Diabetic Retinopathy Detection

Aayush Kapoor (2020007) | Devansh Arora (2020053) | Krishnam Omar (2020309)

1 Problem Statement

Diabetic retinopathy is a leading cause of blindness worldwide, affecting over 93 million people. Almost half of Americans with diabetes have some stage of the disease, which can be difficult to detect until it's too late for effective treatment. Current detection methods require trained clinicians and are resource-intensive, which can limit their effectiveness in areas where diabetes rates are high. Automated detection systems using image classification, pattern recognition, and machine learning have made progress, and this competition aims to push them to their limit for maximum impact on improving DR detection. Winning models will be open-sourced.

2 Related Work

This article focuses on enhancing the discriminative power of the feature representation, a multi-scale attention mechanism is used on top of the high-level representation. The model is trained in a standard way using the cross-entropy loss to classify the DR severity level. In parallel as an auxiliary task, the model is trained using the weakly annotated data to detect healthy and non-healthy retina images.(Al-Antary and Arafa, 2021) The study presents a diabetic retinopathy detection system that uses ultra-wide-field fundus photography and deep learning, which is more efficient than conventional fundus photography used in most automatic systems. The researchers used the early treatment diabetic retinopathy study 7-standard field image extracted from ultra-wide-field fundus photography and found that it outperformed the optic disc and macula-centered image in detecting diabetic retinopathy in experiments. The results suggest that ultra-wide-field fundus photography can be a useful tool for automated screening and grading of diabetic retinopathy.(Alyoubi et al., 2020) The article proposes a deep learning method for detecting diabetic retinopathy that uses a regression activation map (RAM) to provide interpretable features. The RAM is added after the global averaging pooling layer of the convolutional neural network (CNN), allowing the model to identify and localize discriminative regions of the retina image and show the specific region of interest in terms of its severity level. This approach provides insights into why the learning model works and is highly desired in practice, as users are interested not only in high prediction performance but also in understanding the insights of DR detection.(Wang and Yang, 2017)

3 Dataset Description

The dataset (Emma Dugas, 2015) we use was created by Google AI and EyePACS, a non-profit organization that aims to prevent blindness through the early detection and treatment of eye diseases. The dataset was released as part of a Kaggle competition in 2015. The dataset contains over 35,000 retinal images, of which approximately 75% have been labeled for diabetic retinopathy severity on a scale of 0 to 4. The images are in JPEG format and have a resolution of 224 x 224 pixels. The images have been preprocessed to remove patient-identifying information and resized to a standard resolution.

We worked on a subset of the given dataset as the dataset was too big. The class distribution of our dataset can be visualised from figure 1 and figure 2.

We plot the Image Histogram(figure 3) over our dataset. The histograms are created for each channel of the images to see the distribution of pixel values for each class.

4 Methodology

We have used CNN based models such as VGG16, Resnet50, InceptionNetV3 in our baseline approach.(Alyoubi et al., 2020) (Oh et al., 2021) The stacked LSTM model is a type of convolutional LSTM neural network designed for processing 2D

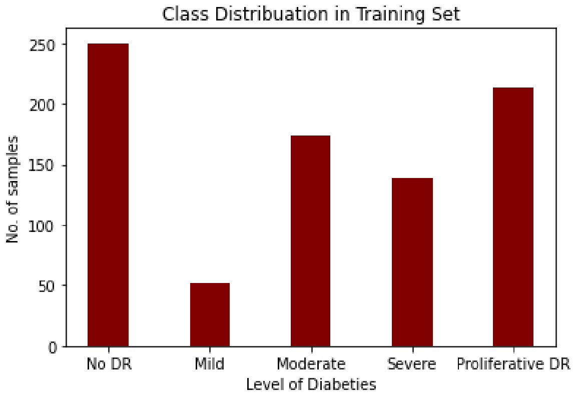


Figure 1: Class Distribution over our Training set

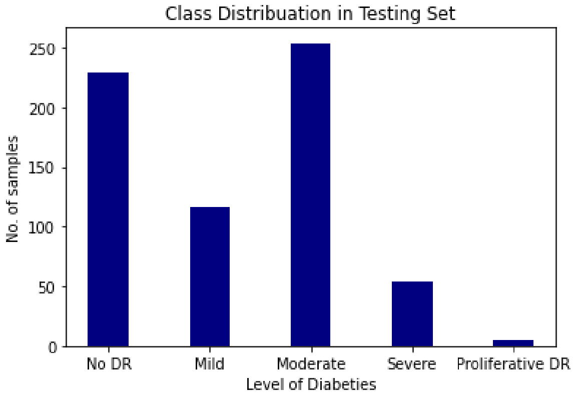


Figure 2: Class Distribution over our Testing set

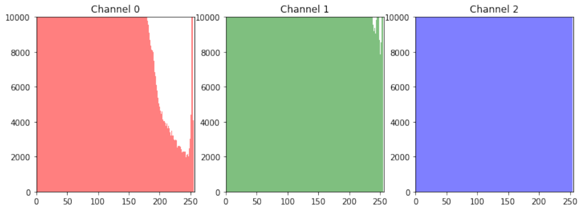


Figure 3: Color distribution over all pixels in our sample dataset

images in a row-by-row and column-by-column manner. The primary processing pipeline of this model includes two convolutional layers followed by two bidirectional LSTM layers for row-wise and column-wise processing of the image. The outputs from both LSTM layers are concatenated and passed through a dropout layer to mitigate overfitting. This mechanism results in a test accuracy of 0.489 and a kappa score of 0.348. The proposed model can be found in figure 4. Confusion matrix on test set for the stacked LSTM is on figure 5.

Now we propose a fusion of Multi scale Attention Network and Vision Transformer Network for the classification of Diabetic Retinopathy Images.

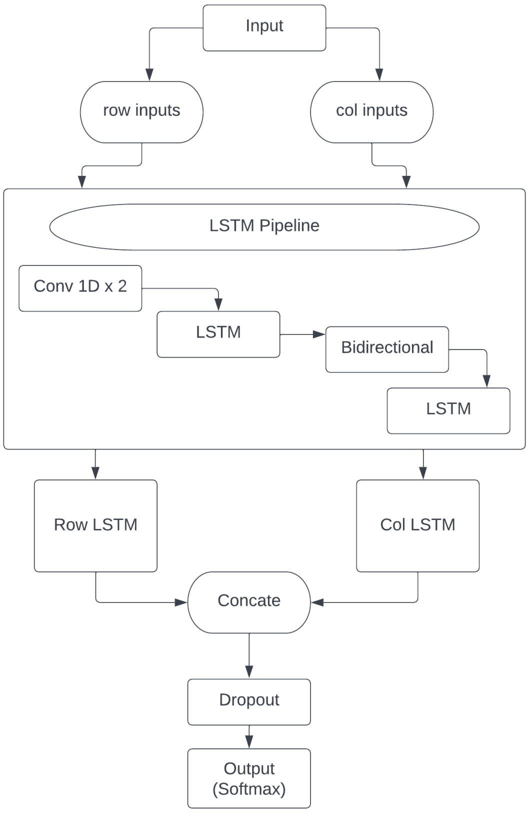


Figure 4: Stack LSTM Architechture

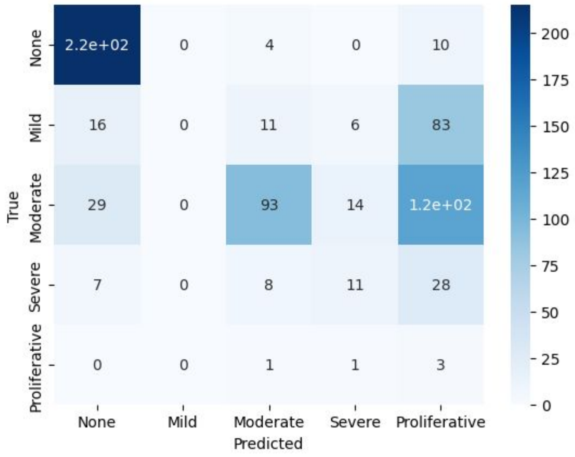


Figure 5: Stack LSTM Confusion matrix on Test set

The Architecture for the proposed model is shown in figure 6. The proposed architecture consists of a multi-modal solution that combines the results of ViTransformer and Multi-Scale Attention Network and applies Late Fusion over them to give the results. The MSA consists of 3 principal components, i.e., The Encoder, Multi-Scale representation, and the Multi-Scale Attention Mechanism. The encoder comprises of Resnet50 model, which captures the images' feature space and has been pre-trained on ImageNet Dataset. A series of convolutions with different kernel sizes processes the feature space given by the encoder network. The different kernel sizes capture the image space's short-term and long-term dependencies. The output of these is then stacked over one other and passed on to a bidirectional LSTM network, which helps calculate the attention weights for each layer. The attention weights define the dependency of which layer is more critical in deciding the result. The final output is then passed on to the last series of convolutions MaxPooling, and dense layers that help in the final prediction of the model. On the other hand, ViT transformer, the input is pre-processed via the ViT feature extractor and then passed on to the model. Finally, after results from both of them have arrived, it gets fused and helps in giving the final prediction.

5 Experimental Setup

Diabetic Retinopathy has been provided with high-resolution images broadly divided into five main categories:

```
0 - Mild
1 - Moderate
2 - No DR
3 - Proliferate
4 - Severe
```

The dataset is organized into TRAIN and TEST folders, each with five sub-folders corresponding to the disease categories. Image paths from each folder are stored in a list and shuffled into the train, validation, and train test split, where 80 percent training and 20 percent validation. Pytorch's Dataset is then applied to this list, which reshapes and applies transformations to ensure uniformity. Data Augmentation techniques such as Normalizing the images using color jitter and Gaussian Filter have been applied to the dataset. After being arranged in their respective sets, this dataset gets passed onto the models described above in batches of 12. The Model reports each epoch's loss, Accuracy, Precision, and F1 scores. Using the trained models, we compute the results for the testing set. Finally, the confusion matrix is plotted for each Model, and the kappa score is the evaluation metric to score the models. This helps to find the degree of reliability among two raters measuring the same quantity.

6 Observations

We have observed that the performance of baseline convolutional neural network (CNN) models in terms of accuracy and kappa score is limited, yielding approximately 40% accuracy and a kappa score of 0.25. This can be attributed to the complex nature of the images and the lack of readily available pre-trained models. In contrast, sequence models that utilize CNNs as their backbone demonstrate improved accuracy compared to traditional models.In the experimental model we stacked Conv1D and LSTM layers over one other and applied attention over those models. This gave an overall accuracy of 49% and a kappa score of 0.348. The incorporation of attention mechanisms in sequence models allows for capturing fine-grained dependencies within the model and focusing on specific layers that may hold greater predictive power. The integration of multi-scale attention mechanism results in an overall accuracy of 63% with a kappa score of 0.48.

8: Upon analyzing the matrix, it is evident that the model exhibits a high degree of accuracy in detecting individuals with no diabetic retinopathy ('No DR'). However, it shows some limitations in accurately identifying the severity level of the condition, particularly in distinguishing between Moderate and Severe types.

7 Error Analysis

The current model significantly increases over the traditional models, but certain drawbacks must be dealt with. The model is very efficient in predicting whether the person has No- DR. Still, it faces inevitable backlash in identifying which stage of DR is present in the person. The model classifies many samples as moderate DR, creating confusion among mild and proliferate. Also, there needs to be more data on severe instances, and the model also falsely classifies proliferate examples with mild and moderate samples. Overall when it comes to the binary classification of being able to tell whether the person has DR or no DR, the model can say to it with 93.7% accuracy.

8 Future Work

We could explore the development of a real-time, early detection system for diabetic retinopathy. The system would use advanced algorithms that can quickly analyze retinal images taken by a fundus camera and provide a diagnosis within a few

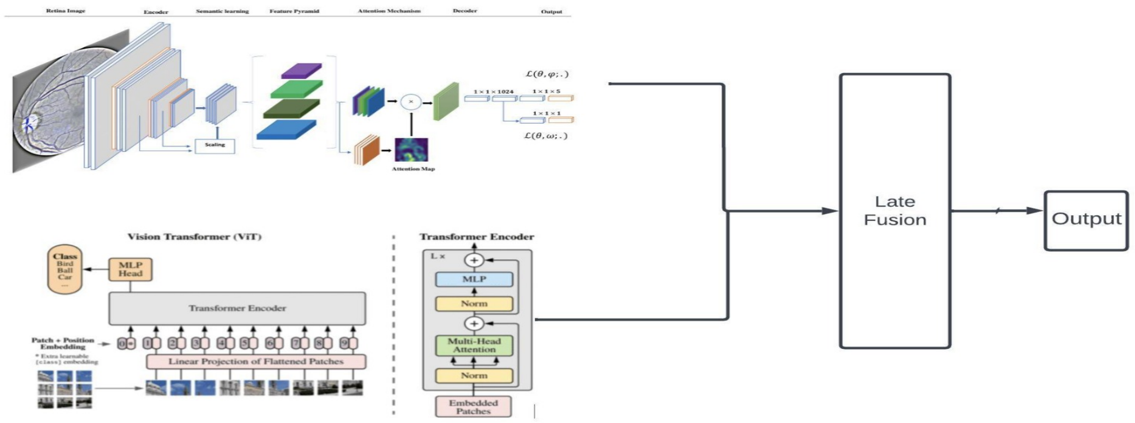
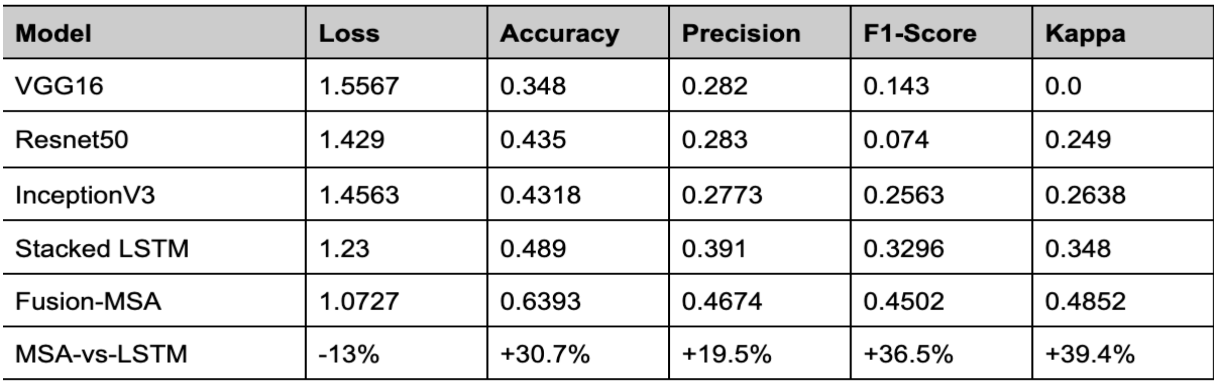


Figure 6: Architecture of the proposed model



Model	Loss	Accuracy	Precision	F1-Score	Kappa
VGG16	1.5567	0.348	0.282	0.143	0.0
Resnet50	1.429	0.435	0.283	0.074	0.249
InceptionV3	1.4563	0.4318	0.2773	0.2563	0.2638
Stacked LSTM	1.23	0.489	0.391	0.3296	0.348
Fusion-MSA	1.0727	0.6393	0.4674	0.4502	0.4852
MSA-vs-LSTM	-13%	+30.7%	+19.5%	+36.5%	+39.4%

Figure 7: Observation table

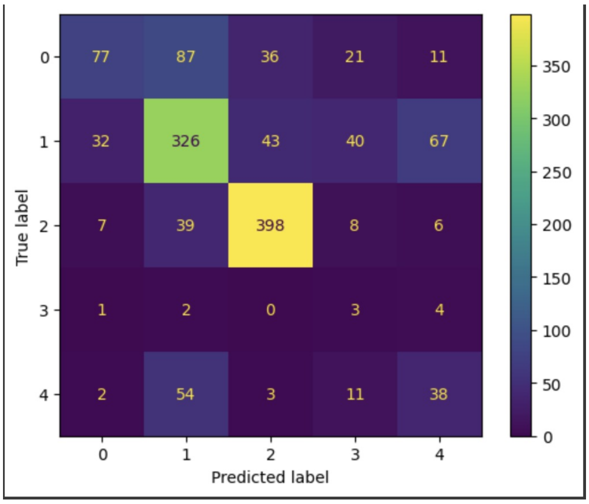


Figure 8: Confusion Matrix for MSA-fusion

seconds of capturing the image. To achieve this, we could explore deep learning techniques that can detect subtle changes in retinal images that indicate the presence of diabetic retinopathy in its early stages. This could involve developing a large dataset of retinal images that include both healthy eyes and eyes with various stages of diabetic retinopathy. The dataset could be used to train and test deep learning models, which could then be fine-tuned for real-time diagnosis.

References

Mohammad T. Al-Antary and Yasmine Arafa. 2021. Multi-scale attention network for diabetic retinopathy classification. *IEEE Access*, 9:54190–54200.

Wejdan L. Alyoubi, Wafaa M. Shalash, and Maysoon F. Abulkhair. 2020. Diabetic retinopathy detection through deep learning techniques: A review. *Informatics in Medicine Unlocked*, 20:100377.

Jorge Will Cukierski Emma Dugas, Jared. 2015. Diabetic retinopathy detection.

Kangrok Oh, Hae Min Kang, Dawoon Leem, Hyungyu Lee, Kyoung Yul Seo, and Sangchul Yoon. 2021. Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images. *Scientific Reports*, 11(1).

Zhiguang Wang and Jianbo Yang. 2017. Diabetic retinopathy detection via deep convolutional networks for discriminative localization and visual explanation. *CoRR*, abs/1703.10757.