Deep Learning to Automate Diagnosis of Diabetic Retinopathy

Alaina Lim[1], Amrit Singh[2] and HongHao Zhen[2]

[1] *Stanford University Department of Computer Science*
[2] *Stanford University Department of Mechanical Engineering*

The use of deep learning algorithms has opened up the doors for increasing accessibility and accuracy in medical diagnosis. Early detection is key for preventing the serious symptoms of diabetic retinopathy. Upon conducting a literature review of deep learning methods on image classification, and specifically diabetic retinopathy, we are highly motivated to reduce the inequities in medical care that people in developing countries face. Using the Kaggle dataset for APTOS 2019 Blindness Detection, we trained and tested VGG-16, GoogleNet, and ResNet-50 with a diverse range of hyperparameters and optimizers, we found that Resnet-50 performed best with 0.842. In the future, given more time and resources, we would be interested in exploring a reliable method to achieve higher accuracy with less training time.

1 Introduction

Diabetic retinopathy is a common and dangerous complication of diabetes, with approximately 1 in 3 diabetes patients contracting diabetic retinopathy during their lifetime. If left undiagnosed or untreated, the disease can cause serious vision loss and eventually, blindness. Diabetic retinopathy and the damage caused by it is permanent, but treatment can help to maintain one's vision and prevent further vision loss. Thus, early detection and treatment is imperative to those suffering from diabetic retinopathy. The current screening process for diabetic retinopathy is rather inaccessible and time-consuming; it requires a well-trained, licensed opthamolgist, costs hundreds of dollars out of pocket, and is prone to misdiagnosis.

Literature shows that much progress has been made towards improving the automation of DR detection through machine learning and image classification. Yet, there remain many challenges with multinomial classification. Such challenges are particularly present for the cases involving early detection, where the identifying features (such as retinal lesions, hard/soft exudates, and haemorrhages) are not as prominent.

The rapid development of AI methodology has proven to be an indispensable, reliable medical assistant. In this paper, we implement a number of a number of deep learning algorithms and architectures including VGG, GoogleNet, and ResNet to automate the diagnoses of diabetic retinopathy. The input of the algorithms are images of patients' eyes, and we use CNNs to output a prediction of the severity of the disease.

2 Related Work

2.1 Background on Diabetic Retinopathy

We conducted an extensive literature review on the different approaches to diabetic retinopathy detection. The five levels of severity to the disease outlined by National Eye Institute: no disease, mild nonproliferative, moderate nonproliferative, severe proliferative, severe proliferative [National Eye Institute, 2022]. Blood vessels in the retina start to swell and distort at the moderate level, and major blockage int the blood vessels as well as leaking into the retinal surface start occurring at the severe nonproliferative stage. Having a medical basis of what diabetic retinopathy does and looks like at different stages was key for feature engineering as well as determining how to transform the images in the dataset. Manually, family doctors diagnose diabetic retinopathy through the features mentioned above, with 87% specificity [Gill JM, 2004].

2.2 Machine Learning Approaches to Diagnosis

Convolutional neural networks have proven to be an incredibly helpful tool in medical image classification in recent years. In 2018, Kwasigroch et al. developed a deep convolutional neural network with a novel class coding technique to minimize the difference between the predicted and target score in the objective function [A. Kwasigroch and Grochowski, 2018]. The basic architecture consisted of a convolutional layer, a ReLU layer, a MaxPooling layer, a dropout layer, and finally a fully connected layer. On a dataset of over 88,000 images of diabetic retinopathy, Kwasigroch et al. produced a model that was able to achieve 82% accuracy in detecting diabetic retionpathy and a 51% in accuracy on the stage of the disease [A. Kwasigroch and Grochowski, 2018]. In 2019, Rehman et al. used a variety of classification tools

including AlexNet, VGG, SqueezeNet, and a custom 5 layered CNN model to classify diabetic retinopathy. The authors created a 5 layered CNN, with each layer having its own specification. The first two layers a process of convolving kernels and reducing their outputs with a pooling layer. The last three layers were fully connected neural network layers with 100, 50, and 10 neurons respectively. The authors' custom 5 layer model performed the best against the baseline, pre-trained models, resulting in 98.15% classification accuracy [Rehman and Rizvi, 2019]. Most recently, in 2021 Dai et al. implemented a deep learning system called DeepDR to detect diabetic retinopathy in a range of stages. DeepDR, a transfer learning assisted multi-task network, consisted of three sub-networks. Each network served a completely different purpose: assessment of the image quality, lesion awareness, and the actual grading of the disease [Dai, 2021]. The authors were able to achieve 94.3%, 95.5%, 96.0%, and 97.2% accuracy on mild, moderate, severe, and proliferative images [Dai, 2021].

3 Dataset and Features

The data used in this study consists of approximately 3600 retina images with varying brightness, focus, and magnification. Each image has a class label ranging from 0 (no symptoms) to 4 (severe DR), with each class being fairly balanced and reflecting the aforementioned levels of diabetic retinopathy severity. The dataset is split into training and validation sets of ratios 0.75 and 0.25, respectively. Prior to training, images were transformed by randomly selecting a central pixel and zooming out such that each transformed image has resolution 224×224 pixels. This minimizes the likelihood of the neural net over-fitting on vascular strictures, which are relatively similar between different retina images, but whose overall trajectory is not a good measure of DR severity.

One defining symptom of diabetic retinopathy is the presence of lesions and nodules on the retina. These appear as small yellow spots that can be difficult to visualize. Applying a Fourier transform to each image and subjecting the transformed image to a high-pass filter would isolate the high frequency components that correspond to abrupt changes in colour (such as edges and spots). The filtered image is then transformed back to Cartesian space, and the lesions and nodules appear to be accentuated, while other non-symptomatic features (such as blood vessels) are suppressed as shown in Figure 1. We postulate that adding a small number of these transformed images to our training dataset could improve the accuracy of our model.

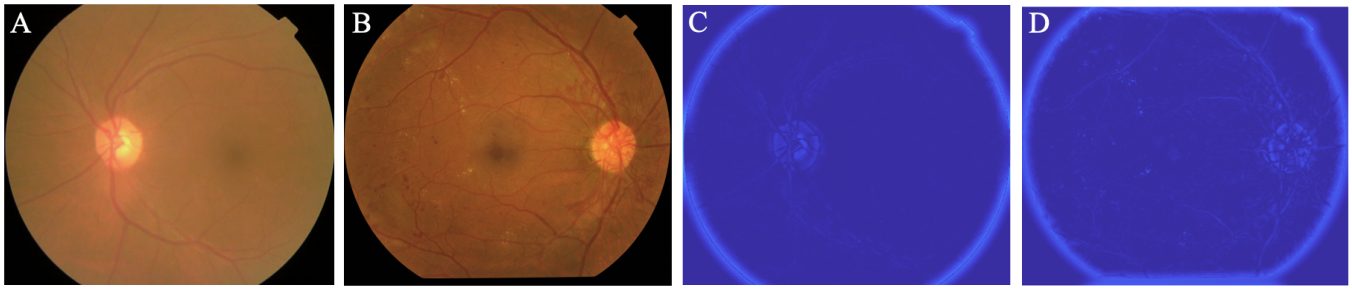

Figure 1: *(A) Class 0 original image. (B) Class 4 original image. (C) Class 0 image after Fourier transform. (D) Class 4 image after Fourier transform. Nodules and lesions on the retina are accentuated in the Class 4 transformed image.*

4 Methodology

Classification of images is always highly complex, with a high degree of pixel varition within each class label. Less complex classification techniques such as multinomial logistic regression would not be suited to this application as they do not capture sufficient variance, while the use of a Gaussian mixture model would be inappropriate since the raw data is not Gaussian in nature. A Convolutional Neural Network is selected as the most suitable classification tool for this application due to its flexibility and ability to capture high degrees of variance. Although this property makes neural networks susceptible to overfitting, careful tuning of hyperparameters and optimization of the network structure can minimize such behaviours.

4.1 VGG

In 2014, Karen Simonyan and Andrew Zisserman from the University of Oxford's Visual Geometry Group introduced VGG for large-scale image recognition. VGG improves from past architectures such as AlexNet with its use of large kernel-sized

filters and an increase in the number of kernel-sized filters, rather than the use of one large filter. VGG-16 consists of 13 convolutional layers and 3 fully connected layers to total 16. The model takes images, and the convolutional layers use a minimal receptive field, 1x1 convolution filters, and ReLU units which significantly reduce training time. By increasing the number of nonlinear layers, the model is able to learn more complex features efficiently. The network was able to achieve 92.7% accuracy on the ImageNet dataset [Karen Simonyan, 2014]. We first used a pretrained VGG-16 to verify the results were comparable to that of GoogleNet and Resnet. We implemented the model, where we found there was negligible effect with switching between ReLU, Softmax, and Linear activations on certain layers.

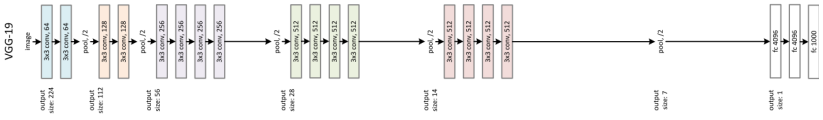

Figure 2: *Overview of VGG architecture [He et al., 2015]*

4.2 GoogleNet

GoogleNet, the first version of the Inception Networks, was developed by a group at Google in 2014 in "Going Deeper with Convolutions". Totalling 22 layers, GoogleNet uses 1x1 convolutions to minimize the number of parameters and increase depth [Szegedy et al., 2014]. Furthermore, the 9 inception modules are sometimes sandwiched between max-pooling layers, which reduces the size of the input. The auxiliary classifier consisting of an average pool, convolutional, fully connected, dropout, and linear layers helps with regularization and prevents overfitting [Szegedy et al., 2014]. GoogleNet requires less memory usage and is generally more cost efficient than other deep learning models. The model won the 2014 ILSVRC image classification challenge. We used weights from a pretrained GoogleNet model and implemented GoogleNet, with a focus on the inception module. For the hidden layer, we tried Softmax and Sigmoid, but quickly realized ReLU led to the best performance and efficiency.

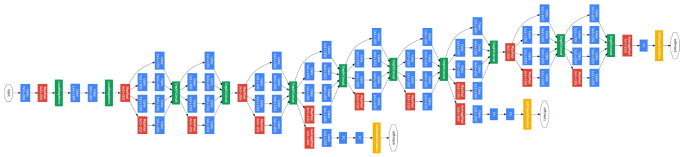

Figure 3: *Overview of GoogleNet architecture [Szegedy et al., 2014]*

4.3 ResNet

Residual Network, or ResNet, was groundbreaking when it was first proposed in 2015 by Microsoft Research. Skip connections, as per their name, connect layers to other layers further in the model's architecture skipping some layers creating residual blocks. This helps prevent the vanishing gradient problem, inherent to training many deeper neural networks, from occurring [He et al., 2015]. ResNets are made up of these residual blocks. ResNet went on to win CVPR 2016 Best paper and first place in multiple tasks for ImageNet competitions. We first ran a pretrained ResNet-18, then used those weights for ResNet-50.

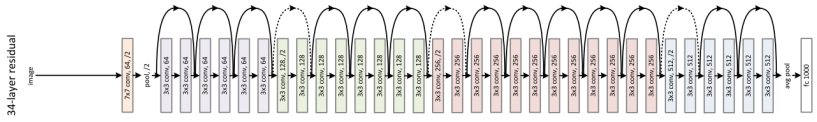

Figure 4: *Overview of ResNet architecture [He et al., 2015]*

5 Results and Discussion

5.1 Hyperparameters and Optimizers

Prior to performing a detailed analysis of each model, we investigated the impact of different pertinant hyperparameters on the GoogleNet model's training performance and overall classification accuracy. Figure 5 outlines the effect of varying learning rates and batch sizes. In general, a learning rate of 0.001 was found to be ideal, with a higher value exhibiting too much volitility, while a lower value increased training time and exhibited significantly lower accuracy, potentially due to its convergence to a sub-optimal local minima.

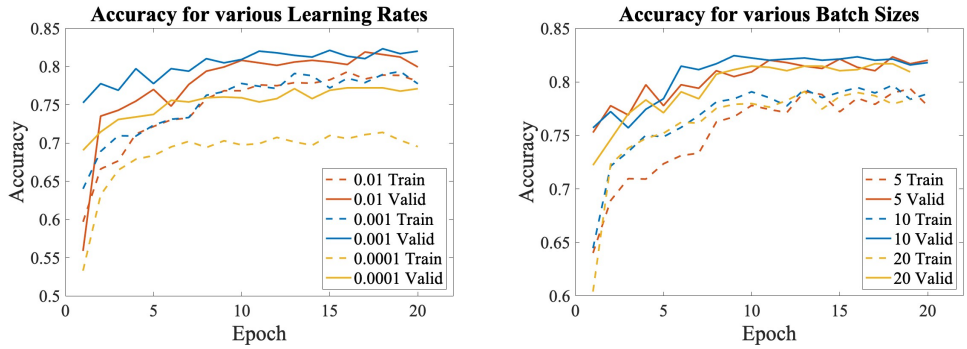

Figure 5: *Effect of varying the learning rate and batch size on the accuracy of the GoogleNet model.*

Our existing experiments with the ResNet-18 model exhibited an increase and subsequent decrease in accuracy during training, indicating a potential issue with overfitting. This was adressed by decreasing the batch size to 5 training examples, resulting in greater noise and variation between samples, providing implicit regularization and directing the model towards a broader, more general minima. The effect of batch size on accuracy was less siginficant for the GoogleNet model, as shown in Figure 5, although the smaller batch sizes of 5-10 performed slightly better.

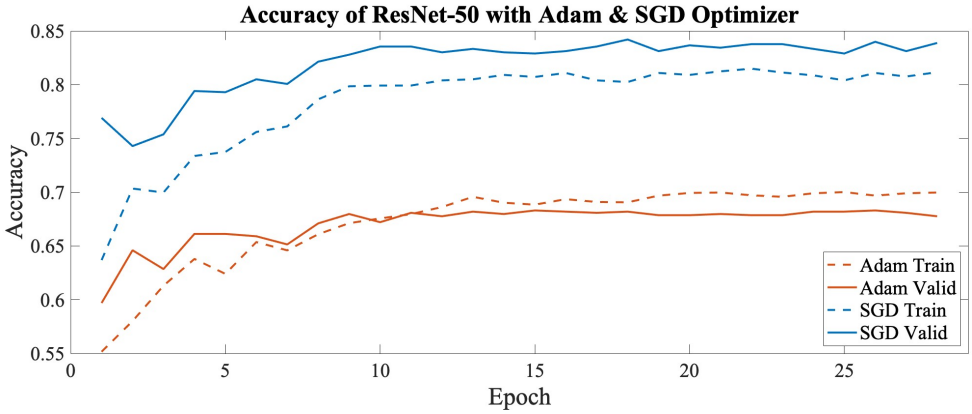

Figure 6: *Accuracy of ResNet-50 with Adam and SGD optimizers.*

We subsequently investigated the effect of different optimizers on the accuracy of the ResNet-50 model, focusing on the traditional Stochastic Gradient Descent (SGD) and increasingly popular Adam optimizers. The SGD optimizer was found to significantly outperform the Adam algorithm (Figure 6), with a performance improvement of nearly 16%. The primary advantage of the Adam algorithm, which calculates adaptive learning rates for individual parameters based on the first and second moments of the gradients, is its rapid convergence rate. For our application, however, run-time performance was almost identical, with the SGD algorithm's better generalization resulting in a significantly more accurate classification.

5.2 Performance of Different Models

Of the 4 different models explored in this study, the modified Resnet-50 with SGD optimizer, a learning rate of 0.001, and a batch size of 5 was the most accurate. Accuracy was calculated based on the percentage of validation examples that are correctly classified as per the provided labels.

Model	Loss	Accuracy
VGG-16	0.434	0.838
GoogleNet	0.472	0.823
ResNet-18	0.432	0.831
ResNet-50	0.467	0.842
ResNet-50 w/ Adam	0.858	0.683

Table 1: *Models with corresponding loss and accuracy.*

We believe this result can be attributed to the ResNet-50's advanced architecture, as well as our use of the weights obtained from the simpler ResNet-18, which may have minimized overfitting and allowed the model to generalize more broadly to previously unseen images.

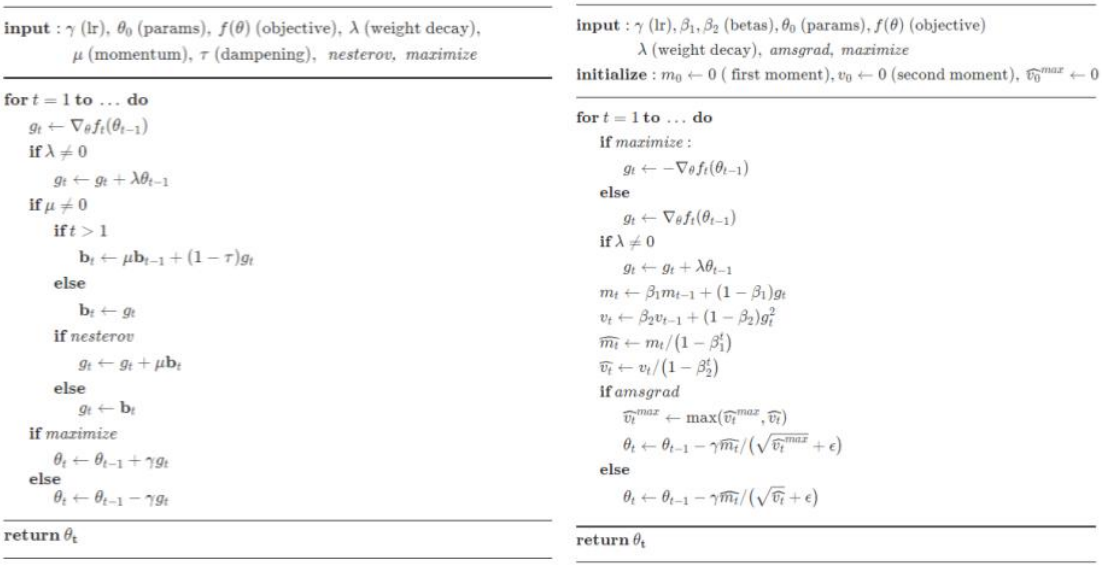

Figure 7: *Left, SGD optimizer. Right, Adam optimizer. Algorithms from PyTorch.*

5.3 Early Detection

A key goal of this project was to improve the detection rate of early stage diabetic retinopathy in the field. We measure this metric by the number of DR-positive cases (class 1-4) that are labelled as any of class 1-4. Our best model detects diabetic retinopathy with an accuracy of 97%, thereby validating Convolutional Neural Nets as a viable means of flagging potential DR cases for further evaluation by a trained physician.

6 Conclusion and Future Work

In this paper, we explored the problems with diagnosing diabetic retinopathy as well as a multitude of fast, cost-efficient solutions. After learning about the deep learning techniques that can be applied, we decided to implement VGG-16, GoogleNet, ResNet-18, and ResNet-50. In the end we found ResNet-50 performed best with an accuracy of 0.842. For next steps, we would like to better understand why our model is unable to classify certain levels of diabetic retinopathy as others, and how to improve this issue. We would use different optimizers and world class architectures that utilize transfer learning for image classification, like Inception-V3 (a more advanced version of GoogleNet). In addition, augmenting the dataset and exploring more advanced transformations (such as further refinement of the Fourier transformed-images) could possibly provide a wider range of training data for the models to learn from.

7 Contributions

As a group, we met up multiple times with and without Jake our TA mentor to scope out the project, work on bugs in the code and the process of setting up GCP, assign tasks, and discuss ideas.

Alaina focused on literature review, exploring background information about diabetic retinopathy and different image classification methodologies. She worked on the VGG and GoogleNet models and running their training.

Honghao setup the GCP VM and performed experiments on the impact of optimizers, in addition to training the ResNet models.

Amrit wrote code to perform Fourier transform and filtering of select training examples, in addition to carrying out experiments on the impact of batch size and learning rate.

Bibliography

A. Kwasigroch, B. Jarzembinski and M. Grochowski (2018). "Deep CNN based decision support system for detection and assessing the stage of diabetic retinopathy". In: *2018 International Interdisciplinary PhD Workshop*, pp. 111–116. URL: doi:10.1109/IIPHDW.2018.8388337.

Dai L., Wu L. Li H. et al. (2021). "A deep learning system for detecting diabetic retinopathy across the disease spectrum." In: *Nature Communications*. URL: https://doi.org/10.1038/s41467-021-23458-5.

Gill JM Cole DM, Lebowitz HM Diamond JJ (2004). "Accuracy of screening for diabetic retinopathy by family physicians." In: *Annual Family Medicine*. URL: doi:10.1370/afm.67.

He, Kaiming et al. (2015). "Deep Residual Learning for Image Recognition". In: DOI: 10.48550/ARXIV.1512.03385. URL: https://arxiv.org/abs/1512.03385.

Karen Simonyan, Andrew Zisserman (2014). "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: *2019 Amity International Conference on Artificial Intelligence (AICAI)*. URL: https://doi.org/10.48550/arXiv.1409.1556.

National Eye Institute (2022). "Diabetic Retinopathy". In: *National Institute of Health*. URL: https://discord.com/channels/1024747154218172476/1024747154838921338/1051034635892699196.

Rehman S. H. Khan, Z. Abbas Mobeen-ur and S. M. Danish Rizvi (2019). "Classification of Diabetic Retinopathy Images Based on Customised CNN Architecture". In: *2019 Amity International Conference on Artificial Intelligence (AICAI)*, pp. 244–248. URL: doi:10.1109/AICAI.2019.8701231.

Szegedy, Christian et al. (2014). "Going Deeper with Convolutions". In: DOI: 10.48550/ARXIV.1409.4842. URL: https://arxiv.org/abs/1409.4842.