Exploratory Analysis on the Design and Implementation of Machine Learning Models for the Detection of Diabetic Retinopathy

COMP0172: Artificial Intelligence for Biomedicine and Healthcare

Name: Osman Hussien Fadl Ali
Student ID: 23083161
Date: 23/11/2023
Project Repository: GitHub Repository

Table of Contents

Chapter 1: Introduction

1.1. Diabetic Retinopathy

Diabetic Retinopathy (DR) is a diabetic disease of the eye in which the retinal blood vessels in the eye are damaged due to long-standing diabetes conditions [1]. Millions of people around the world suffer from diabetic retinopathy, the leading cause of blindness among working aged adults. It was reported by the International Diabetes Federation (IDF) that the number of diabetic patients will increase to 552 million by 2034, underscoring the urgency for effective DR screening methods [1]. DR progresses through four stages, each escalating in severity and risk to vision:

1. *Mild Non-Proliferative Retinopathy*: This is where small swellings in the tiny blood vessels will be formed [2].
2. *Moderate Non-Proliferative Retinopathy*: As the disease progresses, some of the blood vessels that nourish the retina become blocked [2].
3. *Severe Non-Proliferative Retinopathy:* In this stage, a significant number of retinal blood vessels experience occlusion, leading to a substantial reduction in blood supply to various retinal regions. The affected areas of the retina begin to show sign of ischemia (lack of oxygen) such as bleeding of the veins [2].
4. *Proliferative Diabetic Retinopathy*: At this advanced stage, the vasoproliferative factors produced by the retina begin to trigger the growth of new blood vessels. These new blood vessels are fragile and abnormal [2].

The image below illustrates the progression of Diabetic Retinopathy, ranging from a healthy retina with no signs of the disease to the advanced stage of proliferative DR, where the risk of vision loss becomes significantly heightened:

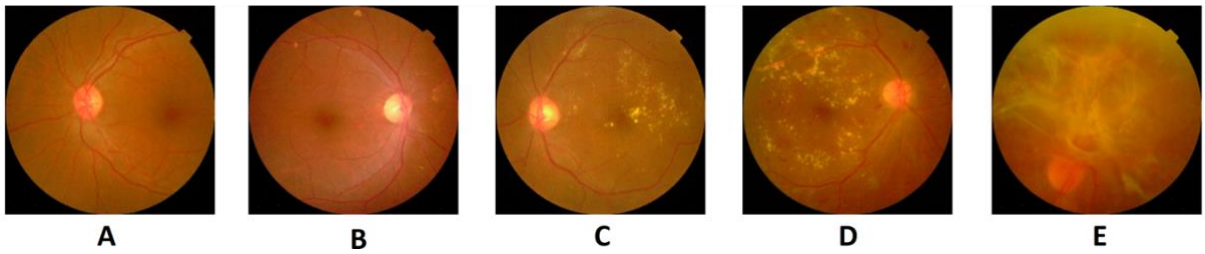

Figure 1: Stages of Diabetic Retinopathy – (A) without DR, (B) mild, (C) moderate, (D) severe, (E) Proliferate [5]

1.2. Machine Learning for DR Detection

Early detection and treatments of DR are therefore crucial steps in preventing irreversible blindness, yet traditional diagnostic methods are resource-intensive, requiring skilled ophthalmologists and extensive examination time. These challenges are particularly acute in rural areas like India, where access to specialised medical care is limited. The dataset central to this study is derived from the Aravind Eye Hospital in India [3]. Aravind technicians travel to these rural areas to capture images and then rely on highly trained doctors to review the images and provide diagnosis. This process, while effective, is not scalable given the limited number of specialists and the growing number of patients in rural areas in India.

This is where Machine Learning (ML), particularly deep learning techniques like convolutional neural networks (CNN), emerges as a transformative solution. By training models on extensive datasets of retinal images labelled with various stages of DR, ML algorithms can learn to identify subtle patterns and indicators of the disease that might elude even experienced clinicians. This capability would enable a more rapid, accurate, and accessible screening process, thus scaling the efforts of Aravind technicians and helping patients in remote areas.

1.3. Personal Motivation

I come from a country, Sudan, where the majority of the population do not have access to public or private healthcare. Having personally observed the hardships faced by those in rural communities, I feel a deep and personal connection with this project. The prospect of leveraging technology to extend healthcare services to those traditionally underserved is both inspiring and motivating for me.

Chapter 2: Dataset Exploration

2.1. APTOS Dataset

The APTOS dataset in this study consists of a large set of retina images taken using fundus photography under a variety of imaging conditions. Each image has a label indicating the severity stage of DR on a scale of 0 to 4:

Class Label	DR Severity
0	No DR
1	Mild DR
2	Moderate DR
3	Severe
4	Proliferate DR

Table 1: Class Labels for Each Severity Stage in the Dataset

This training dataset encompasses 3633 images, each evaluated and assigned a corresponding class label by a medical professional. The class distribution is described below:

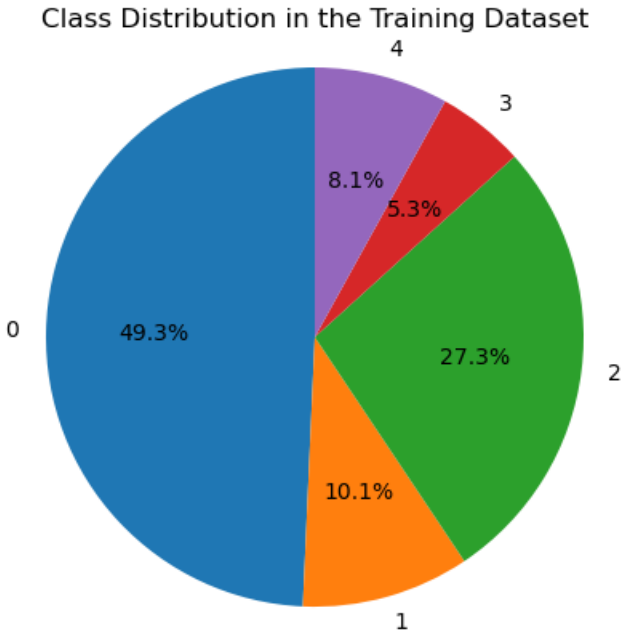

Figure 2: Pie Chart Showing the Class Distribution in the DR Dataset

From the pie chart above, there is a clear imbalance in the distribution of image classes across the training set. Almost half of the images are classified as Class 0, indicating no presence of DR. Class 2 (moderate DR) holds a middle ground in terms of representation but classes 1,3, and 4 are underrepresented in the training set. This could have an impact on the training and performance of the ML model, as ML models tend to bias towards the majority class. In this case, it could become proficient in identifying healthy retinas but could struggle in detecting DR or differentiating between different stages of DR in the images.

2.2. Image Analysis

To carefully analyse the diversity of the data, one must examine the different variations of retina images in the dataset:

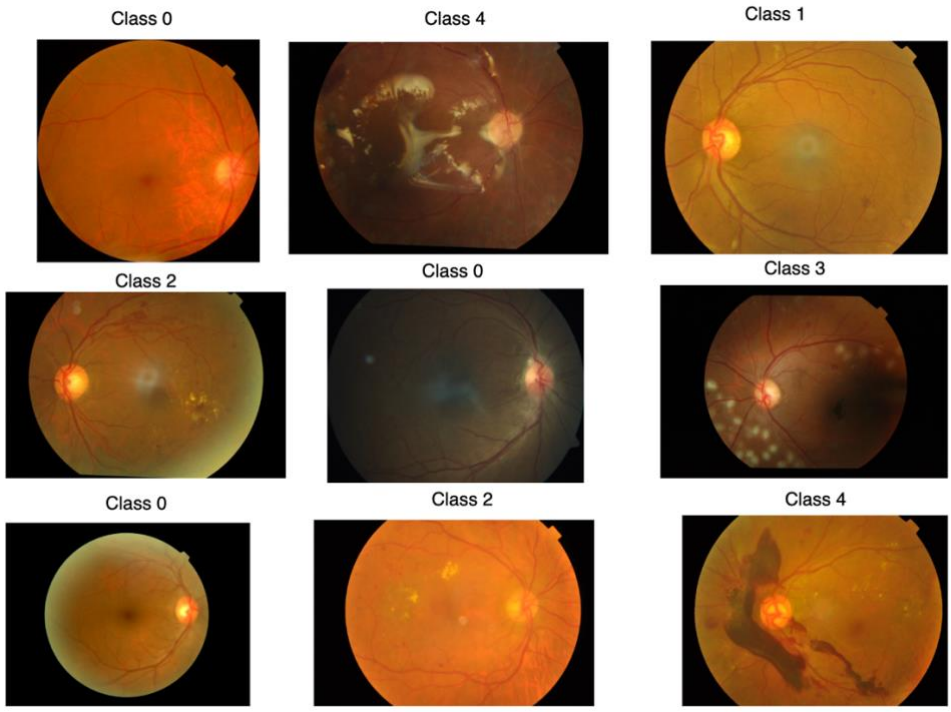

Figure 3: Diversity in Sampled Retinal Images

As seen above, the sampled images exhibit a range of sizes, shapes, and colours, with potential artifacts and varying exposure levels. These variations arise because the retina images in the dataset were gathered from multiple clinics in India using a variety of cameras over an extended period of time, which introduces further variation. To illustrate this further, the figure below shows the distribution of brightness in the first 100 images:

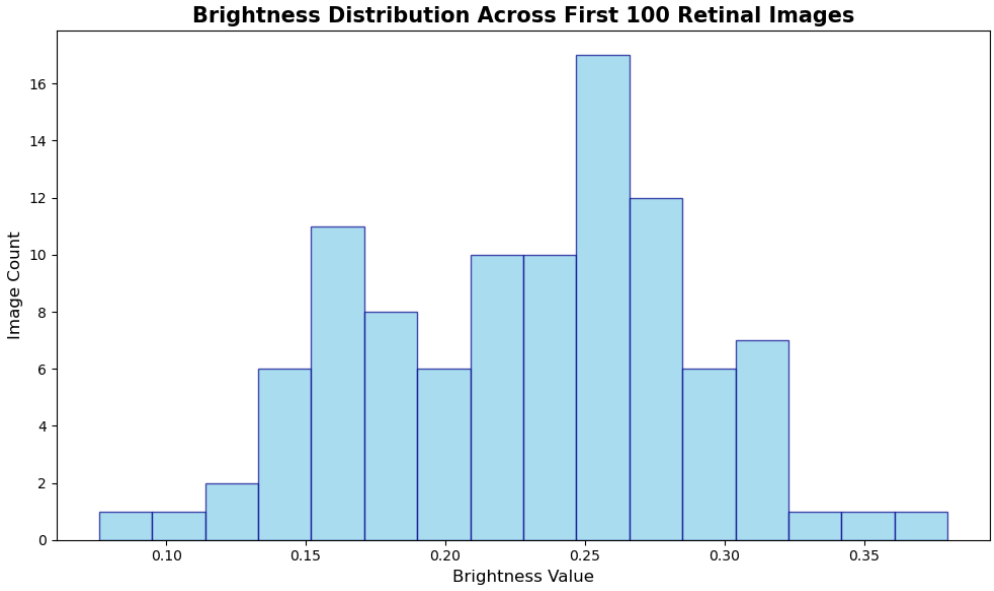

Figure 4: Histogram of Mean Brightness Values

Within the initial subset of 100 images, there is a noticeable variance in brightness, with levels spanning from 0.10 to 0.40 on a scale where 0 represents complete darkness and 1 denotes full brightness. These values showcase a broad spectrum of illumination conditions, accentuating the inherent variability within the dataset. Like any real-world dataset, there will also be noise present in both the images and the labels. To address all of this, it will be essential to implement systematic pre-processing steps that normalize the brightness levels and mitigate the impact of any artifacts on the machine learning model's performance.

Chapter 3: ML Design Methodology and Implementation

3.1. Data Pre-processing

The previous section highlighted the disparities in size, noise, lighting, contrast, and exposure levels of the retina images in the dataset. These variations pose a challenge to automated ML systems, and therefore it is crucial to take systematic pre-processing steps that would bridge the gap between raw data acquisition and model training. The processing steps taken to standardize the retina images were:

- **Resizing:** The fundus images in the dataset were generally very large, and hence all images were resized to 224x224 pixels – the input size of the CNN used in this project. This allows for batch processing and removes the variability in image scale.
- **Gaussian Blur:** A gaussian blur was applied to all images, which helped smooth out the images and reduce the high-frequency noise. This helped improve the model's ability to identify relevant DR patterns present in classes 1 to 4.
- **Greyscale Conversion:** Images were converted to greyscale format to reduce computational complexity and allow the model to focus on extracting structural features on the images like the retinal blood vessels rather than colour information.
- **Normalization:** The pixel values in the images were normalised to a range of 0 and 1. This helps stabilize the learning process by ensuring the gradient updates during the backpropagation stage are not very large.
- **CLAHE (Contrast Limited Adaptive Histogram Equalization):** CLAHE is an advanced image processing technique that improves image contrast and enhances the definitions of edges in different regions of the image [4]. Using CLAHE to enhance retinal image was proven to be very useful in extracting retinal features according a Setiawan's paper on Retinal Image Enhancement [4]. This is depicted in the image below:

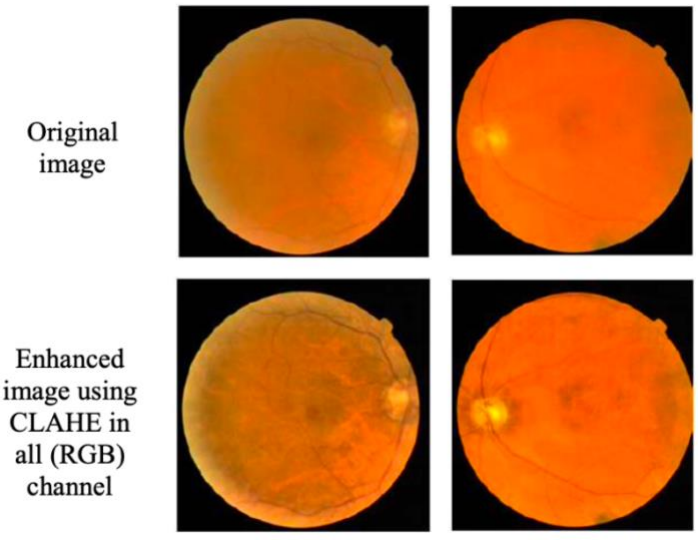

Figure 5: Contrast Enhancement with CLAHE [4]

The image below visually highlights all the pre-processing steps taken to ensure the retinal images in the dataset are standardised and ready to be inputted in the ML model:

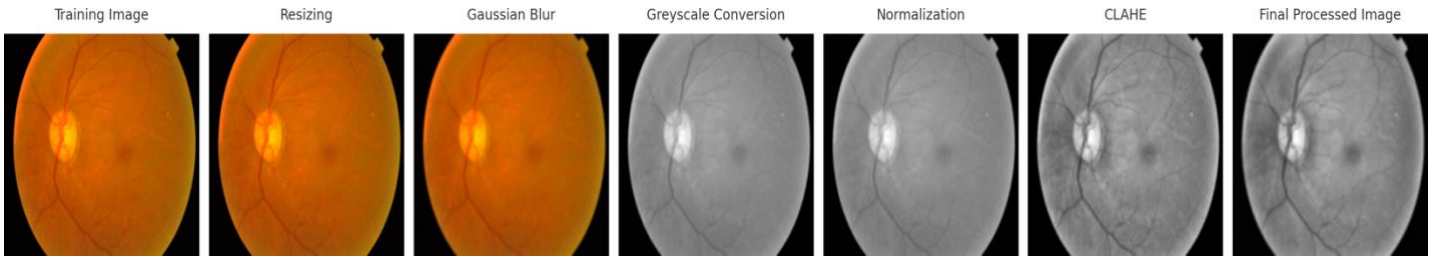

Figure 6: All Pre-processing steps from Resizing and Blurring to Contrast Enhancement

3.2. Machine Learning Implementation: Transfer Learning

To address the challenges in detecting DR, this project employs a strategic transfer learning approach by utilizing a pre-trained CNN – the DenseNet model, which has a densely connected network architecture. DenseNet's architecture facilitates feature reuse throughout the network, making it a suitable candidate for the task of identifying intricate patterns in retinal images indicative of DR. The variant used in this project is the DenseNet-121, which has the architecture below:

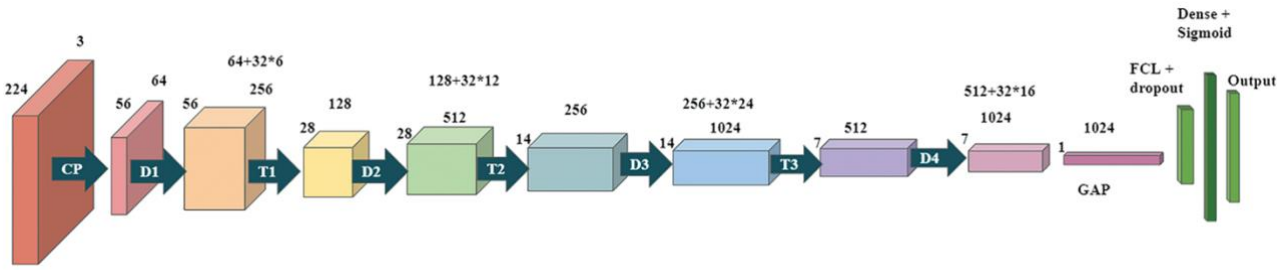

Figure 7: Proposed Dense CNN Model Based on DenseNet-121 [6]

The figure above shows that the DenseNet is made of blocks and transition layers. Each layer is directly connected to each other layer in a feed-forward fashion. This means that the feature maps learned at each layer of the DenseNet model are concatenated and passed on to all subsequent layers. This allows the network to access both the original input data and the deep features learned throughout the network. This structure not only encourages feature reuse, which is particularly beneficial in medical imaging where subtle features can be critical, but also significantly reduces the number of parameters when compared to other deep networks.

To adapt the DenseNet model to the learning task at hand, a few adjustments were made. Given that the processed retina images are in greyscale, the input layer was adapted to accept single channel images, instead of three-channel RGB images. To retain the learned features from ImageNet – a database used to train algorithms in complex image recognition tasks – the layers of the pre-trained DenseNet121 model were frozen. This helped preserve the intricate patterns and knowledge that the network had already acquired. After the pre-trained layers, custom layers were added including a global average pooling layer followed by a dense layer with 1024 neurons and ReLU activation to learn higher-level features specific to the retinal images. A final dense layer with a softmax activation function was added to produce the probability distribution across the five DR severity classes, where each class probability lies between 0 and 1, summing to 1 across all classes.

When it was time to train the model, it was noted that the APTOS test set is unlabelled, and was deemed unsuitable for evaluating model performance. Therefore, the training set was split into a 80:20 (training, test) split. The model was compiled with Adam optimizer and a learning rate of 0.0001, using categorical cross-entropy as the loss function and the labels were one-hot encoded. The model was initially set to train for 50 epochs but ended up training for 45 epochs as the loss stopped improving. The training process took nearly 3 hours.

Chapter 4: Model Evaluation and Discussion

4.1. Training Progression

Model training was evaluated by tracking the progression of training loss and accuracy across successive epochs:

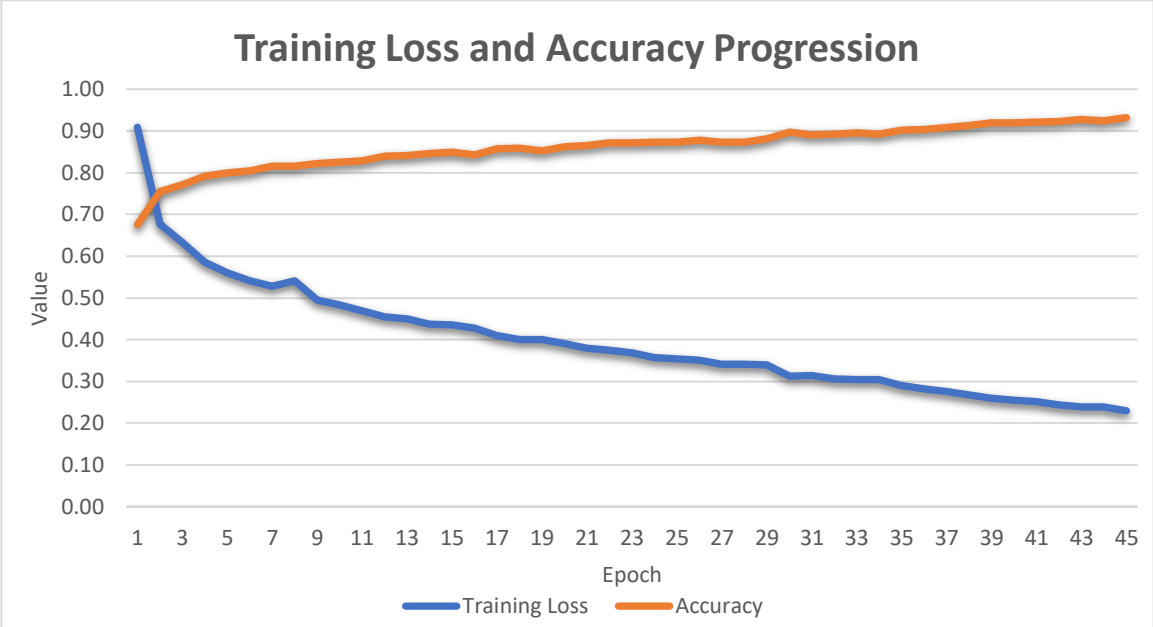

Figure 8: Progression of Training Loss and Accuracy Over Epochs

From the graph, it is apparent that as the number of epochs increases, the training loss decreases, and the training accuracy increases. This is desirable as it indicates that the model is learning and improving its ability to identify DR patterns. The model's improving accuracy is particularly important in this context, as it represents the model's increasing reliability in correctly classifying the severity of retinopathy from retinal images. As training progresses, the loss flattens out, suggesting that the model is starting to converge and that each additional epoch results in smaller improvements.

4.2. Test Sample Prediction

Before analysing how the model performs on the test set, it is important to understand the model output, which is a probability per class that the retina image falls onto:

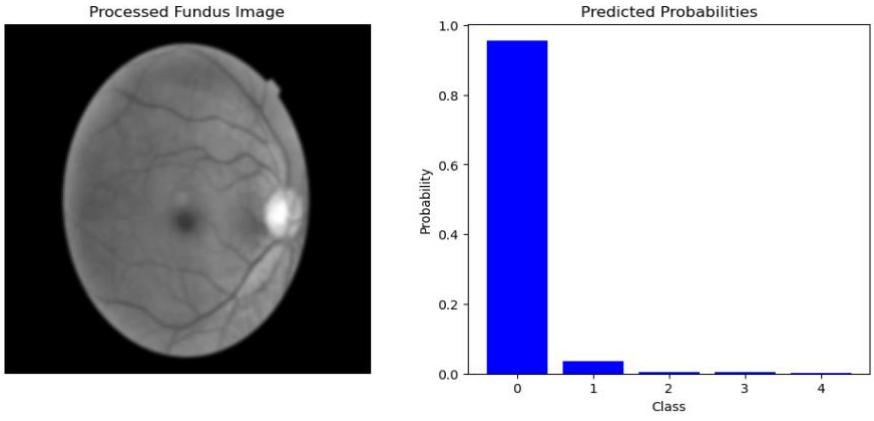

Figure 9: Processed Test Image and Predicted DR Class Probabilities

In this example, the model classifies the test sample as class 0 due to the high predicted probability of class 0, indicating no presence of DR and a healthy retina, which is correct. The next section will highlight how the model performs for all the images in the test set.

4.3. ROC Curve Analysis

The trained model was used to make predictions on the test data and some metrics were calculated using the sci-kit learn library:

Test Data Metric	DR Severity
Test Loss	0.52
Test Accuracy	0.81
Test Area Under Curve	0.97

Table 2: Results on Test Dataset

The training loss graph previously examined indicated a training loss of approximately 0.22, whereas the loss on the test data stands at 0.52. This increase suggests that while the model has learned to predict with a considerable degree of accuracy the training set, it exhibits less precision when applied to unseen data, which is expected as models tend to perform slightly worse on unseen data. Similarly, the training accuracy was slightly higher than the test accuracy of 81%. This means that the model correctly predicts the DR severity class 81% of the time, which is desirable. In the context of DR, the high test AUC value is particularly promising. It suggests that the model is capable of distinguishing between different severities of DR, which is important for accurate diagnosis that can potentially inform treatment decisions. Plottiing the ROC (Receiver Operating Characteristics) curves for each class presented a new perspective:

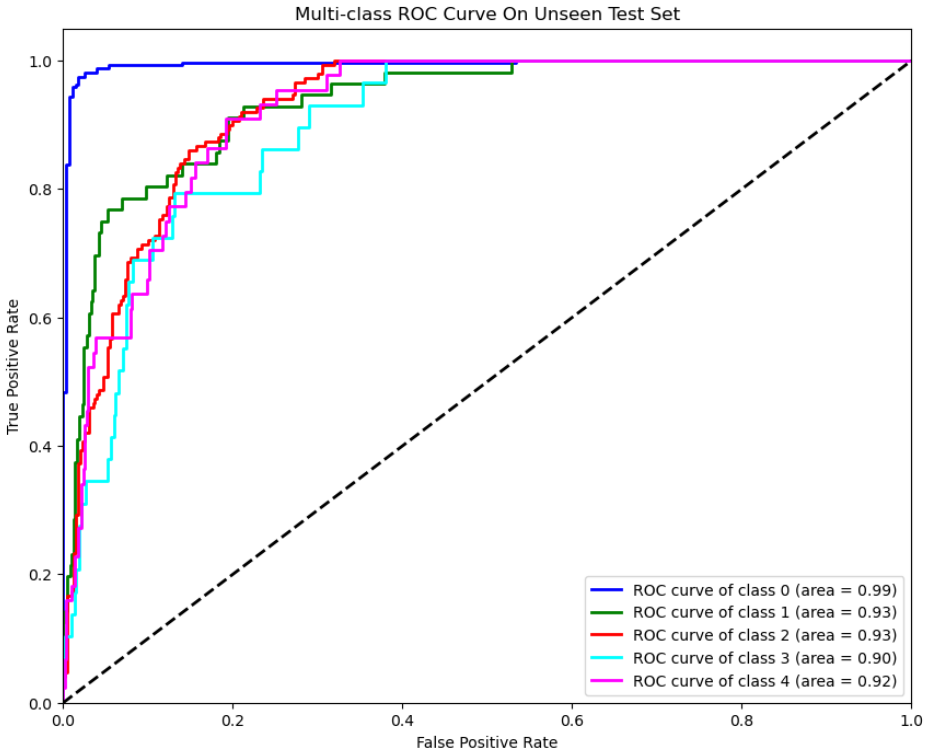

Figure 10: Multi-class ROC Curve on Test Set

Overall, all the ROC curves show a steep rise in the true positive rate, showing that the model correctly identifies DR for most of the true cases for each class. All the AUCs are 0.90 and above, indicating high accuracy and good diagnostic ability for a model that was only trained for 3 hours. However, there is a noticeable disparity between results from class 0, and the remaining classes. The class 0 curve has an area of 0.99 which is almost perfect, so the model has an excellent capacity to distinguish between class 0 from all other classes with high sensitivity and specificity. This discrepancy arises from the distribution of the training data; the model's superior performance on class 0 can be attributed to the fact that nearly half of the training samples belonged to class 0. Therefore, one improvement would be to use a more diverse dataset where all classes are adequately represented. Another improvement would be to fine-tune the DenseNet model by unfreezing the frozen layers and training further. This would improve model robustness and would result in greater accuracy and lower loss across all classes.

Chapter 5: Conclusion

5.1. Conclusion

This project explored the potential of machine learning in the early detection of DR. By adopting a transfer learning approach and harnessing the capabilities of the DenseNet architecture, the study aimed to create an efficient and scalable model for the accurate screening of DR for patients in rural India.

The project commenced with an examination of the APTOS dataset, acknowledging the challenges posed by imbalanced class distributions and the data variability in the retina images. Systematic pre-processing techniques were then employed to ensure appropriate standardisation of the training and test data. The model's training progression demonstrated promising results, with training loss decreasing and accuracy increasing over time. These results, coupled with an 81% test accuracy and an excellent area under the ROC curve of 0.97, substantiate the model's proficiency in DR detection. The ROC curve analysis shows that the model exhibits exceptional performance in distinguishing 'No DR' cases from others, which can be attributed to the substantial representation of this class in the training set. This denotes a significant sensitivity and specificity in the model's diagnostic capability, although it also emphasizes the need for a more balanced dataset to achieve uniform accuracy across all classes.

Future work should focus on augmenting the dataset diversity and further tuning of the DenseNet model to enhance robustness with the aim of helping patients in underserved communities who are at risk of DR.

5.2. Clinical Challenges

Deploying ML models like DenseNet for DR detection in clinical settings can pose many challenges including:

1. **Model Interpretability**: Clinicians need to understand the basis of the model's predictions. DenseNet can act like a black box, making it hard for clinicians to understand and trust the generated results.
2. **Generalisation Across Populations:** Models trained on specific demographics, like working adults in India, might not perform well on other populations due to genetic and lifestyle differences.
3. **Integration with Clinical Systems:** Clinics have established systems and integrating new ML solutions could be difficult due to compatibility issues with imaging equipment and electronic health records.
4. **Data Privacy and Security:** Retinal images are sensitive patient data that needs to be kept secure when training and deploying ML models, which is a big challenge.
5. **Regulatory Approval:** ML models that are used for diagnostic purposes are likely to undergo rigorous scrutiny by regulatory bodies, which could slow down deployment.

Bibliography

[1] L. M. T. M. M. H. Khalifa NEM, “Deep Transfer Learning Models for Medical Diabetic Retinopathy Detection,” *National Library of Medicine: Acta Inform Med,* pp. 327-332, 2019.

[2] W. L. Yun, “Identification of different stages of diabetic retinopathy using retinal optical images,” *Elsevier: ScienceDirect,* vol. 178, pp. 106-121, 2008.

[3] Kaggle, “APTOS 2019 Blindness Detection,” Kaggle, 28 June 2019. [Online]. Available: https://www.kaggle.com/competitions/aptos2019-blindness-detection/overview.

[4] T. R. M. O. S. S. a. A. B. S. A. W. Setiawan, “Color Retinal Image Enhancement using CLAHE,” *International Conference on ICT for Smart Society,* pp. 1-3, 2013.

[5] L. K. Ramasamy, “Detection of diabetic retinopathy using a fusion of textural and ridgelet features of retinal images and sequential minimal optimization classifier,” *PeerJ Computer Science,* no. 10, p. 3, 2021.

[6] P. S. V. P. S. H. a. A. G. R. Nandakumar, “Detection of Diabetic Retinopathy from Retinal Images Using DenseNet Models,” *Tech Science Press,* vol. 45, no. 1, pp. 279-291, 2023.