Retinal Disease Multiclass Classification

Bhagesh Gaur
bhagesh20558@iiitd.ac.in

Madhava Krishna
madhava20217@iiitd.ac.in

Sanyam Goyal
sanyam20116@iiitd.ac.in

Abstract

Retinal diseases like diabetes, glaucoma and age-related macular degeneration (ARMD), among others, can cause loss of vision and permanent blindness. As a result, it is imperative that early detection and diagnosis be carried out effectively and efficiently.
We aim to tackle a multilabel classification problem on the ODIR-2019 dataset, made public by Peking University, China. We implement known and new deep-learning algorithms on the dataset after proper processing and examine how transformations and model layouts change the performance on the dataset.

1 Introduction

Fundus diseases affect millions of people worldwide. They include diseases like diabetic retinopathy, age-related macular degeneration, cataract, glaucoma and hypertensive retinopathy, among others. India has a high prevalence rate of 16.9% for diabetic retinopathy (Vashist, 2021), and 12 million people suffer from glaucoma (George R, 2010). An estimated 100 million people go blind from cataract and 8 million from age-related macular degeneration (ARMD) every year (Sen M, 2022).
As a result, it is imperative that early diagnosis and corrective measures be taken to reduce the impact these diseases have on human life. An overburdened medical system during the COVID-19 pandemic caused a lot of harm indirectly (Raman et al., 2021), and care must be taken to ensure that such a situation does not occur.
Machine learning-based tools can help ease the load on the healthcare sector by assisting in diagnosis to improve accessibility and, in some cases, the accuracy of the treatment method prescribed. Deep-learning-based techniques allow for an even greater level of flexibility in the diagnostic procedure, allowing for multimodal diagnosis and taking inputs from a variety of sources (Ngiam et al., 2011).
We explore various deep-learning algorithms and techniques on the ODIR-19 dataset, provided by the Peking University, China, which has 8 classes of ocular diseases with corresponding fundus images. (ODIR2019, 2019)

2 Related Work

Corbella's MSc dissertation(Coll Corbilla, 2020) had done single-eye multilabel classification. The repository is available and featured transfer-learning methods on models like VGG16, VGG19 and InceptionV3. The full method is well-documented and available on the repository.
Wang *et al.* (Wang et al., 2020) created a similar classifier but using the EfficientNet architecture. They compared various transfer-learning approaches and gave metrics for evaluation. They even used an ensemble-learning approach wherein they co-used EfficientNet-based classifiers on RGB and Grayscale histogram-equalised images respectively, averaging the output at the terminus for class-wise prediction in a multilabel format.
Jinke *et al.* created a multilabel classifier using graph convolutional networks (Lin et al., 2021). They use the ODIR, SSL and GTest datasets, but did not mix the datasets. Li *et al.* used attention mechanisms on binocular fundus images on the ODIR dataset (Li et al., 2022). They used a ResNet50 model in a multilabel fashion.

3 Dataset details

The dataset was taken from the ODIR-2019 challenge by Peking University, China. The dataset instance used was taken from Kaggle (Larxel, 2021). The dataset has a main directory comprised of pre-processed images of dimension 512 x 512, in RGB colour format. It also has two additional directories for the test-train split (pre-split), but we do not use that in our methods. There is an excel file having the dataset description. The following columns were present:

1. ID: the identification number of the patient. IDs range from 0 to 4784.

2. Age: the age of the patient. Minimum 1, maximum 91, median 59.

3. Sex: the sex of the patient (binary: Male/Female). 1885 male patients, and 1615 female patients.

4. Left-Fundus: the image names of the left fundus of the patient.

5. Right-Fundus: the image names of the right fundus of the patient.

6. Left-diagnostic keywords: diagnostic keywords for the left eye.

7. Right-diagnostic keywords: diagnostic keywords for the right eye.

8. Diagnostic columns: There are columns named as "N", "D", "G", "C", "A", "H", "M", "O", denoting the diagnosis for a **patient**.

    - N: Normal eye, normal diagnosis
    - D: Diabetes (diabetic retinopathy)
    - G: Glaucoma
    - C: Cataract
    - A: Age-related macular degeneration
    - H: Hypertension
    - M: Myopia
    - O: Other diseases aside from the above ones.

    The presence of a '1' in these column denotes presence of the disease for the patient and a '0' signifies absence.

3.1 Preprocessing stages

We focused on a single-eye image diagnosis (feeding one eye at a time) instead of a double-eye diagnosis (feeding both images at once). However, the labels were for a patient, instead of an eye, which implied that the defects present in the right eye may not be observed in the left eye. In addition, a patient may have multiple defects, which require parsing the diagnostic labels provided to determine the class.
With assistance in code by Jordi Corbella (Corbella, 2020), we were able to process each eye independently. There were some images that were not present in the dataset, so we excluded them.

InceptionV3 Parameters		
Hyperparameter	In the paper	Ours
Input size	244	299
Transforms	Cropping, mirroring, resize	Cropping, mirroring, resize
Batch size	32	128
Epochs	100	100
Early Stopping Patience	5	5
Optimizer (SGD)	lr = 0.01 decay = 1e-6 mom. = 0.9	lr = 0.01 decay = 1e-6 mom. = 0.9
Loss	Binary Crossentropy	Binary Crossentropy

Table 1: Inception V3 Parameters

In addition, Corbella was able to find out and remove 'bad' images, which had a lower quality than the rest of the images. We blacklisted images as specified in his GitHub repository (low-quality and unavailable images in the CSV).

4 Experimental Setup

We implemented the EfficientNet B3 architecture from the paper by Wang *et al.* (Wang et al., 2020), and InceptionV3 from the paper by Corbilla (Coll Corbilla, 2020).

Training was done on Google Colab notebooks and on a system with a Ryzen 9 CPU and an RTX 3090 graphics card.

For InceptionV3, we flipped right-eye images horizontally to create left-eye-like images. We took the input size as 299x299 in an RGB format (the dissertation took a 224x244 image size); no data augmentation was done. The other parameters are given in table 1.
For EfficientNet B3, we performed ensembling as given in the paper. One of the models was for RGB colour images and the other used images which underwent histogram equalisation to enhance them. The parameters are given in table 2. We ran the code with and without horizontal flipping (as done in InceptionV3 runs). The models were saved and had early-stopping implemented.

EfficientNet B3 Parameters		
Hyperparameter	In the paper	Ours
Input size	299 and 448	299
Transforms	Rotation by 90° or 45°, translation	Horizontal flip*, Rotation by 90° or 45°,
Batch size	15	16
Epochs	30	30 or 100
Early Stopping Patience	Yes	5
Optimizer (Adam)	lr = 0.001	lr = 0.001
Loss	Binary Crossentropy	Binary Crossentropy

Table 2: EfficientNet B3 training parameters. We used horizontal flip for one of the two ensembled models (both of the ensembles were trained on horizontal flipped images, as done in InceptionV3).

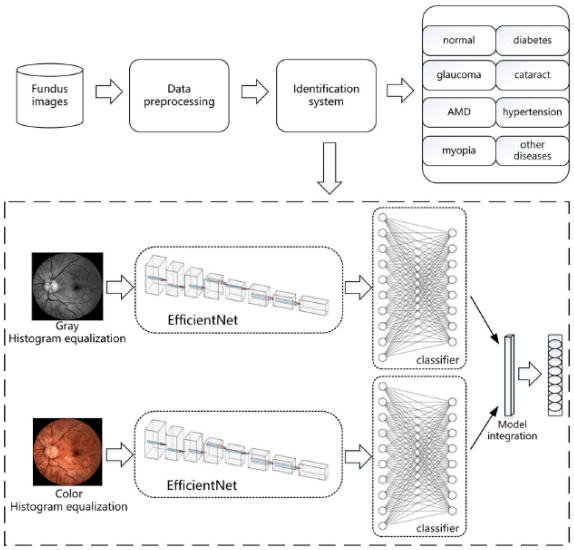

Figure 1: Ensembled Efficient Net Architecture

5 Observation and Future Work

5.1 Observations

The values reported by the papers had reported the validation set images. We took the best results from the papers. The EfficientNetV3 paper had ensembled models trained on 448x488 sized images, however, ours was on 299x299 images for computational reasons.

Overall, we observe close results to the papers, even outclassing the EfficientNetV3 paper but still within error margins. However, we observe that the preprocessed ensemble we implemented did perform poorly.

For InceptionV3, the accuracy, precision and Cohen's kappa are close to the paper implementation, but the the F1, precision and AUC scores fall short.

For EfficientNetV3, the flipped-image results were horrible, as we saw after trying multiple times. However, the non-flipped-image results were very close and even outclassed the model accuracy, precision, AUC and Cohen's Kappa score, falling short marginally in recall and trailing a lot in the F1-score benchmarks.

The results can be found in table 3.

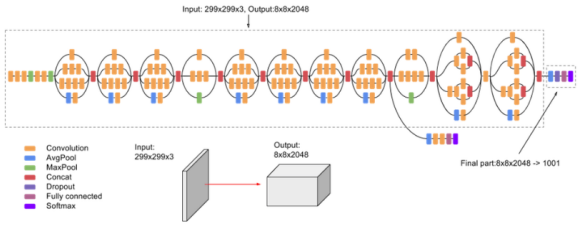

Figure 2: InceptionV3 Architecture

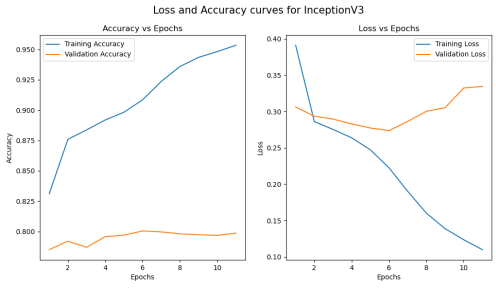

Figure 3: Loss curves for InceptionV3

5.2 Future Work

Right now, we are not taking the other elements (age and sex) into inference, and only taking one-eye into account while inferring decisions. We aim to use both eyes for identifying correlated fea-

Model Performance					
Metrics	**InceptionV3 Reported (val)**	**InceptionV3 Obtained**	**EfficientNet B3 Ensemble Reported**	**EfficientNet B3 Ensemble Obtained (Preproc)**	**EfficientNet B3 Ensemble Obtained (No Preproc)**
Accuracy	0.8984	0.897	0.92	0.874	0.929
Precision	0.6021	0.620	0.71	0.370	0.802
Recall	0.552	0.459	0.66	0.005	0.578
F1-Score	0.8984	0.528	0.89	0.010	0.672
AUC	0.8838	0.710	0.74	0.502	0.779
Kappa	0.5186	0.472	0.52	0.010	0.672

Table 3: Model performance on EfficientNetB3 and InceptionV3 reported and obtained. Our EfficientNet results with preproc had horizontal flipping (as done in EfficientNetV3) and the one without preproc had no horizontal flipping.

tures for better diagnosis of a patient in a binocular decision-making platform.

We aim to explore other modalities and integrate other datasets into the analysis and implement other, newer models.

References

Jordi Coll Corbilla. 2020. Reconeixement intel·ligent de malalties oculars mitjançant arquitectures d'aprenentatge profound.

Jordi Corbella. 2020. Ocular disease intelligent recognition through deep learning architectures. shorturl.at/oxJQZ.

Vijaya L George R, Ve RS. 2010. Glaucoma in india: estimated burden of disease. *J Glaucoma*, 19(6):391–7.

Larxel. 2021. Ocular disease recognition. https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k.

Zhenwei Li, Mengying Xu, Xiaoli Yang, and Yanqi Han. 2022. Multi-label fundus image classification using attention mechanisms and feature fusion. *Micromachines*, 13(6).

Jinke Lin, Qingling Cai, and Manying Lin. 2021. Multi-label classification of fundus images with graph convolutional network and self-supervised learning. *IEEE Signal Processing Letters*, 28:454–458.

Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y Ng. 2011. Multimodal deep learning. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 689–696.

ODIR2019. 2019. https://odir2019.grand-challenge.org/.

Rajiv Raman, Ramachandran Rajalakshmi, Janani Surya, Radha Ramakrishnan, Sobha Sivaprasad, Dolores Conroy, Jitendra Pal Thethi, V Mohan, and Gopalakrishnan Netuveli. 2021. Impact on health and provision of healthcare services during the covid-19 lockdown in india: a multicentre cross-sectional study. *BMJ Open*, 11(1).

Honavar SG Sen M. 2022. Eye care for all.

Suraj S; Gupta Vivek; Manna Souvik; Gupta Noopur; Shamanna B R1; Bhardwaj Amit; Kumar Atul; Gupta Promila2 Vashist, Praveen; Senjam. 2021. Prevalence of diabetic retinopathy in india: Results from the national survey 2015-19. 69(11):3087–3094.

Jing Wang, Liu Yang, Zhanqiang Huo, Weifeng He, and Junwei Luo. 2020. Multi-label classification of fundus images with efficientnet. *IEEE Access*, 8:212499–212508.