Retinal Fundus Multi-Disease Image Classification

Deependra Singh[1] **Saksham Agarwal**[2]
[1]SPS [2]SBS NISER, Bhubaneswar
deependra.singh@niser.ac.in
saksham.agarwal@niser.ac.in

Abstract

We are making a mulri-label classification model for a set of retinal diseases (total 45) using a dataset of 3200 images. Based on the literature review, we found that the best model to apply is convolutional neural networks. For this project, we have applied two specific CNN architectures, one based on a multi images classification model for 20 labels, and the other is a transfer leanrning approach using various pre trained models like VGG19, ResNet50, etc. For midway, we had two goals which were implementation of baselines and finding new datasets. As of midway, we are done with the implementation of baselines, a CNN model and a VGG19 model, with their hyperparameter tunings still left to do. The first model have gave an accuracy of 96% while the VGG model with preset values got a 40% accuracy without fine-tuning.

1 Introduction

A large portion of our population suffer from eye diseases at some point of their lives. These diseases are widespread across India in both rural and urban areas. Unfortunately, India has a scarcity of trained doctors that could diagnose these diseases. Their availability is not evenly distributed and are highly concentrated in urban areas. The goal of our project is to explore the possibility of making a working model that could augment the unavailablity of doctors in rural areas by making accurate predictions for these diseases.

1.1 Literature Review

This review covers three different research papers focused on the development of deep learning-based diagnostic tools for various retinal diseases using fundus images.

The first paper describes the development of a diagnostic tool using convolutional neural networks (CNNs) to classify retinal diseases based on fundus images.[1] The researchers used three different CNN models (VGG19, ResNet50, and GoogLeNet Inception v3) to classify images into two and nine classes. They achieved high accuracy for both classifications, with VGG19 achieving the highest accuracy of 99.12% for the two-class classification and ResNet50 getting 87.42% accuracy for 9 class classification . However, the accuracy improvement using the AI tool was not significant compared to without the tool. Moreover, there were some limitations with the dataset, such as different resolutions and unmatched age of patients.

There was slightly higher accuracy of diagnosis by residents using AI tool but the significant difference was the reduction of time taken in diagnosing the desease (30% to 70%).

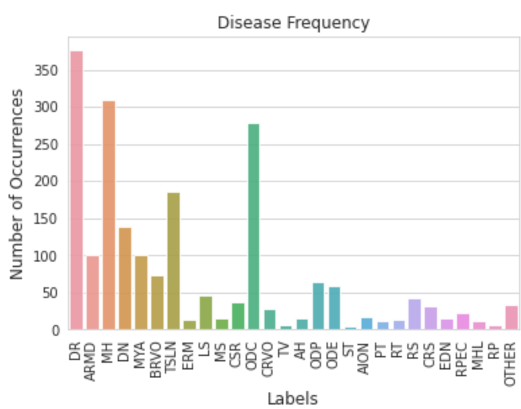
The second paper used a transfer learning approach with the VGG19 model to classify fundus images into three classes: normal, glaucoma, and diabetic retinopathy.[2] The researchers achieved an accuracy of 93.58% with a learning rate of 0.001 and 150 epochs.

The third paper focused on multi-label classification of retinal diseases using transformer-based architecture (C tran).[3] The researchers prepared their custom-made dataset by combining multiple publicly available datasets and pre-processing them for quality images. The model achieved high accuracy (AUC score) for most of the unique labels ($> 90\%$), but two labels showed lower accuracy.

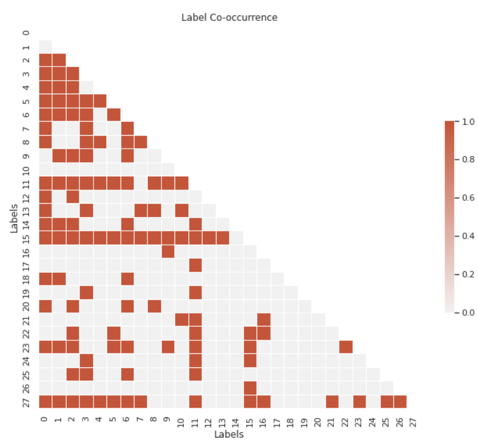
Overall, all three papers indicate that deep learning-based diagnostic tools for retinal diseases can achieve high accuracy and reduce the time taken for diagnosis. However, the limitations of the datasets and some classes with low accuracy show that more work needs to be done to improve the performance of these models.

2 Data Analysis

- RFMiD dataset of 3200 images. 1920 for training and 640 each for validation and testing. These images are marked with 45 or 28 labels. One label being for the normal/abnormal case and the rest being for the diseases.
- We performed exploratory data analysis on the data to gain better insight of the list of minority classes and label co-occurence.



(a) Fundus Image for a healthy eye



(b) Label Co-occurence

- The frequency analysis will be utilised in the sampling techniques and the co-occurence might come handy in the progressive transfer learning approach.

3 Preprocessing

The original fundus images were in resolution $4,288 \times 2,848$ pixels in rgb format and the size was rescaled to 50×50 and 100×100. We then used keras ImageDataGenerator method for augmenting the data. Since ours in a medical data, we have to be very careful regarding augmentation since the alignment of images may be important. We inspected the dataset and find that a significant portion of the images were present in flipped positions (right and left eye) with respect to each other, slight rotations and brightness. This motivated us to apply the various augmentation techniques on the dataset to a certain degree, to make the feature extractor invariant of orientation and position of the object in an image. The zooming of images was omitted. Tuning the augmentation parameters is yet to be done. Thereby we generated a dataset with more examples.

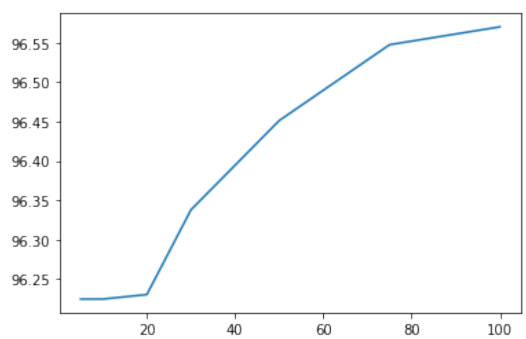
4 Model 1

4.1 Architecture

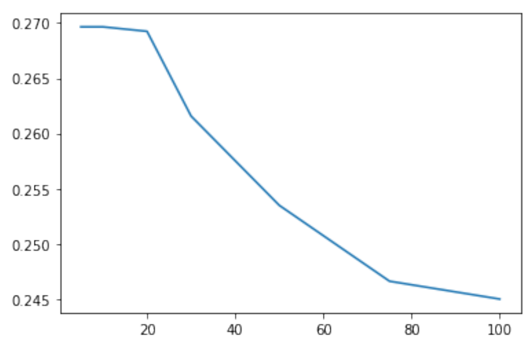
Layer (type)	Output Shape	Param
conv2d (Conv2D)	(None, 96, 96, 16)	1216
max_pooling2d (MaxPooling2D)	(None, 48, 48, 16)	0
dropout (Dropout)	(None, 48, 48, 16)	0
conv2d_1 (Conv2D)	(None, 44, 44, 32)	12832
max_pooling2d_1 (MaxPooling 2D)	(None, 22, 22, 32)	0
dropout_1 (Dropout)	(None, 22, 22, 32)	0
conv2d_2 (Conv2D)	(None, 18, 18, 64)	51264
max_pooling2d_2 (MaxPooling 2D)	(None, 9, 9, 64)	0
dropout_2 (Dropout)	(None, 9, 9, 64)	0
conv2d_3 (Conv2D)	(None, 5, 5, 64)	102464
max_pooling2d_3 (MaxPooling 2D)	(None, 2, 2, 64)	0
dropout_3 (Dropout)	(None, 2, 2, 64)	0
flatten (Flatten)	(None, 256)	0
dense (Dense)	(None, 128)	32896
dropout_4 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 64)	8256
dropout_5 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 28)	1820

4.2 Model Training and Baseline Results

- We have tried running the algorithm for threshold 0.5 to 0.9, and the best accuracy was obtained for threshold greater than 0.7.
- We test the model by rescaling rescaling from 50×50 to 100×100 giving an improvement from 0.32 to 0.86 accuracy for 29 labels.
- However, only a few majority classes had good accuracy, none of the positive cases of minority classes was matched correctly thereby giving a high accuracy. So we changed the labels to 28 by removing disease risk.
- The model showed a slight improvement to 0.96 accuracy for 0.8 threshold. However, it is still had minority class problem.
- We plotted the variation of accuracy and hamming loss with epochs.



(a) Accuracy vs Epoch



(b) Loss vs Epoch

5 Model 2: VGG19

5.1 Architecture

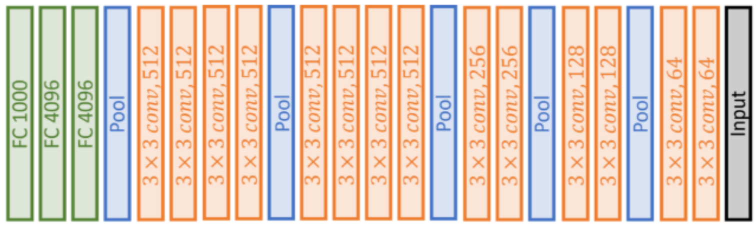
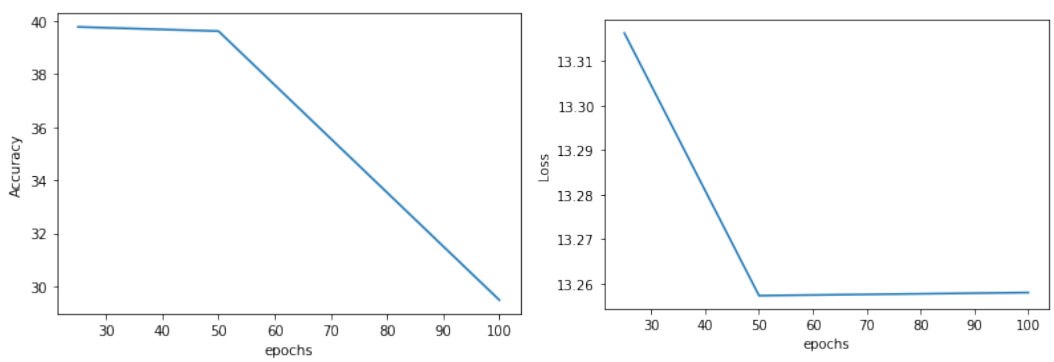


Figure 3: VGG19 architecture
(Source: https://datahacker.rs/deep-learning-vgg-16-vs-vgg-19/)

The size of the input layer is $100 \times 100 \times 3$ and after the sequence of 3×3 convolutional layers followed by a 256 node FC layer with relu activation function and 28 nodes FC layers with sigmoid function for classification.

5.2 Model training and Baseline Results

- We have tried running the algorithm for threshold 0.5, batch size 32 and a learning rate of 0.001, as of now.
- The accuracy if around 40% for a threshold for 0.5. But we know from previous examples the threshold has to be greater. The majority class problem still remains.
- We plotted the variation of accuracy and loss with epochs for threshold 0.5.



(a) Accuracy vs Epoch

(b) Loss vs Epoch

6 Post Midway Work

- Using Sampling methods for handling the minority class problem.
- Find more evaluation metrics and plot corresponding graphs for the problem.
- Apply ResNet50 model for classification. And fine tune both ResNet and VGG for our data.
- Continue on with progressive transfer learning approach in both these models.
- Alternatively use the MuRED data set, for the implementation of transformers approach. [3]
- Addition of Gluacoma label by combining two datasets using Label Superset Transform. And seeing if converting the problem to a multi class classification shows an improvement.
- Hyperparameter Tuning: The important ones are as follows-:

- Number of labels. We will try to sequentially increase the number of labels from majority class to minority class and see how trend for accuracy and loss function.
- Optimization for the augmentation parameters. We will use the first paper as reference.
- Optimization for image size, both resizing and cropping, followed by data cleaning.

7 References

[1] Kim, K. M., Heo, T. Y., Kim, A., Kim, J., Han, K. J., Yun, J., & Min, J. K. (2021). Development of a fundus image-based deep learning diagnostic tool for various retinal diseases. Journal of Personalized Medicine, 11(5), 321.
[2] Das, A., Giri, R., Chourasia, G., & Bala, A. A. (2019, July). Classification of retinal diseases using transfer learning approach. In 2019 International conference on communication and electronics systems (ICCES) (pp. 2080-2084). IEEE.
[3] Rodriguez, M. A., AlMarzouqi, H., & Liatsis, P. (2022). Multi-label Retinal Disease Classification Using Transformers. IEEE Journal of Biomedical and Health Informatics.
[4] Guo, C., Yu, M., & Li, J. (2021). Prediction of different eye diseases based on fundus photography via deep transfer learning. Journal of Clinical Medicine, 10(23), 5481.

7.1 Code files and Other Resources

- Model 1
- Model 2: VGG19
- Data file
- Multi Label Image Classification Blog