DEEP LEARNING MODEL FOR DIAGNOSING DIABETIC RETINOPATHY

Abdullah Amin
Student# 1008803118
abdul.amin@mail.utoronto.ca

Abubaker Jassim
Student# 1008749749
a.jassim@mail.utoronto.ca

Sami Najim
Student# 1008717578
sami.najim@mail.utoronto.ca

Iftier Rahman
Student# 1009121745
iftier.rahman@mail.utoronto.ca

1 INTRODUCTION

Approximately 537 million people globally suffer from diabetes in which 27% percent have to endure diabetic retinopathy (Federation) (Zegeye et al. (2023)). The project undertaken by our team involves the thorough examination of a plethora of retinal scans, and analyzing them in order to give a conclusive deduction on the health of an individual's eyes, as seen in Figure 1. The project was greatly motivated by the team's desire to tackle the challenge of detecting signs of Diabetic Retinopathy (DR) in its different stages as early as possible in order to expedite the treatment process for these patients. Current methods in retinopathy diagnosis are usually time consuming and expensive as it usually requires the manual inspection from ophthalmologists.

Figure 1: Retinal fundus image scans that were analyzed and classified (Karthik (2019))

The goal in this project is to develop an automated system that will be responsible for receiving retinal scan images of patients as input and outputs whether the eye is healthy or not, and if not at what stage of DR is the eye classified as. By utilizing deep learning techniques and networks, our team hopes to provide better facilitation for early prevention and treatment and increase the chances of protecting patient eyesight from blindness when suffering through DR.

The project can prove to be essential and provide important impacts to the overall improvement in the healthcare sector of society. First and foremost, the project directly deals with an issue of blindness that occurs as the result of a particular disease. Thus, our efforts will provide a significant reduction to the risk of blindness from diabetic patients further improving the healthcare quality for patients. Moreover, the success of this project can also prove to increase the accessibility of diagnosis of DR to numerous health care institutions especially in remote locations where ophthalmologists are a rarity. Furthermore, the project could not only prove to be efficient in reducing workload upon specialists, but also prove to be economically friendly upon patients themselves who don't have the means of acquiring retinal screening.

Deep learning techniques have proven to be very effective in image classification and analysis. A well trained and learned neural network can quickly access high level features of retinal scans and classify them from raw pixel data. The subtle differences between retinal scans that require extensive observations from ophthalmologists to be identified, is well suited for deep learning techniques that are proposed to accomplish such tasks. Just by passing retinal scans as input to the network, the network is easily able to analyze different features and output to the user the class representing the stage of DR that the patient possesses.

2 Background and Related Work

The following below are five prior works that were taken into consideration for the development of this project. Most papers explored are a variation of some of the works done below. Some utilize a grading based approach of the 5 stages of DR (Hsu (2024)):

1. No Diabetic Retinopathy
2. Mild Nonproliferative Diabetic Retinopathy (Mild NPDR)
3. Moderate Nonproliferative Diabetic Retinopathy (Moderate NPDR)
4. Severe Nonproliferative Diabetic Retinopathy (Severe NPDR)
5. Proliferative Diabetic Retinopathy (PDR)

The others are only concerned with detection based classification of DR vs. no DR.

2.1 Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

Researchers at Google utilized transfer-learning using the Inception-v3 architecture proposed by Christian Szegedy et al. for classification between referable and non-referable Diabetic Retinopathy. This neural network was then trained on 128,175 retinal fundus images obtained from EyePACS and 3 local eye hospitals in India using 80% of the development data for training and 20% for tuning hyperparameters (Szegedy et al. (2015)). They used ReLU for the activation function, Stochastic Gradient Descent as the optimizer, cross-entropy loss function, and batch normalization. The model was tested on the 9,963 image EyePACS-1 and the 1,748 image Messidor-1 datasets (Gulshan et al. (2016)).

The metric used for testing was Area Under the Curve (AUC), which is used to evaluate the performance of a classification model. It is scored from 0 to 1, models closer to 1 have a near perfect identification model (Google).

Key results from Google's research demonstrate high performance achieving a 0.991 AUC score and 0.990 AUC score respectively with a high sensitivity and specificity rate on both datasets, proving that deep learning techniques are highly effective in this application (Gulshan et al. (2016)).

2.2 Automated Identification of Diabetic Retinopathy Using Deep Learning

A custom convolutional neural network was developed by Rishab Gargeya and Theodore Leng classifying retinal images as having or not having diabetic retinopathy (DR) and an associated probability percentage of the presence of DR using a softmax activation function. This model was trained on 75,137 images obtained from EyePACS using a 5-fold stratified cross-validation approach. They used ReLU activation function, cross-entropy loss function, residual shortcuts, and batch normalization. In addition, a global average pooling layer was used for classification of results through a gradient-boosting classifier and visualizing a heat map of regions of interest that aided in the model's prediction. Tested on the Messidor-2 and E-Ophtha datasets, key results show their model achieved a 0.94 AUC score and 0.95 AUC score respectively with specificity and sensitivity scores in the low 90s (Gargeya & Leng (2017)).

2.3 Using a Deep Learning Algorithm and Integrated Gradients Explanation to Assist Grading for Diabetic Retinopathy

The Inception-v4 architecture proposed by Christian Szegedy et al was used by researchers at Google for the classification of the 5 stages of Diabetic Retinopathy with an associated probability distribution over the 5 stages of DR using a softmax function (Szegedy et al. (2016)). The model was trained on 1.6 million images from EyePACS and 3 indian eye hospitals with 2000 images used for validation and tuning and 1796 images for testing from EyePACS. Integrated gradients were also used to generate a heatmap visualization of ROIs that aided in the model's prediction. Key results show the model had an overall accuracy of 88.4% but a 57.9% accuracy of mild to worse cases of DR and a 96% accuracy for no DR cases. The paper also showed that assistance from the model helped Opthamologists be more accurate in their diagnosis of worse cases than without (Sayres et al. (2019)).

2.4 Deep convolutional neural networks for diabetic retinopathy detection by image classification

In this paper, Shaohua Wan, Yan Liang, and Yin Zhang used the publicly available 35,126 image Kaggle DR dataset to classify the 5 stages of diabetic retinopathy using the ResNet, VGG, GoogleNet, and AlexNet architectures. Considering the imbalance in the dataset, various data augmentation techniques such as noise reduction, flipping, stretching, and rotations were used to balance the class distributions and increase the data set to 20 times its size. In addition, stochastic gradient descent was used as the optimizer. Key results show that amongst the models tested, VggNet-s performed the best with a 5-class overall accuracy of 95.68% (Wan et al. (2018)).

2.5 Deep Learning for the Detection and Classification of Diabetic Retinopathy with an Improved Activation Function

Usharani Bhimavarapu and Gopi Battineni proposed a new activation function $f(x) = \frac{x}{\cos(x)}$ that has been seen to outperform other commonly used activation functions including ReLU for diabetic retinopathy classification. Using transfer learning, they trained Inception-v3, VGG-19, ResNet-50, AlexNet, GoogleNet, SqueezeNet, and ResNet-152 models using the proposed activation function and a variant of the adam optimizer - Nadam optimizer. These models were trained on the Kaggle, CHASE, DRIVE, and DIARETDB0 datasets for a total of 88,904 images. Results showed that when testing the following models using the proposed activation function against other activation functions, the ResNet-152 model using the proposed function performed the best with a 99.41% accuracy (Bhimavarapu & Battineni (2022)).

3 Data Processing

We used the publicly available 35,126 image Kaggle DR (Emma Dugas (2015)), 3662 image APTOS 2019 datasets (Karthik (2019)) for training and validation while the 1748 image Messidor-2 (Patry (n.d.)) (Brain (2021)) was used for testing. All of the aforementioned datasets contain images for the 5 stages of DR.

However, in all of these datasets, there was a significant imbalance in the distribution of data across the 5 stages with most images under the no DR class. Therefore, the following data augmentation techniques will be performed:

1. Image Resizing: All images are to be resized to a dimension of 256x256 to ensure consistent image size/dimension. We tried using 512x512, however due to limitations with memory and colab time constraints, we could not train at this image size reliably.
2. Data Normalization: All image pixel intensities are to be normalized to a similar distribution, based on ImageNet. This is done as ResNet-152 was trained using ImageNet, so basing these normalization of ImageNet will allow for better compatibility for the encoder (transforms.Normalize(mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225])) (He et al. (2015)).
3. Artificial Inflation of Dataset: Due to significant dataset imbalance between no DR and the stages of DR, stages 1,2,3 and 4 have had their images duplicated while class zero had its images reduced so that all classes have the same number of images for training and validation.

In addition to the techniques above, we also used the following data processing techniques used by the winner of the 2015 Diabetic Retinopathy Classification competition:

- Have all images rescaled to have the same radius of about 300 pixels.
- Get the local average coloured subtracted by about 50% greyscale. This is achieved with a combination between the original image and its gaussian blurred image that helps to enhance details like blood vessels, hemorrhages, and exudes which are regions of interest for prediction (right side of Figure 2).
- Cropped fundus images to 90% size in order to reduce effects of black border in each retinal fundus image.

These augmentations were done by using code provided in the winner's paper which uses the openCV and glob libraries to perform these augmentations. Doing all these augmentation and processing techniques ensured better generalization for the model (Community (2015)).

4 Baseline Model

We made use of the support vector machine (SVM) for our choice of baseline model. The SVM was chosen over other models as our baseline model due to its ability to handle multi-class classification and its use in Diabetic

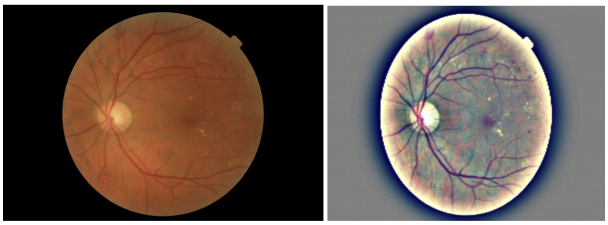

Figure 2: The transformation of original fundus scan (left side) to that used for training models (right side)

Retinopathy classification research such as by M. K. Behera and S. Chakravarty (Behera & Chakravarty (2020)) and by E. V. Carrera, A. González and R. Carrera (Carrera et al. (2017)).

SVMs are generally straightforward to implement and require minimal tuning.They work by finding the most optimal hyperplane, similar to finding the line of best fit, that separates the data points by their respective classes in a higher dimensional space. Figure 3 shows how the hyperplane gives a decision boundary that splits the data for 2 classes with a margin of error. However, unlike CNNs, they do not learn patterns or geometric features as effectively.

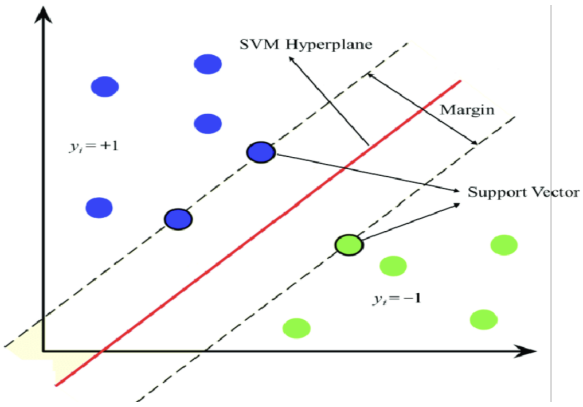

Figure 3: Hyperplane created for decision making by SVM Model (Yang & Prayogo (2022))

We used the SVM classifier built into the sci-kit learn library as our SVM model that we used to train and test.

Before using the SVM, the dataset used, the Kaggle and APTOS DR datasets combined (38,788 images), was processed by grouping images into the 5 respective stages of DR, resizing, normalizing, and performing data augmentations as mentioned above and then converted into image and label numPy arrays. The data is then split into 80% training and 20% validation using train_test_split() from the sci-kit learn library.

A challenge was figuring out what hyperparameters to initialize and use in our SVM. We ended up automating the hyperparameter tuning process using GridSearchCV which allows one to input a set of parameters and cross-validate each combination to find the best combination that optimizes a model's performance. Using GridSearchCV allowed us to experiment with multiple values of C-scores and Gamma values at once using a RBF kernel. There are other kernels available such as polynomial and linear, however, RBF kernel is best for our application for its ability to generalize multi-class non-linear data effectively while requiring fewer parameters to tune.

Another challenge was ensuring our model did not use up all the RAM available. Despite resizing the images to 256x256, the system would use up all the RAM available and crash when performing hyperparameter tuning using GridSearchCV. One way we mitigated this is by using Principal Component Analysis (PCA) which reduces the dimensions of the images in the dataset. In our case, we reduced our dimensions to 100 principle components that capture the most significant information available from the dataset, where each component captures the highest sets of variance available in the data for each feature (i.e. pixels with the biggest variance in data). This as a result helped reduce the computational load when training the SVM.

As seen below in Figure 4, our baseline SVM has a validation accuracy of about 73.4%. Given the dataset is significantly imbalanced and the nature of what is being classified, the recall gives a better idea of how the model is performing as we want to see how well the model is correctly identifying positive cases. The recall of the SVM is also 73.4%. Upon analyzing the precision and recall for each class, it is seen that the model is predicting cases of no DR (stage 0) with high recall of 100%, identifying few false positives and virtually no false negatives.

However, the model is struggling more to identify the other four stages of DR as seen by their lower precision and recall scores. This is especially apparent with stage 4 where the recall is 1% and the precision is 3%.

The possible reasons for these results come down to the following: the capabilities of the SVM, the massive dataset imbalance. Unlike neural networks, SVMs are not the best when it comes to capturing details and patterns as it requires more computations to come up with the optimal hyperplane as a result of more features to process. A lot of research in using SVMs in diabetic retinopathy such as in M. K. Behera and S. Chakravarty and by E. V. Carrera, A. González and R. Carrera, employs feature engineering techniques to extract the exudates in retinal fundus scans - a key indicator for diabetic retinopathy which is not done in our case (Behera & Chakravarty (2020)), (Carrera et al. (2017)). This is one advantage of CNNs as it can do such feature extractions on its own from an image.

The dataset itself is also significantly imbalanced, as seen in Figure 5, where stage 0 makes a large majority of the dataset while class 1, 2, 3, and 4, the stages of DR, make are significantly smaller. This is seen through the model performance as it can learn stage 0 samples with ease but struggles to correctly identify stages of DR. This means that we may have to artificially increase the dataset on stages 1,2,3, and 4 samples through data augmentations to reduce the impact of the imbalance. We may also have to find more datasets to train the model on or change our classification from the 5 stages of DR to distinguish between no DR, non-proliferative DR (stage 1 and 2), and severe & proliferative DR (stage 3 and 4). This may also apply to our primary model as it too can struggle classifying the stages of DR as a result of the class imbalance. We may retrain our SVM if we decide to implement these changes for the primary model.

At the current stage, we will proceed with the results defined in Figure 4 as our benchmark in comparing our neural network primary model.

```
Validation Accuracy: 0.7338030589448359


              precision    recall  f1-score   support

           0       0.74      1.00      0.85      4143
           1       0.43      0.06      0.10       422
           2       0.53      0.12      0.19       944
           3       0.83      0.03      0.06       160
           4       0.50      0.01      0.03       150

    accuracy                           0.73      5819
   macro avg       0.61      0.24      0.25      5819
weighted avg       0.68      0.73      0.65      5819


Precision: 0.683081001973129
Recall: 0.7338030589448359
F1 Score: 0.6477256922656597
```

Figure 4: The SVM model achieved a Validation accuracy of 73.38%, Precision of 68.31%, Recall of 73.38%, and F1 Score of 64.77%

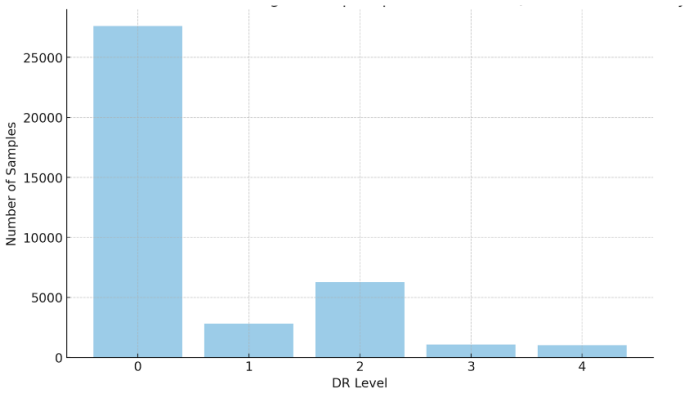

Figure 5: The breakdown of the distribution of images across the 4 classes: Stage 0: 71.19%, Stage 1: 7.25%, Stage 2: 16.22%, Stage 3: 2.75%, and Stage 4: 2.59%

5 Final Model Architecture

After numerous iterative processes of tuning our hyperparameters and neural network architecture, we have reached a conclusive neural network model which accurately and efficiently classifies the DR images. The network is composed of a fixed pretrained ResNet 152 model connected to a fully connected classifier, as seen in Figure 6. The pre-trained ResNet 152 model uses 5 sets of convolutions that are repeated many times as can be seen in Figure 6. These help the model break down the characteristics of the image, which the fully connected classifying layers can then use to further distinguish the fundus scans. The first layer is imputed 2048 output features from the ResNet model (seen in Figure 6). There are 1024 neurons in this first hidden layer, which pass their output to the second hidden layer which is composed of 512 neurons. These neurons connect to the output layer composed of 5 neurons for the 5 different classifications. Each fully connected layer utilizes a ReLu activation function, as well as, 50 percent dropout regularization for countering over and under fitting. This is the most successful model that our team developed, and will be the model that our quantitative results will represent.

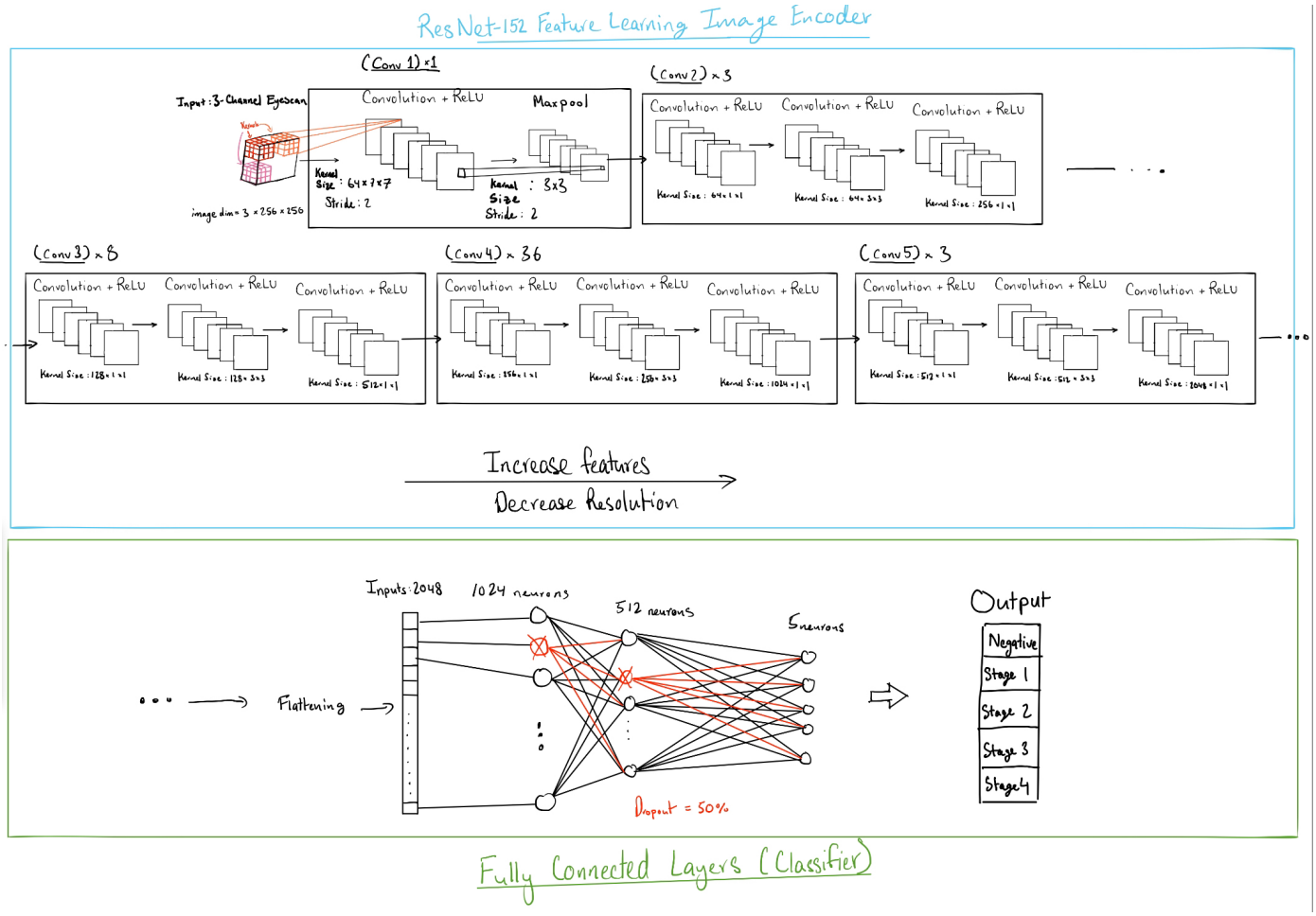

Figure 6: The neural architecture for the diabetic retinopathy classifier neural network developed.

5.1 Model Complexity

Calculations showed that the tested model had the following parameters:

- ResNet-152: 60 million parameters
- Fully Connected Layers: 2,625,541 parameters

The model therefore had a grand total of around 62,625,541 parameters.

6 Results

6.1 Quantitative Results

Based on training and validation, our models accuracy improved from approximately 45% to 82% over 30 epochs, before signs of overfitting as seen in Figure 7. Conversely, our model's loss decreased from approximately 1.27 to 0.45. Both showing positive signs that the model was able to learn and predict data as seen in Figure 8.

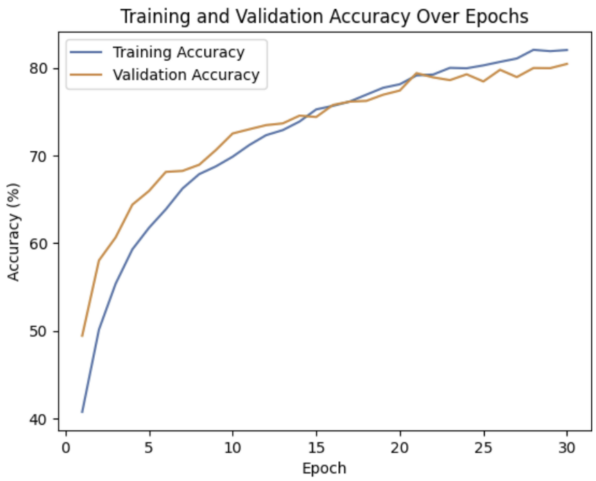

Figure 7: The training and accuracy plotted with respect to epochs of training that have passed.

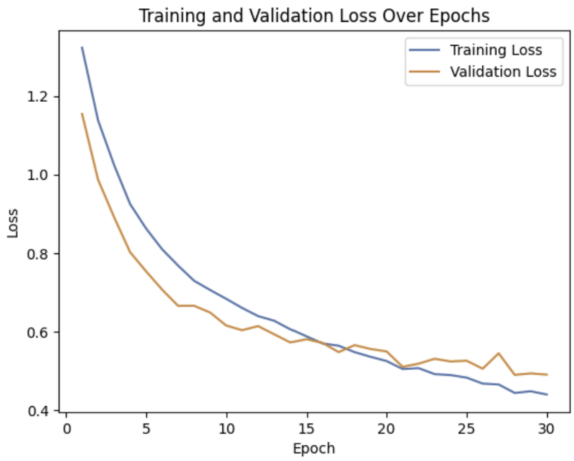

Figure 8: The training and validation loss plotted with respect to epochs of training passed.

Further inspecting the confusion matrix outputted on the test data further shows that the model has an average F1 score of 82.3%, seen in Figure 9. It can be also noted from Figure 9 that the model has the highest F1 score for classes 3 (severe DR) and 4 (proliferative DR), 97.8% and 98% respectively, while classes 0 (no DR) and 3 (moderate DR) had the lowest F1 score, 64.2% and 66.6% respectively.

6.2 Qualitative Results

Our testing dataset composed of images that our model has never seen consists of 5744 images ranging across all 5 classes of DR severity. We notice from the results displayed in Figure 9 that our model's recall is significantly high for severe and proliferative cases with 95% compared to the more mild and moderate cases of 60%. This indicates that the amount of images that our network had incorrectly predicted for severe and proliferative cases are significantly less compared to the rest of the cases. This coincides with the accuracy values illustrated in the quantitative results in which severe and proliferative cases of DR have an accuracy in detection of 99% while the milder versions of DR have an accuracy ranging in the 60 percentile. The success of the model can be attributed

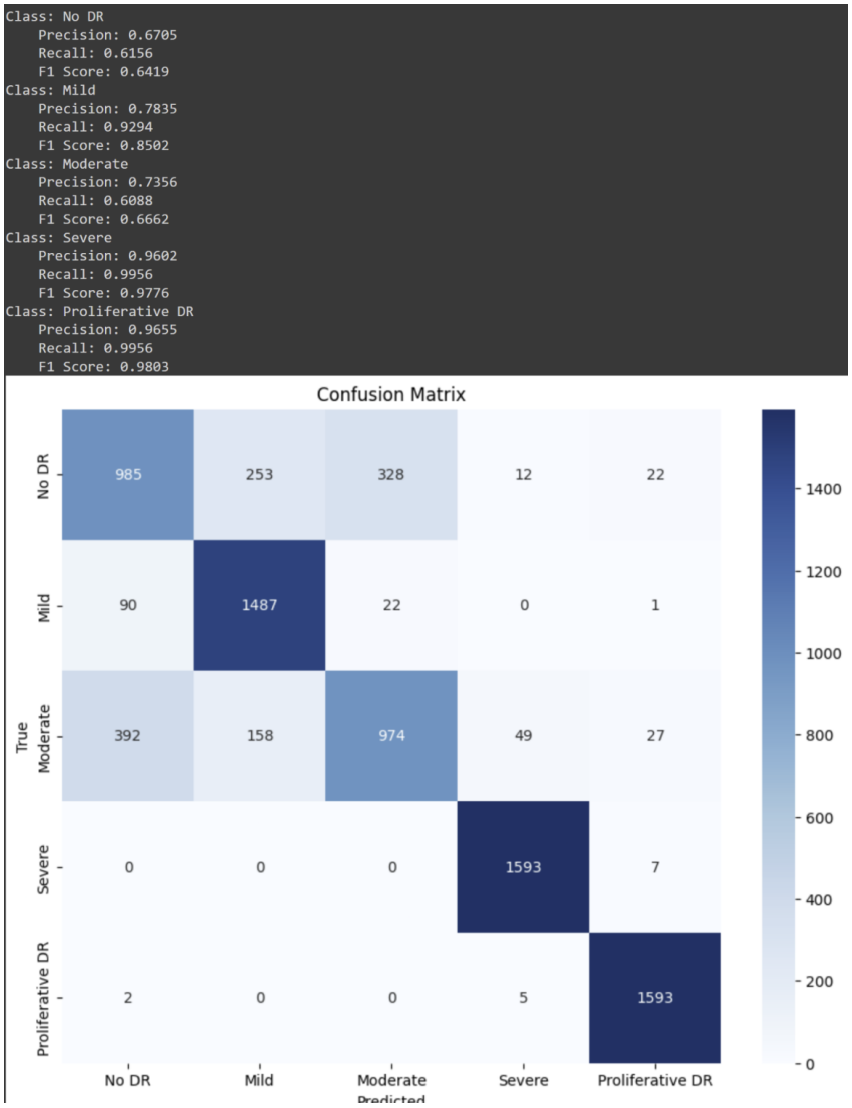

Figure 9: The confusion matrix illustrates the distribution of predictions of classes based on true labels.

to the augmentations applied to the dataset which aided in distinguishing subtle features, coupled with developing a more balanced dataset.

As can be seen in the Figure 9, our model has high recall for classes even when its precision goes down for a class. For example, the mild DR stage (class 1) has a lower precision (78.35%) but a much higher recall (92.9%). High recall is particularly crucial for our application due to the prevalence and severity of DR. DR is a leading cause of blindness worldwide, making it essential that the model is highly sensitive to detecting the disease, especially in its more severe forms. High recall ensures that most, if not all, cases of severe DR are identified, which is vital for intervention and treatment. By accurately detecting these severe cases, our model can potentially save patients' sight by prompting earlier and more effective medical intervention. This capability to reliably detect severe cases highlights the practical significance of our model in clinical settings, where missing a severe case could have dire consequences for the patient.

6.3 Collection of Test Set

A test set was put together with fundus samples that were not seen by the model during training and validation to ensure the final results accurately show the model's performance. 1748 images from the Messidor-2 dataset were merged together to create this test set. This accumulation for the test set was designed to provide an evaluation of the model's generalization capabilities.

Based on the evaluation on this set, we seem to reach similar results as evaluation on our validation set as well. With precision ranging from 75-80% as seen in Figure 9, we can see that the model is proficient at diagnosing

severe and proliferate DR (classes 3 and 4), and not so proficient with classifying no DR. In the case of DR, higher recall values are more important than precision. This is because DR can be so common, sometimes can be overlooked, so it's more important there are more true positives even if it raises the number of false positives. One other thing to mention, is that DR is really diagnosed with visual inspection of doctors, so there are variances in what doctors really categorize between its 4 stages. It is highly likely doctors themselves may fail to recognize signs of DR and label a fundus scan to have no DR. The fact that our model seems to predict so many no DR cases as mild or moderate DR is proof of that.

7 Discussion

After thorough analysis of the quantitative and qualitative outcomes from our neural network model on the testing dataset, it was apparent to us that our model overall had an excellent accuracy rate of 82.3% in correctly classifying images spanning all 5 classes. Our team has established an arbitrary accuracy rate threshold of 80% to determine if our model is successful, we find that our model exceeds these parameters, which our model successfully succeeded. This success is attributed to our team's efforts to balance our data samples for all classes, utilizing the power of the ResNet-152 architecture to encode the images, and applying specific augmentations to the images to make blood vessels more defined.

However, what was surprising and interesting about our results is that our model was not equally accurate in classifying each of the different classes. It was far more accurate when classifying severe and proliferative DR images with accuracy rates higher than 98%, but it is less proficient when classifying no DR images with an accuracy rate of 62%. This skew in the accuracy rates for the different classes results in a more biased model towards the more severe cases of DR. Having said that, the reason for this is likely due to the varying differences between having no DR, mild, and or moderate DR, as these classes are seemingly and can be difficult for even the most experienced doctors to distinguish between. However, high overall recall of our model approximately (as seen in Figure 9), which shows that this model can help for early dr detection.

Overall, the biggest takeaways for our team from this project is the complexity involved in producing a large scale neural network model that is accurate in classification from general fundus scans.

8 Link to Github or Colab Notebook

https://github.com/UltimateForce21/DiabeticRetinopathyCNN

REFERENCES

Manoj Kumar Behera and S. Chakravarty. Diabetic retinopathy image classification using support vector machine. In *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*, pp. 1–4, 2020.

Usharani Bhimavarapu and Gopi Battineni. Deep learning for the detection and classification of diabetic retinopathy with an improved activation function. *Healthcare (Basel, Switzerland)*, 11(1):97, 2022. doi: 10.3390/healthcare11010097. URL https://doi.org/10.3390/healthcare11010097.

Google Brain. Messidor-2 dr grades dataset. Kaggle, 2021. URL https://www.kaggle.com/datasets/google-brain/messidor2-dr-grades. Accessed: 2024-08-14.

Enrique V. Carrera, Andrés González, and Ricardo Carrera. Automated detection of diabetic retinopathy using svm. In *2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, pp. 1–4, 2017.

Kaggle Community. Diabetic retinopathy detection - discussion. Kaggle, 2015. URL https://www.kaggle.com/competitions/diabetic-retinopathy-detection/discussion/15801. Accessed: 2024-08-14.

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

Internation Diabetes Federation. IDF Diabetes Atlas 2021 – 10th edition. https://diabetesatlas.org/idfawp/resource-files/2021/07/IDF_Atlas_10th_Edition_2021.pdf. accessed Jun. 6, 2024.

Rishab Gargeya and Theodore Leng. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*, 124(7):962–969, 2017. ISSN 0161-6420. doi: https://doi.org/10.1016/j.ophtha.2017.02.008. URL https://www.sciencedirect.com/science/article/pii/S0161642016317742.

Google. Classification: Roc curve and auc — machine learning — google for developers. https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc. accessed Jun. 6, 2024.

Varun Gulshan, Lily Peng, Marc Coram, Martin C. Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, Ramasamy Kim, Rajiv Raman, Philip C. Nelson, Jessica L. Mega, and Dale R. Webster. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*, 316(22):2402–2410, 12 2016. ISSN 0098-7484. doi: 10.1001/jama.2016.17216. URL https://doi.org/10.1001/jama.2016.17216.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015. URL https://arxiv.org/pdf/1512.03385.

Jason Hsu, MD. Diabetic retinopathy. *American Academy of Opthalmology*, 04 2024. URL https://eyewiki.aao.org/Diabetic_Retinopathy.

Sohier Dane Karthik, Maggie. Aptos 2019 blindness detection, 2019. URL https://kaggle.com/competitions/aptos2019-blindness-detection.

Guillaume Patry. Messidor-2. ADCIS, n.d. URL https://www.adcis.net/en/third-party/messidor2/. Accessed: Jun. 6, 2024.

Rory Sayres, Ankur Taly, Ehsan Rahimy, and Kimberly Blumer. Using a deep learning algorithm and integrated gradients explanation to assist grading for diabetic retinopathy. *Ophthalmology*, 126(4):552–564, 2019. doi: 10.1016/j.ophtha.2018.11.016. URL https://doi.org/10.1016/j.ophtha.2018.11.016.

Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. *arXiv.org*, 2015. URL https://arxiv.org/abs/1512.00567v3. Accessed: Jun. 6, 2024.

Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alex Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv.org*, 2016. URL https://arxiv.org/abs/1602.07261. Accessed: Jun. 6, 2024.

Shaohua Wan, Yan Liang, 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: https://doi.org/10.1016/j.compeleceng.2018.07.042. URL https://www.sciencedirect.com/science/article/pii/S0045790618302556.

I-Tung Yang and Handy Prayogo. Efficient reliability analysis of structures using symbiotic organisms search-based active learning support vector machine. *Buildings*, 12(4), 2022. ISSN 2075-5309. doi: 10.3390/buildings12040455. URL https://www.mdpi.com/2075-5309/12/4/455.

Abayneh F. Zegeye, Yared Z. Temachu, and Chalie K. Mekonnen. Prevalence and factors associated with diabetes retinopathy among type 2 diabetic patients at northwest amhara comprehensive specialized hospitals, northwest ethiopia 2021. *BMC Ophthalmology*, 23(1):9, 2023. doi: 10.1186/s12886-022-02746-8. URL https://doi.org/10.1186/s12886-022-02746-8.