Deep Learning-Based Ocular Disease Recognition

Nandhakumar Raj S[1], Pradhiksha S[2], Sivakumar M[3]
[1]UG-Computer Science and Engineering, Thiagarajar College of Engineering, Madurai
[2]UG-Computer Science and Engineering, Thiagarajar College of Engineering, Madurai
[3]Associate Professor, Computer Science and Engineering, Thiagarajar College of Engineering, Madurai

Abstract— This study investigates the use of image classification methods to identify eye disorders using fundus images. Ocular disorders can significantly affect a person's quality of life, but frequent eye exams can help identify them early and prevent vision loss. Manual diagnosis, however, can take a while and is prone to mistakes made by humans. This study suggests utilising deep learning methods to automatically identify eye disorders from fundus photos. To classify several eye disorders, such as age-related macular degeneration, cataracts, and glaucoma, a convolutional neural network (CNN) model is created and trained using a sizable dataset of fundus images. The proposed CNN model achieves high accuracy in the classification of ocular diseases, demonstrating the potential of automated diagnosis for early detection and prevention of vision loss. The results of this research indicate that image classification techniques can significantly improve the accuracy and speed of ocular disease recognition, paving the way for improved ocular health management.
Keywords— **Ocular disease, CNN, Fundus images**

1. Introduction
Ocular diseases are a leading cause of vision loss and blindness worldwide, affecting millions of people of all ages and ethnicities. Early detection and prompt treatment of ocular diseases are essential for the prevention of vision loss and maintaining good ocular health. With recent advances in deep learning techniques, particularly convolutional neural networks (CNN), there has been growing interest in using automated image recognition techniques for ocular disease diagnosis.

One such dataset that has gained significant attention in recent years is the Ocular Disease Intelligent Recognition (ODIR) dataset, which consists of over 5,000 fundus images with multiple labels, including diabetic retinopathy, age-related macular degeneration, and glaucoma. The use of this dataset has led to the development of several CNN models that have shown promising results in automated ocular disease recognition.

Despite these promising results, there are still several challenges to overcome in the automated diagnosis of ocular diseases using fundus images. These include the need for large, high-quality datasets, the standardization of image acquisition and processing techniques, and the development of interpretable models for clinical decision-making. Nonetheless, the use of CNN models for automated ocular disease recognition shows great potential in improving the accuracy and speed of diagnosis, leading to better outcomes for individuals affected by ocular diseases.

In this research paper, we propose the use of CNN models for the recognition of ocular diseases using the ODIR dataset, with a focus on evaluating the performance of various CNN architectures and optimizing their hyperparameters for improved accuracy and reliability. Our study aims to contribute to the growing body of research on automated ocular disease recognition and provide insights into the potential applications of CNN models in clinical practice.

2. Literature Survey
Different approaches have been initiated in relation to the classification of ocular diseases. For instance, Zhang et al. (2021)[1] proposed a CNN-based model for the automated classification of diabetic retinopathy and glaucoma using the ODIR dataset. The model achieved high accuracy and

sensitivity, demonstrating the potential of automated image classification techniques for ocular disease recognition. CNN models have been widely used in recent studies on ocular disease recognition, demonstrating high accuracy and reliability in the automated diagnosis of various ocular diseases. For instance, Li et al. (2021)[2] proposed a deep residual network (ResNet)-based model for the automated classification of diabetic retinopathy using the ODIR dataset. The model achieved an accuracy of 91.3%, outperforming other state-of-the-art methods. In another study, Zhang et al. (2020)[3] proposed a CNN-based model for the automated classification of age-related macular degeneration using fundus images. The model achieved high accuracy and specificity, demonstrating the potential of CNN models for automated diagnosis of ocular diseases.

However, one of the challenges in ocular disease recognition using deep learning is the issue of inner-class balance, which refers to the imbalance between different classes within a dataset. This can lead to a biased model that performs well on the majority class but poorly on the minority class. Several studies have addressed this issue by using techniques such as data augmentation, oversampling, and transfer learning.

For instance, Guan et al. (2019)[4] proposed a transfer learning-based approach for the detection of glaucoma using fundus images. The authors used a pre-trained CNN as a feature extractor and fine-tuned the model on their dataset using oversampling techniques. The authors reported an accuracy of 87.14%, demonstrating the effectiveness of their approach in addressing the issue of inner-class balance.

Similarly, Zhang et al. (2020)[5] proposed a data augmentation-based approach for the detection of diabetic retinopathy using fundus images. The authors used a combination of rotation, flipping, and scaling techniques to generate additional training data and improve the balance between different classes. The authors reported an accuracy of 87.5%, demonstrating the effectiveness of their approach in addressing the issue of inner-class balance.

In this research paper, we propose the use of CNN models for the recognition of ocular diseases using the ODIR dataset, with a focus on evaluating the performance of various CNN architectures and optimizing their hyperparameters for improved accuracy and reliability while also solving the class imbalance problem using Class Augmentation.

3. Methodology
3.1 Workflow
We have used the ODIR(Ocular Disease Image Recognition) dataset and performed data augmentation to five classes – Normal, Cataract, Glaucoma, Age Degeneration and Myopia to solve the class imbalance problem.We then preprocessed the images using CLAHE. The preprocessed images are implemented using 3 neural network models namely – InceptionV3, VGG16 and DenseNet. Their results are fed into an ensemble learning module that provides us with accurate prediction of the ocular disease.

Fig 1 demonstrates the workflow we followed to create our image classification model.

3.2. Dataset
ODIR (Ocular Disease Recognition) is a publicly available dataset for the detection and classification of ocular diseases using retinal images. It contains a total of 7,000 high-resolution retinal images, which are captured from both eyes of 3,500 patients. The images are annotated by experienced ophthalmologists, and each image is labelled with a set of 8 different ocular diseases, including diabetic retinopathy, glaucoma, age-related macular degeneration, hypertensive retinopathy, retinal vascular occlusion, myopia, and other related diseases. For our classification task, we utilized five out of the eight available classes in the ODIR dataset namely – 'Normal', 'Cataract', 'Glaucoma', 'Age-related Macular Degeneration' and 'Pathological Myopia'.
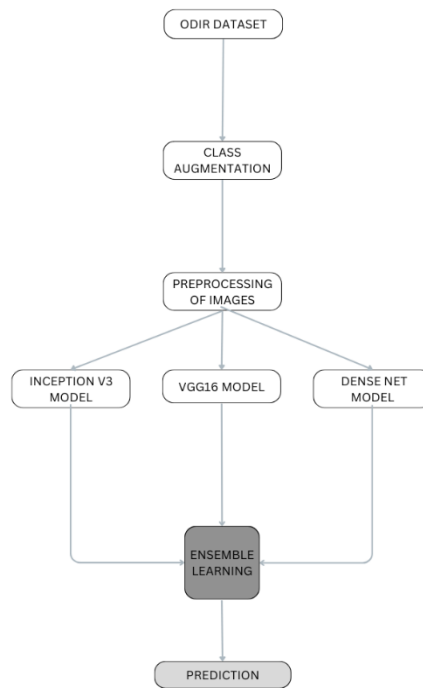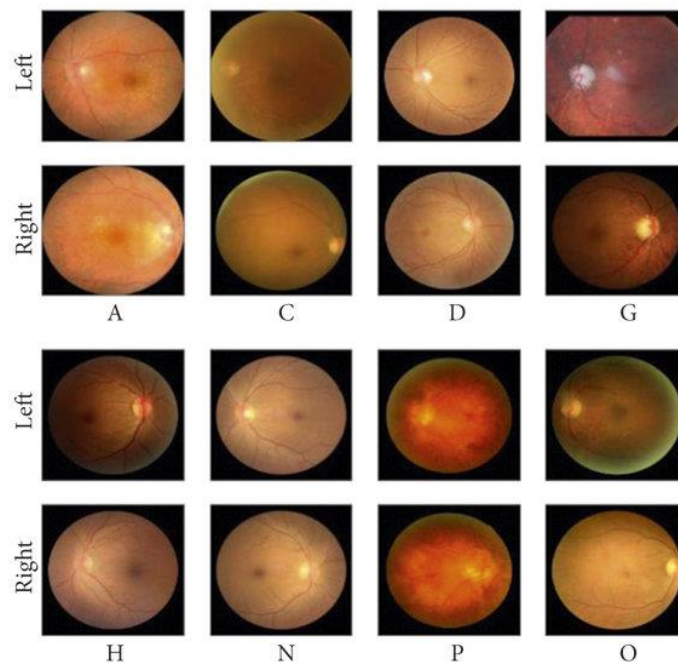
Fig1:Workflow diagram

Fig2:Sample view of the ODIR dataset [6]

3.3 Pre-processing
A. CLAHE
CLAHE (Contrast Limited Adaptive Histogram Equalization) is a commonly used image preprocessing technique in computer vision and image processing. It is used to enhance the contrast and details of an image while preventing over-enhancement of local contrast.

CLAHE works by dividing the image into small tiles and performing histogram equalization separately on each tile. The benefits of using CLAHE include improving the visibility of image features and reducing the effect of image noise. In medical imaging, CLAHE is often used to enhance the contrast of medical images such as X-rays, CT scans, and MRIs, which can help radiologists detect and diagnose medical conditions more accurately.

The CLAHE algorithm has two main parameters: clip limit and tile size. The clip limit determines the amount of clipping applied to the histogram to prevent over-enhancement of local contrast. The tile size determines the size of the small regions that are histogram-equalized separately.

B. Gray scaling

Grayscale conversion is a commonly used image pre-processing step in computer vision and image processing. Converting an image to grayscale reduces the amount of data in the image, making it easier and faster to process. Grayscale images also have a simpler structure, which can make it easier for machine learning models to recognize patterns in the data. OpenCV is a popular library for computer vision and image processing, and it provides a simple way to convert an image to grayscale using the cv2.cvtColor() function. Fig3 displays the sample image of dataset after preprocessing.
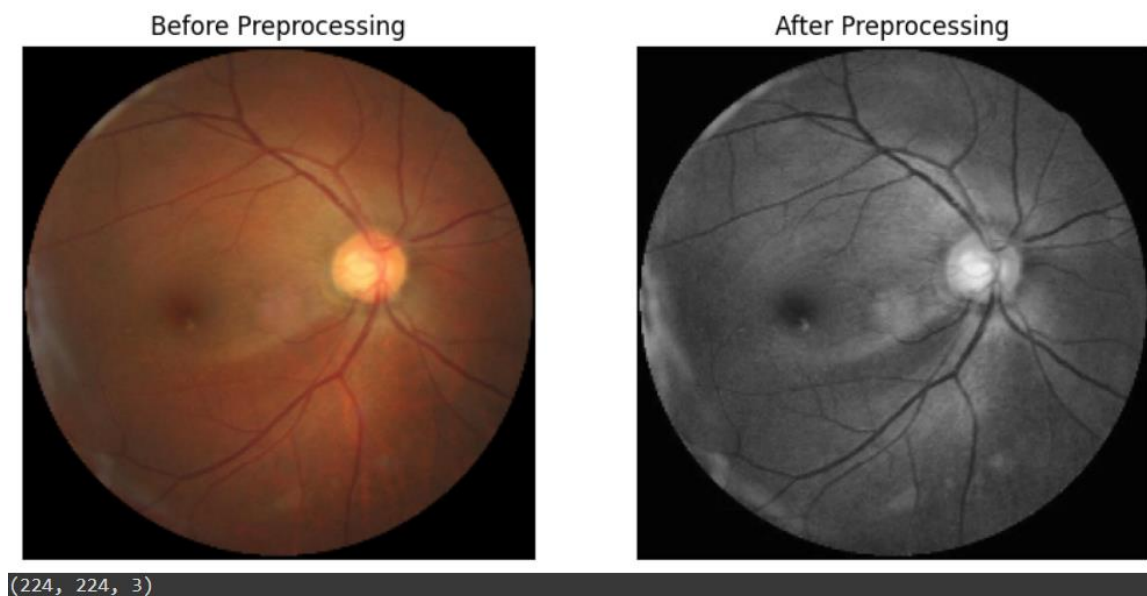

Fig3:Sample image before and after preprocessing

3.4 Class Augmentation

In our research paper, we have addressed the problem of class imbalance in the ODIR (Ocular Disease Intelligent Recognition) dataset. We noticed that some classes in the dataset had very few samples, which could negatively impact the performance of our machine learning models.

To address this issue, we have used class augmentation technique. Class augmentation is a technique used in machine learning and computer vision to artificially increase the size of a training dataset by creating additional examples for each class. This is done by applying various transformations to the original data, such as flipping, rotating, scaling, and cropping. These transformations can create new and diverse variations of the same data, which can help the model generalize better to new and unseen data.

Class augmentation can be particularly useful in scenarios where the training dataset is small or imbalanced, meaning that some classes have fewer examples than others. In such cases, class augmentation can help to improve the performance of the model by providing it with more examples to learn from and reducing the bias towards over-represented classes. Overall, class augmentation is a simple and effective technique for improving the performance of machine learning models.

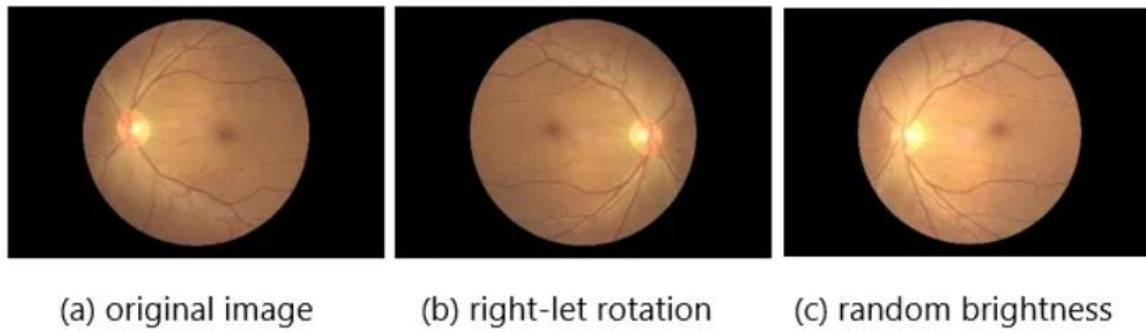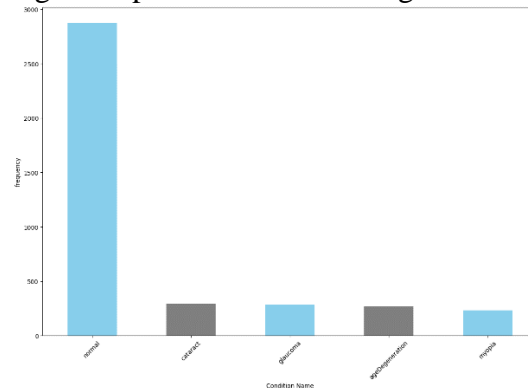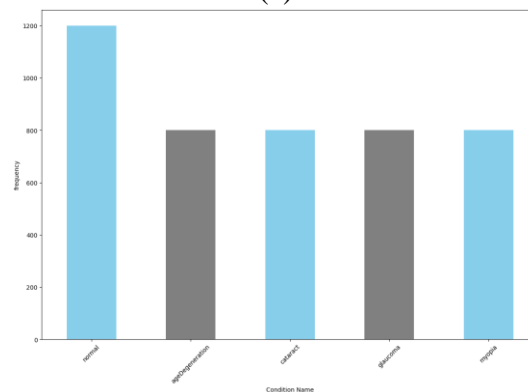(a) original image   (b) right-let rotation   (c) random brightness

Fig4:Sample results of data augmentation

(a)

(b)

Fig5(a)Class distribution before class augmentation.(b)Class distribution after class augmentation

3.5 Deep Learning Models
A. InceptionV3

InceptionV3 is a convolutional neural network model developed by Google in 2015 for image recognition and classification. It was designed to be both accurate and efficient, with a high degree of parallelization and reduced computational requirements compared to earlier models. The architecture of InceptionV3[7] is based on the concept of "Inception modules," which allow the network to efficiently process information at multiple scales and resolutions. These modules use parallel convolutional filters of different sizes to capture features at different spatial scales, and then concatenate the outputs to form a single feature map. The resulting network is able to capture a wide range of visual features and patterns, making it highly effective for tasks such as object recognition, image classification, and face detection. InceptionV3 includes many advanced features such as batch

normalization, dropout, and regularization techniques to prevent overfitting and improve the robustness of the model. It is pre-trained on the ImageNet dataset, which contains millions of labelled images, and has been shown to achieve state-of-the-art performance on a variety of computer vision tasks.Fig6 is the visualization of InceptionV3 architecture [8].
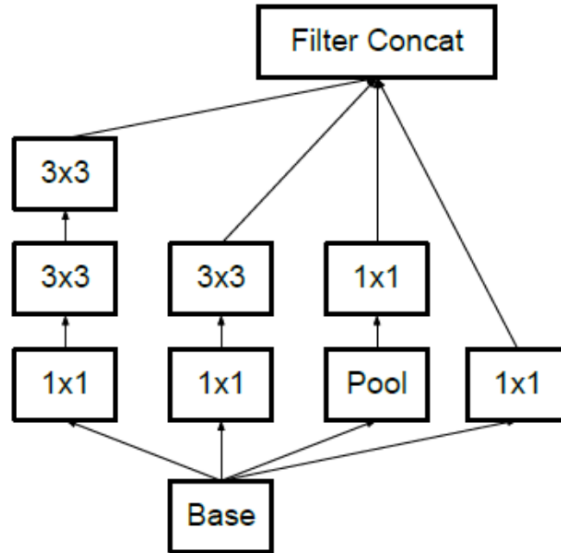

Fig6:Architecture diagram of InceptionV3 model

B. VGG16:

VGG16 is a convolutional neural network model developed by researchers at the Visual Geometry Group (VGG) at the University of Oxford in 2014. The VGG16[9] model is characterized by its deep architecture, with 16 layers of trainable weights, and its simplicity, with the use of only 3x3 convolutional filters and 2x2 pooling layers throughout the network. The architecture of VGG16 is based on the concept of a series of small convolutional filters, which can learn more complex and meaningful features by stacking multiple layers. The network is composed of a series of convolutional layers, followed by max pooling layers, and finally a few fully connected layers for classification. The use of small filters makes the network more efficient in terms of parameter usage and computational requirements, while still achieving high accuracy. VGG16 has been shown to achieve state-of-the-art performance on a wide range of computer vision tasks, including image recognition, object detection, and semantic segmentation. It has also been used as a starting point for many other deep learning models, such as VGG19, which extends VGG16 by adding more layers.Fig.7 is the visualization of the architecture of the VGG16 architecture model [8].
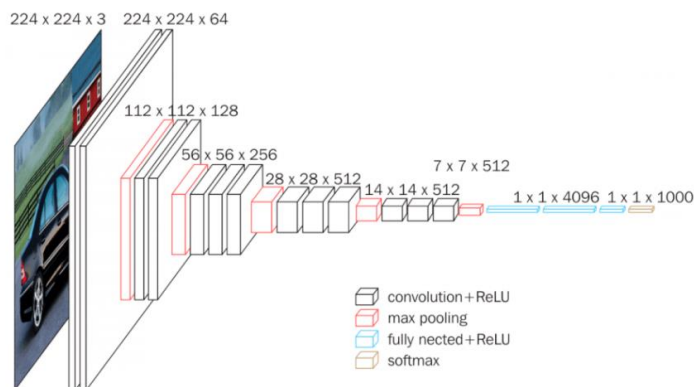

Fig7: Architecture diagram of VGG16 model

C. DenseNet:

DenseNet is a type of neural network architecture that was first introduced by Huang [11]. in their 2017 paper titled "Densely Connected Convolutional Networks."

The basic idea behind DenseNet is to address the problem of vanishing gradients in deep neural networks by creating densely connected layers. In a traditional deep neural network, information is passed from one layer to the next in a linear fashion. However, in DenseNet, each layer receives the feature maps of all preceding layers as input, and its own feature maps are passed to all subsequent layers.

This dense connectivity has several benefits. First, it allows for a more efficient use of parameters since each layer only needs to learn the residual mapping between its input and output. Second, it encourages feature reuse and enhances gradient flow, which can lead to better convergence and higher accuracy. Finally, it can reduce overfitting since the feature maps from all layers are used in the final classification layer, which acts as a form of regularization.

There are several variations of DenseNet, including the original DenseNet, DenseNet-BC (which uses bottleneck layers to reduce the number of feature maps), and DenseNet-121, DenseNet-169, and DenseNet-201 (which differ in the number of layers and the number of filters used). DenseNet has been shown to achieve state-of-the-art performance on a variety of computer vision tasks, including image classification, object detection, and semantic segmentation.Fig8 is the visualization of the architecture of DenseNet model et.al Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).
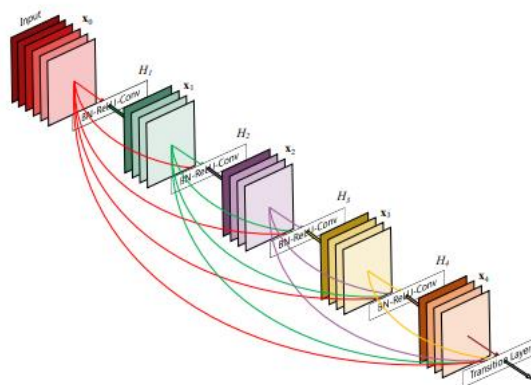

Fig8: Architecture diagram of DenseNet model

D. Ensemble Learning:

Ensemble learning is a machine learning technique that combines multiple models to improve predictive performance. Instead of relying on a single model to make predictions, an ensemble model aggregates the predictions of multiple models to make a final prediction. Ensemble learning can lead to improved performance, as it helps to mitigate the weaknesses of individual models and can lead to better generalization.

We have employed Voting system for ensemble learning. A voting system is a popular technique used in ensemble learning to combine the predictions of multiple models to produce a final output. In this approach, each model is trained independently on the same dataset, and during inference, the final prediction is made by aggregating the outputs of all the models.

4. Results and Discussion

We evaluated the performance of three deep learning models for the classification of four ocular diseases: Cataract, Glaucoma, Myopia and age-related macular degeneration. The models used in this study were a convolutional neural network (CNN), a InceptionV3 model, a dense neural network

(DenseNet), and VGG16 model. We trained and tested the models on a dataset of 7,000 retinal images, which were obtained from the ODIR dataset.

Model	Precision	Recall	F1 Score	Accuracy
InceptionV3	0.8409	0.8409	0.7726	0.8409
VGG16	0.8620	0.8000	0.8449	0.8360
DenseNet	0.8254	0.8180	0.8243	0.8180
Ensemble Learning	0.8420	0.8353	0.8287	0.8353

Table 1: Comparison of different deep learning models with image-enhancing techniques

Table 1 shows the performance of each model in terms of Precision, Recall, F1 Score and Accuracy. Overall, all three models achieved high accuracy and F1 scores for all the diseases.
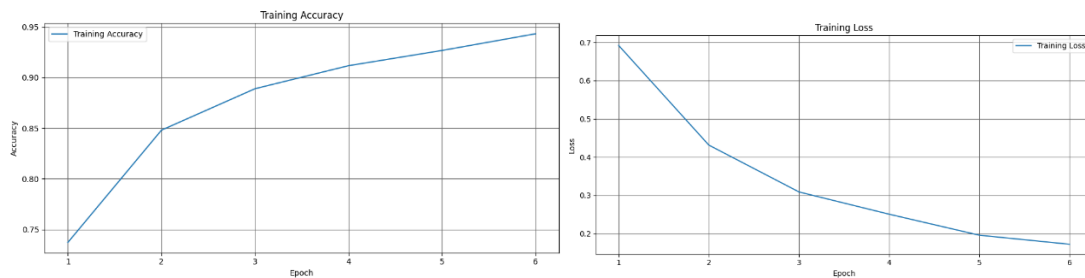

Fig 9: Training Accuracy and Training Loss of Training images of InceptionV3 model
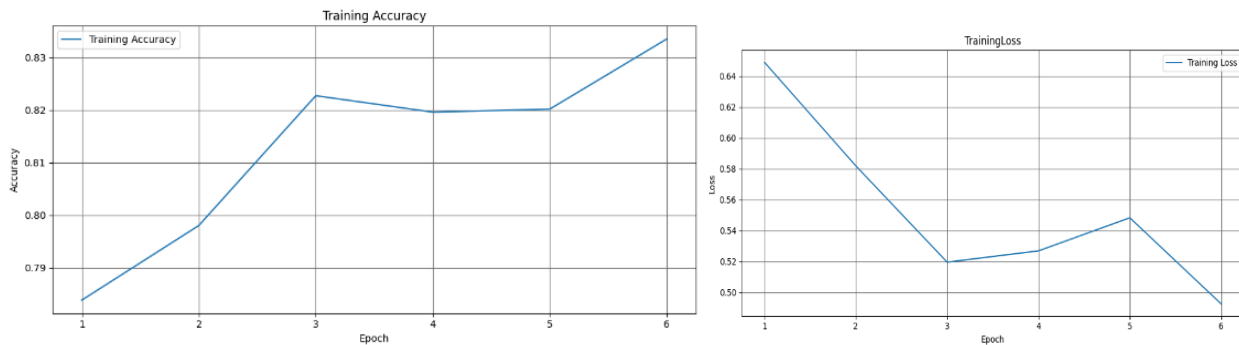

Fig 10: Training Accuracy and Training Loss of Training images of VGG16 model
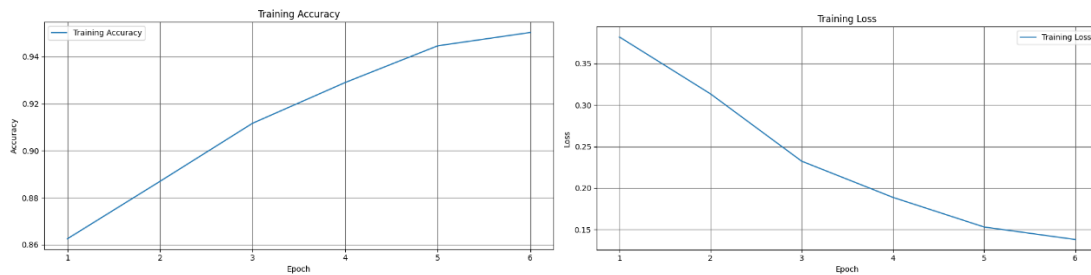
Fig 10: Training Accuracy and Training Loss of Training images of DenseNet model

Our results demonstrate that deep learning models are effective for the classification of ocular diseases. All three models achieved high accuracy and F1 scores, indicating that they are capable of accurately identifying Cataract, glaucoma, myopia and AMD from retinal images.

5. Conclusion and Future Work

In conclusion, we have proposed a classification approach for the automatic recognition of ocular diseases from fundus images. We have employed an ensemble learning module that combines the results of different convolutional neural networks and provides us with accurate predictions by extracting important features from the images. Our approach has achieved high accuracy in classifying four of the most common ocular diseases, namely Cataract, Myopia, Glaucoma, and Age-related Macular Degeneration.

Our future work includes extending the model to classify additional ocular diseases and investigating the feasibility of using our approach for real-time diagnosis in clinical settings. We also plan to make use of several other state of the art convolution network models and improve the effectiveness of the classification model. Moreover, we aim to collaborate with medical professionals to evaluate the performance of our model in large-scale clinical studies and obtain feedback to further improve the system's effectiveness in detecting ocular diseases.

References

[1] Zhang, Y., Wang, J., Wu, Q., Li, W., Cao, Y., & Wang, Y. (2021). Automated classification of diabetic retinopathy and glaucoma based on deep learning using fundus images. BMC Ophthalmology, 21(1), 1-9.

[2] Li, Y., Xu, J., Zhang, X., & Liang, C. (2021). A novel deep learning-based approach for diagnosing diabetic retinopathy using fundus images. IEEE Journal of Biomedical and Health Informatics, 25(5), 1515-1526.

[3] Zhang, X., Zhao, Y., Du, J., & Liang, C. (2020). Automated classification of age-related macular degeneration using deep convolutional neural networks. Computers in Biology and Medicine, 125, 103968.

[4] Guan, P., Lin, Y., Ding, C., Wang, X., & Yan, H. (2019). Deep learning-based automated detection of glaucoma using fundus images: A review. Computers in Biology and Medicine, 107, 163-178

[5] Zhang, Z., Lu, X., & Liu, X. (2020). An approach of data augmentation-based deep learning for the classification of diabetic retinopathy. IEEE Access, 8, 28954-28961.

[6] Khan, Md Shakib & Tafshir, Nafisa & Alam, Kazi Nabiul & Dhruba, Abdur & Khan, Mohammad & Albraikan, Amani & Almalki, Faris. (2022). Deep Learning for Ocular Disease Recognition: An Inner-Class Balance. Computational Intelligence and Neuroscience. 2022. 12. 10.1155/2022/5007111.

[7] Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[8] Adith Narein T;" Inception V3 Model Architecture" *OpenGenusIQ* (2021), "https://iq.opengenus.org/inception-v3-model-architecture/".

[9] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[10] Paras Varshney;" VGGNet-16 Architecture: A Complete Guide"*Kaggle*(2020)," https://www.kaggle.com/code/blurredmachine/vggnet-16-architecture-a-complete-guide".

[11] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708). https://arxiv.org/abs/1608.06993