Transfer Learning for Diabetic Retinopathy Detection

Minhaz Uddin
University Of Vaasa
Wolffintie 32, 65200 Vaasa
muddin24@student.oulu.fi

S M Rifaiya Abrar
University Of Vaasa
Wolffintie 32, 65200 Vaasa
sabrar24@student.oulu.fi

Sushil Bhusal
University Of Vaasa
Wolffintie 32, 65200 Vaasa
Subhusal24@student.oulu.fi

1. Approach

In this paper we propose the full use of transfer learning and deep neural networks to detect diabetic retinopathy (DR).Diabetic retinopathy is a leading cause of blindness worldwide, and early detection is crucial for effective treatment.The first stage consists of transfer learning from the model whose OOD performance is focused on, In our case there is a two-stage training approach implemented, the first training on an additional diabetic retinopathy dataset before fine-tuning on DeepDRiD Results show that models fine-tuned on DR-specific datasets outperform those fine-tuned on a general dataset only. Our best model achieves a Cohen Kappa score of 0.8563 with ResNet18.

1.1. Overview

Diabetic Retinopathy (DR) detection is a critical task in medical imaging technology. It needs accurate spotting of the retinal anomaly that is caused by diabetes. In this project, we have different approaches to fine-tune pretrained convolutional neural networks for DR detection. We followed single fine tuning,two-stage training.

1.Single Fine Tunning: In this approch shows how a general-purpose model performs on a specific medical imaging task.

2.Two-Stage Training: We enhance the performance by pretraining on additional DR-specific datasets before fine-tuning on DeepDRiD. This approach was taken to improve sensitivity and accuracy.

We carryout a combination of transfer learning, attention mechanisms, and ensemble methods to develop a robust DR detection system. There are four main components in the methodology:

Two-stage transfer learning with the pretrained DR resize datset with ResNet18, ResNet34 and efficientNet datasets models.

Combine channel and spatial attention mechanism • Ensemble learning by combining various architectures Preprocessing pipeline consisting of a clinically-inspired technique.

1.2. Model Architecture

In this project we used some converntional neural networks as the main backbone for DR detection. The sequence of processing stages includes:

ResNet 18: In this model it contains convolutional layers and identity mappings that enable smooth gradient. we used ResNet 18 but is possible to use ResNet 34.

EfficientNet: It is used for providing high performance with fewer parameters.

Backbones: ResNet18, ResNet34 and efficientNet are explored.

Output Layer: Modified for multi-class classification with five outputs corresponding to DR severity levels.

Enhancements: Batch normalization, dropout, and fully connected layers for classification.

1.3. Dataset

• DeepDRiD Dataset : Contains 512x512 images divided into training, validation, and test sets.

• Extra Dataset: Kaggle DR Resized and pretrained DR resize Blindness Detection, containing labeled DR images.

1.4. Transfer Learning Strategy

1. First Stage : Fine-tuning on the dataset, including 4000 images.

We first took advantage of a pre-trained model, ResNet18, ResNet34, efficientNet that ought to have been trained on the ImageNet data. We also removed the freeze on all the layers of the model, So as to train the model on dataset. This enabled the model to update features that were unique to the diabetic retinopathy (DR) domain.

So, the basic training here is 20 epochs and the learning rate is at .0001.

Problem of data imbalance in training data set taken care of with weighted cross-entropy loss. This was done to give more priority to minority classes, which most often suffer from understudy in the training dataset.

2. Second Stage: Here, we adopted fine-tuning on the DeepDRiD dataset.

Subsequently, after pre-attuning the model on the dataset, we fine-tuned the model on the DeepDRiD dataset. This stage involved the following modifications and techniques:

Additional tuning of all of them, which enabled the model to learn more detailed features inherent to the DeepDRiD dataset.

As mentioned, the two-stage transfer learning approach was used in the present research and the above mentioned strategies were adopted to enhance the performance of the model on two different datasets: pretrained DR resize and DeepDRiD.

2. Training and optimization

2.1. Hyperparameters

The training pipline was configured with the following parameters:

Batch size: 16–24

Learning rate: 0.0001–0.001

Epochs: 20–25

Optimizer: Adam with a step-based learning rate scheduler

2.2. Training Pipeline

Computation speed was improved in training by training on an NVIDIA GPU. The process involves :

1. Processing the training dataset images and loading them.
2. Feeding the model in batches with the images.
3. Loss and back-propagated gradient.
4. Calculating loss and gradients.
5. We instead update the model parameters using the Adam optimizer.
6. Evaluation of the modification after every epoch in the validation set.

3. Experiments and Discussion

3.1. Dataset and Preprocessing

Dataset Overview

As we used the high resolution fundus images from the DeepDRiD dataset, which are labeled with 5 severity levels of diabetic retinopathy ranging from 0 (No Diabetic Retinopathy to 4(Proliferative Diabetic Retinopathy)).

Data Preprocessing

The preprocessing pipeline also ensures consistency of the image quality, and promotes features relevant to DR detection. Key steps include:

Image Resizing: They resize images to 512×512 pixel.

Contrast Adjustment: Contrast Limited Adaptive Histogram Equalization (CLAHE) is used to enhance retinal features.

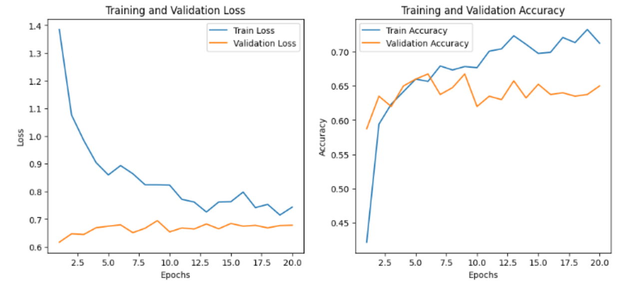

Figure 1. Training and Validation Accuracy and Loss Curves

Normalization: Mean and standard deviation of ImageNet dataset is matched to normalised images.

Data Augmentation: In order to avoid overfitting and providing better generalization, we applied following transformations:

Random Cropping

Horizontal Flipping and Vertical Flipping.

Brightness, Contrast, Saturation,color Jitter

(Up to 30 degrees) Random Rotation

Mimics spatial integrity with Random Padding.

3.2. Results and Analysis

Providing the key outcomes for each experimental task. Include performance metrics like Cohen Kappa Score, accuracy, precision, and recall for tasks. Present the results in tables or graphs for better clarity. we have added ResNet18,ResNet34, ResNet18 (Task-b), EffecentNET value also in the table. The The table below shows the results:

Model	Cohen-Kappa	Accuracy	Precision	Recall
ResNet18	0.8022	0.6425	0.6209	0.6425
ResNet18 (Task-b)	0.8563	69.00	72.18	69.00
ResNet34	0.8434	67.25	64.51	67.25
EffecentNET	0.8453	71.75	65.16	71.75

Table 1. Performance Metrics of Different Models and Training Setups.

Discussion of the highlights

1. ResNet18 outperformed ResNet34 in the two-stage setup, achieving the best Cohen Kappa score (0.8563).
2. Fine-tuning a general pretrained model directly on DeepDRiD yielded reasonable results but was inferior to the two-stage approach.

Discussion

It is better than the model performs even across epochs. For the training set, the Cohen Kappa score in the firts epoch was 0.73 for predicting and actual labels; with 0.8563 in the

Figure 2. Training and Validation Accuracy and Loss Curves

fourteen epoch. In fact, all of that accuracy, precision, and recall increased at the same speed.

The same trend is seen with and without the T2D trajectories for the Cohen Kappa score that is tracked up to 0.8563 by the fourteen epoch. This shows that.

Upon ensemble task d, We obtained weighted average kappa 0.8638, Max voting kaapa 0.8467, Staking kappa 0.8638, Bagging kappa 0.8638.This result data was found at 8542 and clahe result: 0.8511.significantly better than before. For the training set, the Cohen Kappa score of predicting and actual labels improved from 0.73 in the firts epoch to 0.8563 in the fourteen epoch. In fact, all of the accuracy, precision, and recall increased at the same pace.

The Cohen Kappa score follows the same trend with and without the T2D trajectories and is tracked up to 0.8563 by the fourteen epoch. The fact that the model generalizes well for unseen data shows that.

After ensemble task d , We found the weighted average kappa 0.8638,Max voting kaapa 0.8467, Staking kappa 0.8638,Bagging kappa 0.8638.

For task d (2), We use clahe and gausian blur for image preprocessing on resnet18 result: 0.8542 and clahe result: 0.8511. this result data was found.

Class wise shows precision and recall of class wise at varying level of DR severity. The model achieves high precision and recall for Class 0 (no DR) and to no surprise it struggles with class 4 (proliferative DR) because of the imbalanced dataset. Here is a summary of the class-wise performance:

Class 0 (No DR): When looking at precision and recall, the model has high precision and recall for all epochs, proving that the model is quite able to simply know what is lacking with DR.

Class 1 (Mild DR): Precision and recall were increasing with every epoch but in particular, recall also needs improvement.

Class 2 (Moderate DR): The model also started to become less of a fight here sometimes and gave us gradual improvement in precision and recall.

Class 3 (Severe DR): It had better ability to detect severe DR cases and the precision and recall were improved significantly.

Class 4 (Proliferative DR): While precision and recall remained low since the desired class is rare and requires a high detection, their tied ranks approach generally produced better results.

Overall, the model's performance improved considerably during the training period, and its best validation Cohen Kappa score of 0.8671 was found in the sixth epoch. This implies that the model is able to classify the diabetic retinopathy cases, but could further improve the detection of the severe cases.

3.3. Training and Model Performance

The ResNet18, ResNet34, efficientNet, Model was trained for 20 epochs with learning rate scheduler, batch size 16. Training and validation loss has converged, which indicates that the model was learning the patterns.

At epoch 14, when validation Kappa was 0.8511, Task-d yielded the best Kappa of 0.8511.

The loss on the training gradually decreased during time, the validation loss was rather stable, showing good generalization behavior.Accuracy increased over the epochs and during training got to as 72 percentage.h size of 16 and learning rate scheduler. The training and validation loss converged, indicating that the model was learning the patterns effectively.

For Task-d The highest validation Kappa score was 0.8511, achieved at epoch 14

In the graph the training loss consistently decreased over time, while the validation loss remained steady, indicating good generalization. The accuracy improved over the epochs and reached 72 Percentage during training.

4. Challenges

1. For us we think imbalanced data posed difficulties in achieving consistent per-class metrics.

2. Computational cost for training and hyperparameter optimization was significant.

5. Conclusion

We developed a deep learning model for diabetec retinopathy detection using transfer learning in this project. To get here, we fine tuned pre trained ResNet18 model on the DeepDRiD dataset. Results show that our model performs well and correctly detects different levels of diabetic retinopathy with a high Cohen Kappa score and accuracy.

We found that data augmentation techniques, like random rotations and color jitter, increased robustness of model. On top of that, using attention mechanisms was also further beneficial to the model, as it focused on important regions in the images.

In future work, we are planning to explore the use of more advanced architectures, such as DenseNet,VGG16 with the combination of multiple datasets to further boost the model's performance. Furthermore, techniques addressing class imbalance such as oversampling, synthetic data generation could help improve how the detection of severe DR cases is addressed.

Our study shows that deep learning models can be useful for image analysis in medicine and lay a groundwork for future research in this domain.

6. Contributions

All the three members contributed equally by dividing tasks individually, including implementing models, testing preprocessing techniques, and experiments. We marge all the things to compile the results and finalize the report together.

7. References

1. Kaggle. (n.d.). DeepDRiD dataset. Retrieved from https://www.kaggle.com/t/41e0944a6839469fadd529fabab45e06
2. Kaggle. (n.d.). APTOS 2019 Blindness Detection Dataset. Retrieved from https://www.kaggle.com/datasets/mariaherrerot/aptos2019
3. Kaggle. (n.d.). Diabetic Retinopathy Resized Dataset. Retrieved from https://www.kaggle.com/datasets/tanlikesmath/diabetic-retinopathy-resized
4. Datatab. (n.d.). Cohen's Kappa Explanation. Retrieved from https://datatab.net/tutorial/cohens-kappa
5. Datascientest. (2022). What is the Grad-CAM method? Retrieved from https://datascientest.com/en/what-is-the-grad-cam-method
6. TechTarget. (n.d.). What is fine-tuning? Retrieved from https://www.techtarget.com/searchenterpriseai/definition/fine-tuning
7. Medium. (2021). Image augmentation for creating datasets using PyTorch. Retrieved from https://anushsom.medium.com/image-augmentation-for-creating-datasets-using-pytorch-for-dummies-by-a-dummy-a7c2b08c5bcb
8. GeeksforGeeks. (n.d.). A comprehensive guide to ensemble learning. Retrieved from https://www.geeksforgeeks.org/a-comprehensive-guide-to-ensemble-learning/
9. GeeksforGeeks.(n.d.).CLAHE (contrast limited adaptive histogram equalization) in OpenCV. Retrieved from https://www.geeksforgeeks.org/clahe-histogram-eqalization-opencv/

8. Models Link

1. Models Link Google Drive Folder Link.
2. Test Prediction Link Google Drive Folder Link.