APS360 PROJECT: DIABETIC RETINOPATHY

Hannah Ye
Student# 1009924548
hannahh.ye@mail.utoronto.ca

Gordon Jeon
Student# 1008990451
gordon.jeon@mail.utoronto.ca

Annabel Liu
Student# 1010274711
jianuo.liu@mail.utoronto.ca

Alex Zhang
Student# 1009825405
alexjm.zhang@mail.utoronto.ca

ABSTRACT

An augmented and balanced dataset collection containing the images of retinal scans of different diabetic retinopathy (DR) stages are subjected to train a series of Convolutional Neural Networks including VGG16, AlexNet, EfficientNet, and SqueezeNet. Multilayer Perceptrons (MLPs) were used to classify the severity of DR. VGG16 and a three layer MLP performed the best with an testing accuracy of 53.2%. In addition to presenting data processing and the implementation of the baseline and primary model, this report is expected to present the objective of the project, quantitative and qualitative results, ethical considerations, and draw conclusions to this project. —-Total Pages: 9

1 INTRODUCTION

Our vision is an essence of human vitality. For people with diabetes, diabetic retinopathy (DR) is an eye condition that causes vision loss and is the major cause of blindness (NIH, 2023). DR is composed of the following five stages, ranked in terms of severity:

0. No DR
1. Mild
2. Moderate
3. Severe
4. Proliferative

As prevalence has been rising rapidly in low and middle-income countries (WHO, 2022), the incidence of DR is also rising, hence the importance of early detection and diagnosis. DR can be controlled and treated if it is detected and diagnosed in its early stages (1-2). Given that our eyes play a crucial role in our everyday lives, developing a model that can help predict the stages of DR and improve patient outcomes is the greatest motivation for this project. With deep learning models, we can handle large datasets of images, learn from pre-existing patterns and build relationships between diabetic retinal photos to help diagnose patients. For example, our model will be able to distinguish between a healthy and diseased retina and classify the stages of DR. Unlike manual examination by optometrists that may be time-consuming, subjective, and prone to error, deep learning models can continuously learn, optimize the indicators, and possibly discover hidden patterns, thus maximizing prediction accuracy and allowing for improvement overtime. Therefore, our team's goal is to develop a robust and accurate model that can aid optometrists by speeding up the diagnosis process and removing subjectivity and error, leading to timely treatment.

2 ILLUSTRATION

An illustration of our process and a sample image of a retinal scan is presented in Figure 1.

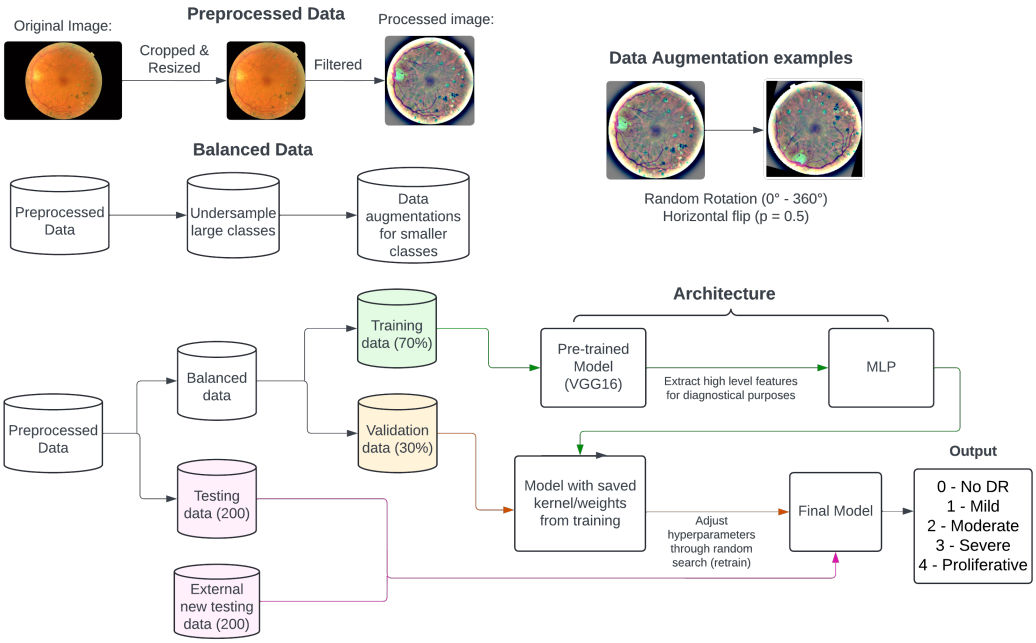


Figure 1: Illustration of the high-level overview of the project

3 Background & Related Work

Numerous research studies have been conducted in hopes to optimize detection and classification models for the stages of DR. In such studies, different methods including algorithms for classification, pre-processing, feature extraction, etc., were tested in order to achieve a higher accuracy in diagnosing DR.

Sebastian et al. proposed the usage of CNN architectures and several pre-processing techniques such as grayscale, resizing, and conversion to enhance feature extraction and eliminate unnecessary noise from the data. The model reached a testing accuracy of 97% in diagnosing and classifying the stages of DR (Al-Maadeed et al., 2023).

AbdelMaksoud et al. integrated two CNN models, namely EyeNet and DenseNet based on transfer learning. This resulted in an optimized hybrid E-DensNet model which could accurately distinguish between a healthy and unhealthy retina, and identify the stages of DR with an average accuracy of 91.2% (AbdelMaksoud et al., 2022).

In addition to using CNNs and transfer learning for DR classification, Wan et al. adopted pre-processing techniques such as normalization schemes and data augmentation to compensate for the lack of input images in their dataset. Following the parameter-tuning of the proposed CNN, the model yielded a classification accuracy of 95.68% (Liang et al., 2018).

Deepa et al. utilized the pre-existing Xception model and a novel clustering algorithm, hierarchical clustering by Siamese network, to classify the different stages of DR. The model tested several hyperparameters, such as the optimizer, batch size, learning rate, etc., and obtained a 96% accuracy (Deepa et al., 2022).

The DR detection model proposed by Khalifa et al. utilized a deep transfer learning CNN to reduce training time by transferring learning weights from pre-existing models such as AlexNet and SqueezeNet. Data augmentation techniques were also adopted to prevent overfitting by increasing the amount of images in the dataset, leading to a significant improvement in the model's testing accuracy. The highest accuracy of 97.9% was achieved using an eight layer AlexNet model (Khalifa et al., 2019).

Meenakshi et al. focused on a two step method in which a boosting based ensemble learning method was used first to predict the presence of DR, then a CNN model was used to categorize the stages of DR. The resulting model gave a 96% accuracy in detecting DR and a 98% accuracy in the categorization (Meenakshi & Thailambal, 2022).

4 Data Processing

Our datasets were taken from Kaggle, a data science competition platform under Google LLC that provides free datasets to the public. The dataset we chose was Diabetic Retinography Arranged (Neo, 2021), by Aman Neo which he made the data from DR (resized) (ILoveScience, 2019) easily readable through pytorch and arranged the data based on the class. The root source of these data are from a DR Detection (Cukierski et al., 2015a) Competition, sponsored by the California Healthcare Foundation. The retinal images are provided by EyePACS, a free platform for retinopathy screening.

Since the images varied between sizes, lighting, camera resolution, and colour, the team decided to preprocess the dataset using the steps listed below. Ben Graham (Graham, 2015), the winner of the DR Detection Competition, utilized techniques 3-5 in image preprocessing. We found that the effects of the techniques were ideal through our own experimentation, therefore we implemented the idea.

1. Resize the data to 224 by 224 to ensure consistency in input dimensions
2. Center Crop
3. Calculate the local average colour using Gaussian blur (helps with noise reduction)
4. Enhance the contrast by mapping the local average colour to 50% grey (makes blood vessels and lesions more prominent)
5. Subtract the mapped local average colour from the image (Graham, 2015) (normalizes the illumination and colour variability)

An example of the original and preprocessed image is in Figure 2.

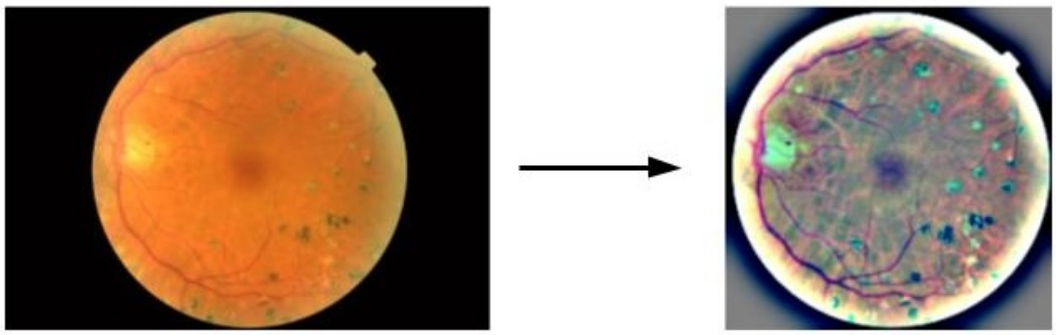


Figure 2: Original (left) and preprocessed (right) image comparison

This data was separated into the 5 stages mentioned in section 1.0. This dataset is severely unbalanced with over 70% of the images in stage 0 (No DR), as shown in table 1. To address this problem, we set a limit for images in each class and undersampled the data from classes that had more images than the limit (5200). Following, the team performed data augmentation on stages 1, 3, and 4 which increased the amount of data in those classes and gave us a balanced dataset. The augmentation techniques performed were rotation of up to 360 degrees and horizontal flipping. This was inspired from Goceri (Goceri, 2023), who researched multiple techniques such as scaling, rotation, flipping, etc., on different types of medical images including retinal scans. We ignored scaling because we recognized that the images in the dataset all had varying magnifications, therefore scaling (enlarging/reducing the size) was not necessary. We were aware that if class 0 and 2 were not augmented, bias may have been introduced, thus they were also augmented, leading to 6000 images in each class. That amount was specifically chosen because a total of 30000 images was the maximum number of

images our computers could run without crashing. Statistics for the dataset balancing can be found in table 1.

Table 1: Dataset statistics during preprocessing and augmentations.

	Number of images				
Class	Original	Undersampling & preprocessing	Preprocessed images moved to testing	Rest of preprocessed images (to be augmented)	Augmentation
0	25810	5200	200	5000	6000
1	2443	2443	200	2243	6000
2	5292	5200	200	5000	6000
3	873	873	200	673	6000
4	708	708	200	508	6000

Lastly, the balanced dataset was then capped at 2000 due to hardware limitations, then separated into training and validation datasets, with 0.7 and 0.3 ratios respectively. In addition, note that 200 samples from each class were separated out of the original dataset to form our internal testing dataset. A percentage wasn't used as the team wanted a balanced test set to provide a fair assessment of performance on all classes, especially the ones underrepresented. The team also obtained an entirely new dataset to test our model, named external testing dataset in table 2. Further details will be described in section 9 of this report.

Table 2: Dataset statistics for validation, training and testing datasets.

	Number of images			
Class	Training	Validation	Non augmented internal testing dataset	External testing dataset
0	1400	600	200	200
1	1400	600	200	200
2	1400	600	200	200
3	1400	600	200	200
4	1200	600	200	200

5 ARCHITECTURE

Our team utilized transfer learning, ie. using pre-trained networks to extract high level features from the images, then run those features through a classifier to make a diagnosis on the severity of the disease. This approach leverages the pre-existing knowledge of the pre-trained models, as they were trained over large datasets. In addition, it also saves computational power, which is unfortunately lacking for our team. By using transfer learning, the team can focus more on optimizing the classification part of our model.

The pre-trained network selected by our team is VGG16. The "features" section of the architecture was utilized to extract high level features from the filtered samples. It consists of a series of thirteen convolutional layers separated in five blocks, which increases the depth (number of convolutional filters) for every odd layer starting with 64 and doubling after each block until 512. Each convolutional layer is followed by a ReLU activation function and at the end of every block there is a maxpool layer for reducing resolution. This type of architecture (Figure 3) allows it to capture 512 high level features from a single image, which would make it more efficient for classification tasks.

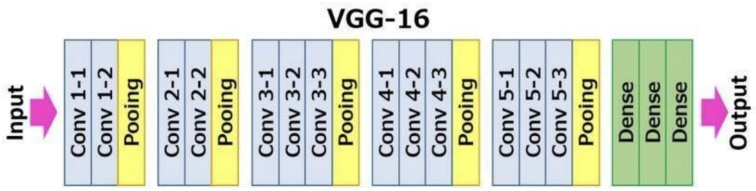


Figure 3: VGG-16 architecture (Yawar, 2024)

The high-level features extracted by the above are then passed through a classifier (Figure 4) with three fully-connected layers. The hidden layers have 512 and 128 neurons respectively and the output layer contains 5 neurons, which represents the probability of each of the five classes. Regularization techniques such as layer normalization and dropout (50%) were utilized on the two hidden layers.

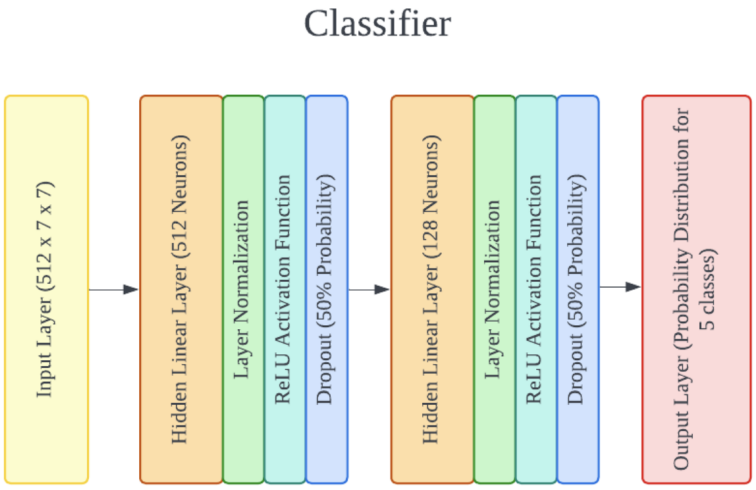


Figure 4: Classifier architecture

6 Baseline Model

For this project, the baseline model chosen is Random Forests, which is a combination of multiple decision trees (Figure 5). Each individual tree gives a yes or no answer and the random forest would combine the predictions from each individual tree to make a final prediction (IBM, 2023).

Samples from the same training dataset for the primary model were passed through the same filters, then the pre-trained VGG16 model, and finally fed into the random forest classifier to mimic the transfer learning utilized by our primary model. Since the effectiveness of transfer learning was already acknowledged from Lab 3 and Tutorial 3b, we made this choice so we can directly compare the effectiveness of the classifier of our primary model. The hyperparameters were tuned using Random Search CV, and the most optimal was a tree depth of 14 with 273 different trees. This model achieved a training accuracy of 99%, a validation accuracy of 45%, and a testing accuracy of 38.3% using the internal testing dataset.

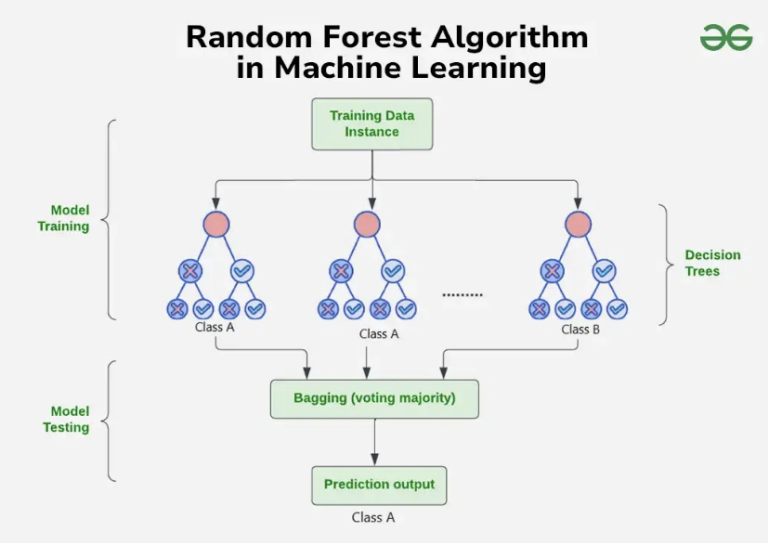
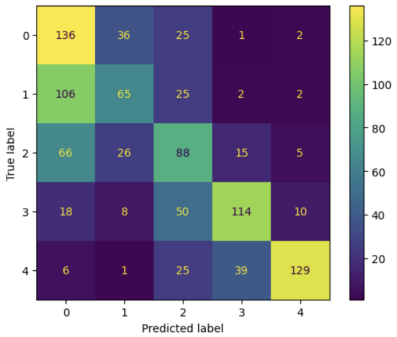


Figure 5: Random forest algorithm (GeeksforGeeks, 2024)

7 Quantitative Results

To evaluate the effectiveness of the final model, we had it run over two sets of data that it hasn't seen before while collecting quantitative measures. The first set is a pre-determined testing dataset consisting of samples from the same source as the training and validation dataset, while the second set is from a completely new source. We took this approach because it reflects the model's performance over real world scenarios, where the model's ability to generalize over unseen data is much more important than memorizing.

The quantitative results that we collected include the accuracy of the model, as well as the confusion matrix, which we will then use to calculate recall, precision, and the F1 score (Figure 6). Recall that the original dataset is significantly imbalanced, and even though downsampling and augmentation techniques were used to create a balanced dataset, there may still be imperfections between classes that could be captured with the measurements mentioned above. In the same logic, accuracy for predicting each class was also calculated by referencing the confusion matrix. The following results contain the model's performance over our internal testing dataset. It consists of 200 samples from each class, which we then preprocessed the same way as the training/validation dataset.



Testing accuracy: 0.532
Precision: 0.5678027117763695
Recall: 0.532
F1 Score: 0.5360535289736754
False Positives per Class (samples):
0: 196 | **1**: 71 | **2**: 125 | **3**: 57 | **4**: 19
False Negatives per Class (samples):
0: 64 | **1**: 135 | **2**: 112 | **3**: 86 | **4**: 71
Accuracy for each class:
0: 0.68 | **1**: 0.325 | **2**: 0.44 | **3**: 0.57 | **4**: 0.645

Figure 6: Confusion matrix and relevant calculations

The high number of false positives for class 0 and 2 shows that the model is overconfident in guessing samples to be from that particular class, while the high number of false negatives for class 1 and 2 shows that the model is failing to identify actual samples from those classes. When you put in perspective that we only have 200 samples per class, it shows that the model lacks understanding of

the above classes mentioned. In the same sense, the model's predictions for class 4 are very precise. The above statistics also shows that the model has a relatively higher precision compared to recall, which means that the model is usually accurate when predicting a positive case. However, the model has a higher tendency to miss positive cases, which is unideal for medical diagnoses.

8 Qualitative Results

To illustrate the performance of our model, a preprocessed image was passed through VGG16, which was the pre-trained model chosen, then visualized using the feature maps outputted after the convolutional layers. Sample images of this process are shown in Figure 7.

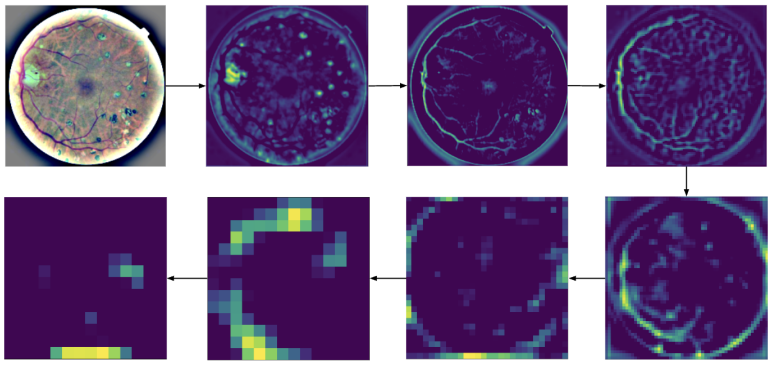


Figure 7: Visualizations of the preprocessed image after the convolutional layers of VGG16

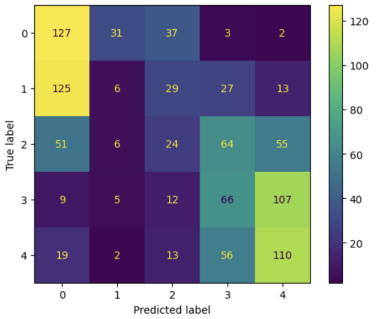
Through the sequence of images in the figure above, we observe that the model initially captures finer details, and as we move deeper into the model, the features begin to become more abstract. Extracting these abstract features helps the model better generalize and classify other retinal images.

Relating to the quantitative results in the section above, we can see that our model performs better on classifying classes 3 and 4 (severe and proliferate DR). This may be because classes 3 and 4 have more distinguishable features, such as lesions and macular holes, in comparison to classes 1 and 2. Since those features are more apparent in classes 3 and 4, the model should be able to extract them with greater effectiveness. In addition, those classes initially had less than 1000 images, which were then augmented to 6000. The extensive augmentation likely also helped the model extract more relevant features, causing better generalization of this data and improved performance.

Furthermore, the results indicate that the model primarily confuses adjacent classes. This may be due to the extreme similarities between data in neighboring classes. Such similarities would lead the model to capture more analogous features, making it difficult to distinguish between the classes.

9 Evaluate Model on New Data

To ensure that the model's performance on new data is accurately represented, the team chose to acquire additional testing data from a completely different source. Evaluating the model on an external dataset facilitates an assessment of its ability to generalize and classify, closely simulating real-world applicative conditions. For this purpose, we selected the dataset from the 2019 Asia Pacific Tele-Ophthalmology Society (APTOS) Blindness Detection competition (Society, 2019). Due to the unavailability of the competition's testing dataset solutions, the team repurposed the provided training dataset for model testing. To maintain consistency with our internal test set, we selected 200 samples from each class, replicating the approach used for our internal evaluation and training. On this new dataset, our model achieved a testing accuracy of 33%, with the results presented in Figure 8.



Testing accuracy: 0.33
Precision: 0.280217115491484
Recall: 0.33
F1 Score: 0.28967988178133786
False Positives per Class (samples):
0: 204 | **1**: 44 | **2**: 91 | **3**: 150 | **4**: 177
False Negatives per Class (samples):
0: 73| **1**: 194 | **2**: 176 | **3**: 133 | **4**: 90
Accuracy for each class:
0: 0.635 | **1**: 0.03 | **2**: 0.12 | **3**: 0.33 | **4**: 0.55

Figure 8: Results from the external test dataset

10 Discussion

Based on the final model's results, we concluded that its performance was satisfactory but limited. In comparison to the competition leaderboard, our solution would have ranked within the top 10th percentile. The performance of our highly effective baseline model, which achieved 38.3% accuracy through the implementation of both filters and transfer learning, underscored the significant challenges inherent in this project. Despite multiple rounds of pre-processing and augmentation (refer to Qualitative Results), we faced the challenge of developing a model capable of effectively capturing features and generalizing patterns for such a delicate set of data. This ultimately resulted in a classic case of overfitting, where validation accuracy plateaued, and the model ceased to learn or extract useful information. Despite these setbacks in developing a robust model, the team adapted and refined methodologies to optimize the model to the best extent of our knowledge and hardware's capabilities.

As reflected in the training log (Figure 9), our initial approach involved heuristic tuning through a trial-and-error method to assess the impact of each hyperparameter on model performance. While this method offered simplicity, it failed to ensure robust and stable outcomes. Consequently, we explored alternative tuning strategies, including adjustments to optimizers, pooling methods, and regularization techniques, as well as the implementation of cyclic and decaying learning rates. Conducting such rigorous experimentation and testing led us to the hyperparameters and properties of our final model.

Model	Aug	Percent	Batch Size	Epochs	Learning Rate	Cap (amt. in each class)	Training Acc.	Val. Acc.	Training %	Validation %
VGG16	Y	50	128	20	0.01	N	0.631	0.599	0.7	0.15
VGG16	Y	50	128	10	0.001	N	0.583	0.552	0.7	0.15
VGG16	N	100	64	20	0.001	709	0.589	0.387	0.7	0.15
VGG16	N	100	64	50	0.002	709	0.955	0.383	0.7	0.15
VGG16	N	100	64	50	0.003	709	0.974	0.358	0.7	0.15
MOBILENET	N	100	64	50	0.001	709	0.844	0.421	0.7	0.15
MOBILENET	N	100	64	50	0.002	709	0.93	0.415	0.7	0.15
MOBILENET	N	100	64	50	0.003	709	0.999	0.43	0.7	0.15
RESNET50	N	100	64	50	0.001	709	0.468	0.368	0.7	0.15
RESNET50	N	100	64	50	0.002	709	0.539	0.413	0.7	0.15
RESNET50	N	100	64	50	0.003	709	0.6	0.404	0.7	0.15
VGG16	Y	100	64	50	0.001	5200	0.895	0.517	0.7	0.15
VGG16	Y	100	64	50	0.002	5200	0.937	0.511	0.7	0.15
VGG16	Y	100	64	50	0.003	5200	0.962	0.521	0.7	0.15
VGG16	Y	100	64	50	0.005	5200	1	0.562	0.7	0.2
VGG16	Y	100	64	50	0.005	5200	1	0.572	0.6	0.3
VGG16	Y	100	64	50	0.01	5200	1	0.572	0.6	0.3

Figure 9: Sample entries from the training log

From a medical standpoint, our model remains insufficiently effective. With half of the predictions resulting in misdiagnoses, its application in clinical settings would be too risky, as any misdiagnosis could have severe consequences for a patient's vision and then some. Additionally, our model underpredicts the severity of DR with the external dataset (ie. predicts false negatives). This would severely impact the patients, as the model identified 125 class 1 samples as class 0's (no DR). Informing a patient that they don't have DR imposes more risks than overpredicting a patient's severity.

A comparison of the model's predictions on the internal and external testing datasets revealed markedly different behaviors. When working with samples from the internal dataset, the model exhibited a higher tendency to classify them under class 0. Contrastingly, when working with samples from the external dataset, the model showed an increased likelihood of predicting classes 3 and 4. These observations suggest that the model lacks sufficient generalization. This discrepancy may be attributed to the nature of the internal dataset, potentially limiting the model's ability to generalize. Additionally, the subtle features between stages of DR, such as tiny blood vessels in the eyes, pose a significant challenge for accurate detection and in turn, generalization. Even minor variations in the input data can mislead the model. If the project timeline were not a constraint, further improvements could be pursued, such as maintaining the original resolution of the images and employing more extensive data augmentation techniques. These strategies could help the model learn the intricate details necessary for accurately diagnosing diabetic retinopathy.

Based on our understanding of classification problems and the additional research we conducted, we have exhausted all feasible approaches for this classification task. Although the results did not meet our expectations, they represent our best effort given our current knowledge and hardware constraints. As discussed in the baseline model section, our classifier still demonstrates a notable improvement, exceeding the baseline model's performance by over 10%. Similarly, when considering the performance in the competition from which we sourced the dataset, the majority of participants—ranging from professionals, academics, to hobbyists in applicative AI—achieved results comparable to ours, with most participants (400/600) reaching only 10% (Cukierski et al., 2015b). Therefore, we suggest that the final accuracy we achieved should be considered within the context of the specific challenges posed by this problem. While our results may not meet the rigorous standards of efficacy seen in the medical field, they are more appropriately compared to the outcomes of similar competitions. The medical field demands an exceptionally high level of precision, which differs significantly from the benchmarks typically used in other domains.

11 Ethical Considerations

The dataset does not contain information regarding where the data was collected. This means that the model could have potential representation bias due to limited geographical diversity to cover the general population. In other words, the model may perform better or worse based on the individual's eye colour (Blake et al., 2003).

The model is only capable of outputting the stage of DR the patient is in. However, it is not possible for the model to explain the reasons behind making the diagnosis to the patient. In addition, there are rare occurrences that the model can make mistakes. These are all reasons why the model cannot replace the job of a professional diagnostician. The model should only be used as a reference. A trained professional who thoroughly understands the disease should be present at all times when using the model.

This model should be able to fit all ethnicities and gender. Apart from iris colour, there are minor variances of internal structure of the eyes that are relevant for DR diagnosis (Blake et al., 2003). These factors should not cause issues for the model as the model is primarily focusing on spotting areas of yellow exudates or breakage of blood vessels (Star Retina, 2022). In addition, unlike usual medical data, this dataset contains even ratios of labels. This means that we do not need to worry about an inaccurate measure of performance of the model when calculating its accuracy, and that we do not have to resort to calculating the recall of the model.

References

Eman AbdelMaksoud, Sherif Barakat, and Mohammed Elmogy. A computer-aided diagnosis system for detecting various diabetic retinopathy grades based on a hybrid deep learning technique. *Med Biol Eng Comput*, 60:2015–2038, 2022. doi: 10.1007/s11517-022-02564-6.

Somaya Al-Maadeed, Noor Almaadeed, Omar Elharrouss, and Anila Sebastian. A survey on deep-learning-based diabetic retinopathy classification. *Diagnostics*, 13(3), 2023. ISSN 2075-4418. doi: 10.3390/diagnostics13030345. URL https://www.mdpi.com/2075-4418/13/3/345.

C. Richard Blake, Deepak P. Edward, and Wico W. Lai. Racial and ethnic differences in ocular anatomy. *International Ophthalmology Clinics*, 2003. doi: 10.1097/00004397-200304340-00004.

Will Cukierski, Emma Dugas, Jared, and Jorge. Diabetic retinopathy detection. https://kaggle.com/competitions/diabetic-retinopathy-detection, 2015a.

Will Cukierski, Emma Dugas, Jared, and Jorge. Diabetic retinopathy detection leaderboard. https://www.kaggle.com/competitions/diabetic-retinopathy-detection/leaderboard?tab=public, 2015b.

V. Deepa, C. Sathish Kumar, and Thomas Cherian. Automated grading of diabetic retinopathy using cnn with hierarchical clustering of image patches by siamese network. *Phys Eng Sci Med*, 45: 623–635, 2022. doi: 10.1007/s13246-022-01129-z.

GeeksforGeeks. Random forest algorithm in machine learning. https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/, 2024.

Evgin Goceri. Medical image data augmentation: techniques, comparisons and interpretations. *Artif Intell Rev*, 56:12561–12605, 2023. doi: 10.1007/s10462-023-10453-z.

Ben Graham. Kaggle diabetic retinopathy dection competition report. https://storage.googleapis.com/kaggle-forum-message-attachments/88655/2795/competitionreport.pdf, 2015.

IBM. What is random forest? https://kaggle.com/competitions/diabetic-retinopathy-detection, 2023.

ILoveScience. Diabetic retinopathy resized. https://www.kaggle.com/datasets/tanlikesmath/diabetic-retinopathy-resized, 2019.

Nour Eldeen M. Khalifa, Mohamed Loey, Hamed Nasr Eldin Taha Mohamed, and Mohamed Hamed N. Taha. Deep transfer learning models for medical diabetic retinopathy detection. *Acta Informatica Medica*, 27(5):327–332, 2019. doi: 10.5455/aim.2019.27.327-332.

Yan Liang, Shaohua Wan, and Yin Zhang. Deep convolutional neural networks for diabetic retinopathy detection by image classification. *Computers & Electrical Engineering*, 72:274–282, 2018. ISSN 0045-7906. doi: 10.1016/j.compeleceng.2018.07.042. URL https://www.sciencedirect.com/science/article/pii/S0045790618302556.

G. Meenakshi and G. Thailambal. Categorisation and prognostication of diabetic retinopathy using ensemble learning and cnn. In *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 1145–1152, 2022. doi: 10.1109/ICOEI53556.2022.9777156.

Aman Neo. Diabetic retinopathy arranged. https://www.kaggle.com/datasets/amanneo/diabetic-retinopathy-resized-arranged, 2021.

NIH. Diabetic retinopathy — national eye institute. https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/diabetic-retinopathy, 2023. [Accessed 06-06-2024].

Asia Pacific Tele-Ophthalmology Society. Aptos 2019 blindness detection. https://www.kaggle.com/c/aptos2019-blindness-detection/data, 2019.

Star Retina. What is diabetic retinopathy? https://starretina.com/diabetic-retinopathy/, 2022.

WHO. Diabetes — who. https://www.who.int/news-room/fact-sheets/detail/diabetes#:~:text=Over%20time%2C%20diabetes%20can%20damage,blood%20vessels%20in%20the%20eyes, 2022. [Accessed 06-06-2024].

Md Yawar. Vgg-16 - cnn model. https://www.naukri.com/code360/library/vgg-16---cnn-model, 2024.