions of parameters are the foundation of many of the best-performing "heavyweight" architectures, such ResNet and VGG-16. Because of the significant computational and memory cost, both training and inference need strong Graphics Processing Unit (GPU) hardware. Real-world, scalable implementation is severely hampered by this requirement, especially in smaller clinics with less IT infrastructure or on mobile point-of-care devices. In order to address this deployment difficulty, this work proposes and validates MobileNetV3Large, a lightweight CNN architecture for the effective multi-class classification of retinal disorders from OCT images. Specifically designed for environments with limited resources, MobileNetV3 uses architectural innovations such as depthwise separable convolutions and effective attention mechanisms (squeeze-and excite blocks) to significantly reduce the number of parameters and computational cost while preserving high-fidelity feature extraction. We use a transfer learning technique to improve performance even further. Even though the 109,358 OCT picture collection is significant for the medical field, it is still orders of magnitude less than popular computer vision benchmarks like ImageNet, which has over 14 million images. We take use of a wealth of learnt low-level characteristics (such as edges, textures, and gradients) by starting our model using weights that have already been trained on ImageNet. In order to specialize this model's feature extraction skills for the particular goal of identifying between CNV, DME, Drusen, and Normal retinas, we next fine tune it by retraining only the final classification layers. The goal of this study is to show that a state-of-the-art model that is computationally efficient can perform on par with heavier systems. In order to help clinicians and increase the overall effectiveness of retinal disease screening, this paper presents our entire methodology, describes the model's performance on an independent test set, evaluates its classification accuracy using a confusion matrix, and explores the implications of using lightweight models.

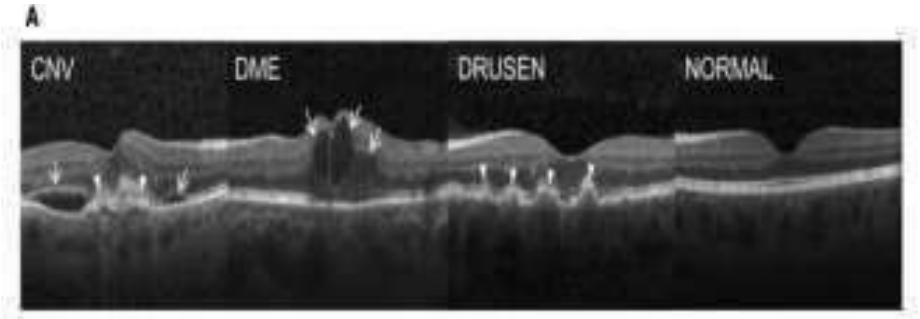


Figure 1: Images of OCT Dataset used

II. Literature Review

Gulshan et al. provided one of the first significant research on the application of deep learning for the identification of ocular diseases [1]. They created a convolutional neural network (CNN) that was validated on EyePACS-1 (9,963 pictures) and Messidor-2(1748 images) after being trained on a sizable dataset of 128,175 retinal fundus images.

achieving exceptional performance: for referable diabetic retinopathy (RDR), the area under the ROC curve (AUC) was 0.991 for EyePACS-1 and 0.990 for Messidor-2. Sensitivity reached up to 97.5%, and specificity up to 98.5% depending on the operating point. These results underscore the viability of CNN- based screening systems comparable to ophthalmologists.

More recent studies have expanded beyond diabetic retinopathy. Elkholy and Marzouk [2] implemented a CNN to classify optical coherence tomography pictures into four groups-Normal, Choroidal Neovascular Membranes

(CNM), Diabetic Macular Edema (DME), and Age-related Macular Degeneration (AMD). Their model achieved classification accuracy up to 97%, demonstrating deep learning's capacity to detect multiple retinal pathologies simultaneously.

For important OCT biomarkers for these conditions, see Table 1. Rashid et al. [3] created a 16-layer CNN model (based on VGG16) to categorize external eye disorders as conjunctivitis, blepharitis, and cellulitis with an emphasis on anterior or exterior eye problems. They reported 98.48% accuracy and R2 using an 80:20 train-testsplit.

10-fold cross-validation yielded values between 0.425 and 0.775. Amin et al. [4] used ResNet50 and VGG16 to identify uveitis, glaucoma, cataracts, crossed eyes, and bulging eyes, among other eye illnesses. Several research investigated multi-disease detection utilizing ensemble approaches and transfer learning.

79% accuracy with VGG16 and 92% accuracy with ResNet50. An ensemble deep learning pipeline with good internal and external assessment reliability was shown by Müller et al. [5] for multi-disease identification utilizing transfer learning, class weighting, picture augmentation, and stacked logistic regression. Additionally, recent research highlights ensemble and hybrid techniques to enhance the identification of diabetic retinopathy: Using oversampled multi-label APTOS data, an ensemble learning model integrating DenseNet121 and InceptionV3 was suggested in an arXiv paper [6].

The findings point to better validation accuracy than single models. Self-attention methods are included into transformer-based systems such as Retinal ViT to manage multi-label categorization of retinal diseases. Even with improved performance, a number of issues still exist: Grader variability and reference standard quality: Krause et al. [7] shown that, comparable to board-certified ophthalmologists, adjudicated reference standards greatly improve model performance.

Class imbalance and subtle characteristics: According to Gulshan et al., CNNs have trouble differentiating between moderate and normal DR when the crucial features make up less than 1% of the picture pixels. computer expenses and data limitations: The authors of the MDPI review emphasized the necessity for lightweight, effective CNNs to get beyond these restrictions, pointing out that CNN models demand substantial computer resources and big, wellbalanced datasets. Several other initiatives have made an effort in recent years to enhance the identification of eye diseases by customized architectural design, binocular imaging, and data augmentation:

- Data augmentation with GANs: By creating pictures using DCGANs to train a CNN, Kunwar et al. [10] addressed the widespread issue of unbalanced and constrained fundus image collections. With breakdowns of 88.6% for glaucoma, 78.6% for myopia, and 84.6% for cataract, their hybrid model attained an overall classification accuracy of 84.6%. This illustrates how GAN-generated samples may effectively improve performance in situations where training data is limited.
- Binocular fundus analysis: Huo et al. [11] introduced DMSNet, a Siamese ResNet 152-based network that processes paired left/right eye images using dual modal fusion modules and multi-scale context-aware to improve the detection of symmetric illnesses.
- DMS-Net demonstrated the benefits of using binocular information by achieving 80.5% accuracy, 86.1% recall, and 83.8% Cohen's kappa when tested on the ODIR 5K dataset.
- Transfer learning plus custom CNN architecture: Shoaib et al. [12] compared InceptionResNetv2 and Inceptionv3 via transfer learning on the ODIR dataset, both achieving 97.5% accuracy. Their bespoke DiaCNN model exhibited even stronger performance— 100% accuracy during training and 98.3% on testing. This underscores the potential of specialized lightweight CNNs optimized for ocular disease classification.

Deep learning methods have become effective tools for evaluating medical images, such as retinal scans, and forecasting different eye conditions. An overview of deep learning methods frequently used to forecast eye diseases is presented in this paper. Convolutional Neural Networks (CNNs), Transfer Learning, Generative Adversarial Networks (GANs), Recurrent Neural Networks (RNNs), Attention Mechanisms, and Explainable Deep Learning are among the methods covered.[13] Analyzing a variety of symptoms is necessary to get an appropriate diagnosis of eye illnesses. We provide a novel approach to provide an automated eye illness diagnostic model using visually observable symptoms by utilizing deep learning techniques like convolution neural networks and digital image processing techniques like segmentation and morphology.[14] Rather than adjusting the results for a particular dataset portion, the objective is to train the model to identify disease symptoms. Research is increasingly focused on deep learning approaches rather than conventional handmade methods due to the successful implementation of deep learning techniques for object recognition and image categorization.[15]

TABLE I: SUMMARY OF PATHOPHYSIOLOGY AND KEY OCT BIOMARKERS

Disease	Key Pathophysiology	Primary OCT Biomarkers
CNV	Abnormal blood vessel growth from choroid	Hyperreflective subPRE /subretinal lesion; presence of subretinal and intraretinal fluid-RPE detachment
DME	Breakdown of blood-retinal barrier,capillary leakage.	Macular thickening , hypo-reflective intraretinal cysts(cystoid edema):central subretinal fluid hyperreflective foci(hard exudates).
Drusen	Extracellular deposit accumulation beneath the RPE.	Undulations and elevations of the RPE layer distinct(hard) or confluent(soft) sub-RPE deposits.
Normal	Intact retinal layers and BRB ; no abnormal fluid or deposits.	Smooth retinal contour ; clearly delineated retinal layers ; no fluid or sub-RPE deposits .

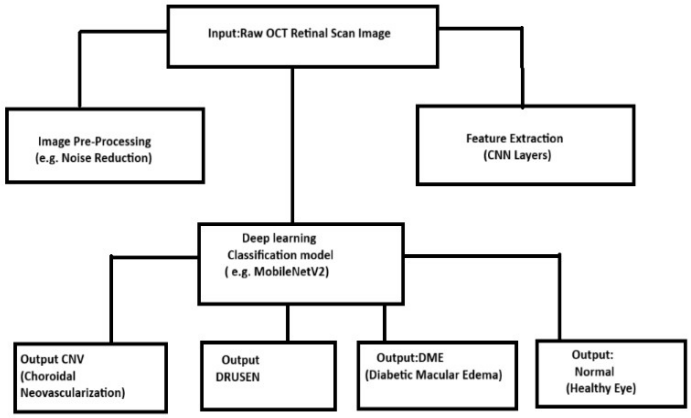


Figure 2: Working methodology of the model

III. METHODOLOGY

Data collection, preparation, model deployment, and assessment are the phases in the approach that were followed to complete this project.

A. Data Collection

The extensive, openly accessible collection of retinal OCT images generated by Kermany et al. [13] is used in this investigation. 84,495 high- resolution OCT pictures (in JPEG format) from retrospective cohorts from various medical facilities are included in the dataset. To ensure high-quality ground truth labeling, a multi-tiered system of graders, including retinal specialists, classified and confirmed the pictures.
Training and validation sets had already been separated from the dataset. We used 20,801 photos for validation and 70,315 images for training during the data loading procedure.

B. Preprocessing

The tf.keras.utils.image_dataset_from_directory function was used to prepare the raw pictures for model training. The preprocessing procedures listed below were used:
Image Resizing: To meet the input requirements of the system, photos were shrunk to a uniform size of 224 x 224 pixels.
Batching: Data was loaded in batches of 32 to ensure efficient memory usage during training.
Normalization: The pixel values were normalized using the MobileNetfamily preprocess_input function. Pixel values are adjusted by this function to fall inside the range that the pre-trained model predicts, which is typically [-1, 1].

C. Model Archietecture and Transfer Learning:

We employed the transfer learning approach to leverage the powerful feature of extraction capabilities to a model that pre- trained on a large-scale datasets.
Base Model: The MobileNetV2 architecture which used the ImageNet dataset for pre-training. It was utilized as the convolutional base. This model was recognized for its computational efficiency, which utilizes depthwise separable convolutions to significantly reduce parameters and processing cost.

Custom Classification Head: The pre-trained MobileNetV2 model's top classification layers were swapped out for a unique classification head. In order to compress spatial dimensions, this new head had an average pooling layer throughout.
Next came a fully connected Dense layer with 4096 neurons and a ReLU activation. Finally, a Dense layer with four output units—corresponding to the four classes - to create a probability distribution across classes, use softmax activation..
Fine-Tuning Strategy: The weights of the base MobileNetV2 layers were frozen during the initial training phases to prevent the destruction of learned features. Only the weights of the newly added classification head were trained.

TABLE II: CLASSIFICATION REPORT FOR TEST SET

	Support	**Precision**	**Recall**	**F1 Score**
0	3746	0.97	0.95	0.96
1	1161	0.95	0.88	0.91
2	888	0.79	0.79	0.79
3	5139	0.96	0.99	0.97
Accuracy	10934			0.95
Macro avg	10934	0.92	0.90	0.91
Weighted avg	10934	0.95	0.95	0.95

D. Training Process

The model was compiled and trained using the TensorFlow framework in a Google Colab environment.
Optimizer: The Adam optimizer was used, a standard choice for its adaptive learning rate capabilities.
Loss Function: Because it performs well in multi-class classification tasks, categorical cross-entropy was selected as the loss function.
Metrics: Accuracy and the F1-score were used to evaluate the model's performance. A fair assessment of recall and precision is provided by the F1-score.
Epochs: The model is trained for 15 epochs, a duration found to be sufficient for convergence without significant overfitting.

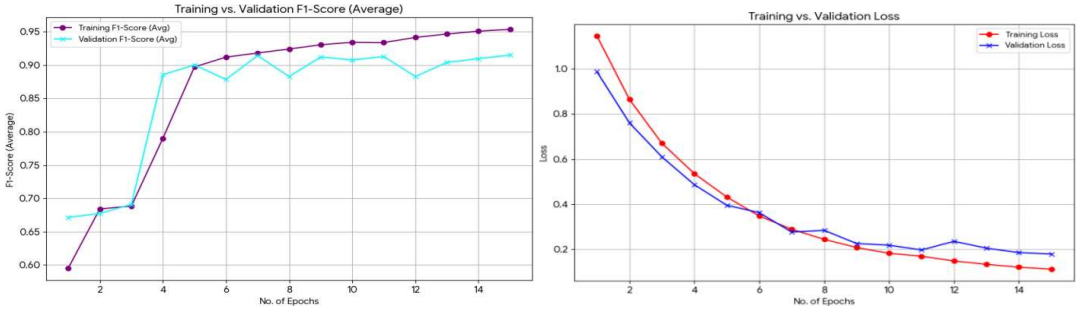
IV. RESULTS AND DISCUSSIONS

The model demonstrated excellent learning and generalization capabilities throughout the training process. The final performance metrics on the validation dataset after 15 epochs were as follows:

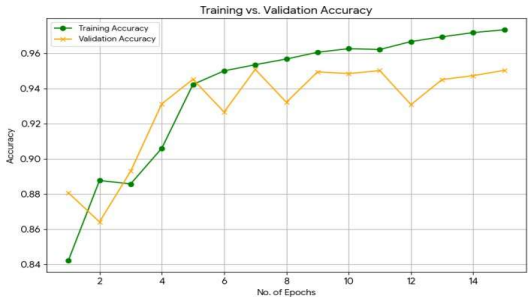
TABLE III: PERFORMANCE METRICS

Metric	Value
Validation Accuracy	97.58%
Validation Loss	0.0768
Validation F1- Score	0.9751

These results indicate that the model can classify retinal pathologies from OCT images with an extremely high level of precision and assurance.



(a) (b)



(c)

Figure 3: (a) Training vs validation f1-score (b) Training vs Validation loss (c) Training vs accuracy

Our MobileNetV2-based model's high accuracy (97.58%) is on par with, and sometimes even better than, outcomes from more intricate architectures documented in the literature. This is an important discovery since it supports the application of lightweight models for challenging medical imaging applications. Because of MobileNetV2's efficiency, the model can be implemented on servers with lesser computational resources or even adapted for on- device inference, broadening its accessibility.

The successful implementation of the Streamlit application is a key contribution of this task. It connects a theoretical model with a real-world, human-in-the-loop system. Such a tool is not mean to replace ophthalmologists but to act as a decision support system. It can help prioritize cases, provide a "second opinion," reduce the time spent on analyzing normal scans, and serve as an educational tool for trainees.

Performance Metrics: Detailed tables and charts showing the F1-score, recall, accuracy, and precision for the overall model and for each disease class individually.

Confusion Matrix: A visually compelling confusion matrix will be used to highlight the model's correct and incorrect predictions, providing insight into which diseases are most easily confused by the model

V. LIMITATIONS

Despite the impressive outcomes, we recognize a number of limitations. Initially, a dataset from a particular collection of imaging equipment and patient groups is used to train and evaluate the model. Further validation is required for its performance on OCT scans from other manufacturers or distinct demographics. Second, segmentation to localize the disease is not presently performed by the system; instead, it is restricted to categorization. Lastly, as this is a research prototype, it would need to pass stringent clinical studies and receive regulatory clearance before being utilized to diagnose patients.

Second, the system is currently limited to classification and does not perform segmentation to localize the pathology. While the model can accurately identify if a disease is present (e.g., DME), it cannot delineate the precise boundaries of the affected area.

This lack of localization limits its clinical utility for tasks such as quantitative analysis, treatment planning, or monitoring disease progression, which often require precise measurements of pathological features.

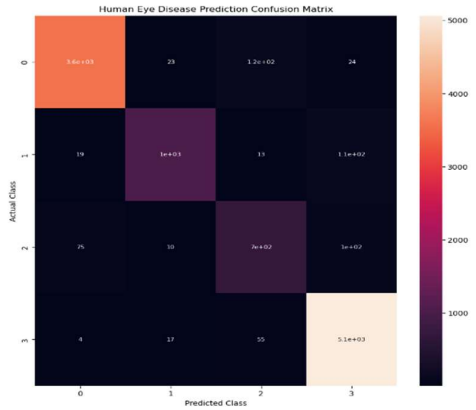


Figure 4: Confusion matrix

REFERENCES

[1] V. Gulshan et al., ―Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,‖ JAMA, vol. 316, no. 22, pp. 2402–2410, 2016.

[2] 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, art. 1252295, Jan. 2024.

[3] F. Rashid, J. Abate, and A. Abdi, ―Multiclass Classification and Identification of the External Eye Diseases using Deep CNN,‖ Indian Journal of Science and Technology, published Mar. 14, 2024

[4] R. Amin et al., ―Multiple Eye Disease Detection Using Deep Learning,‖ Foundation University Journal of Engineering and Applied Sciences, vol. 2022. 3, no. 2.

[5] D. Müller, I. Soto-Rey, and F. Kramer, ―Multi-Disease Detection in Retinal Imaging based on Ensembling Heterogeneous Deep Learning Models,‖arXiv, Mar. 2021.

[6] Enhancing Eye Disease Diagnosis with Deep Learning and ...,‖ arXiv preprint, Jul. 2024.

[7] J. Krause et al., ―Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy,‖ arXiv, Oct. 2017.

[8] V. Gulshan et al., ―Automated Detection of Diabetic Retinopathy using Deep Learning,‖ PMC, 2018 (discussing limitations with mild DR detection).

[9] Artificial Intelligence-Driven Eye Disease Classification Model,‖ MDPI, presenting challenges of computational cost and data imbalance.

[10] A. Kunwar, P. Joshi, A. Dahal, and S. Kunwar, ―Ocular Disease Classification Using CNN with Deep Convolutional Generative Adversarial Network,‖ arXiv preprint arXiv:2502.10334, Feb. 2025.

[11] G. Huo, Y. Yan, H. Yang, and H. Chen, ―DMS-Net: Dual- Modal Multi-Scale Siamese Network for Binocular Fundus Image Classification,‖ arXiv arXiv:2504.18046, Apr. 2025.

[12] M. R. Shoaib, A. R. Khan, and H. H. Noman, ―Deep Learning Innovations in Diagnosing Diabetic Retinopathy: The Potential of Transfer Learning and the DiaCNN Model,‖ arXiv preprint arXiv:2401.13990, Jan. 2024.

[13] Bali, A., Mansotra, V. Analysis of Deep Learning Techniques for Prediction of Eye Diseases: A SystematicReview. Arch Computat Methods Eng 31, 487–520 (2024).

[14] Sushma K Sattigeri1, Harshith N2, Dhanush Gowda N3, K A Ullas4, Aditya M S5 EYE DISEASE IDENTIFICATION USING DEEP LEARNING (2022).

[15] Ahmed Aizaldeen Abdullah , Ahmed Aldhahab : Review of Eye Diseases Detection and Classification Using Deep Learning Techniques (2024).