IJNRD.ORG ISSN : 2456-4184

International Research Journal
IJNRD
Research Through Innovation

INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Prognostication of Diabetic Retinopathy using AlexNet

[1]D. Umamaheswari, Research Scholar, Annamalai University
[2]Dr. N. Nacammai, Associate Professor, Department of Electronics and Instrumentation, Annamalai University
[3]Dr. S. Anita, Associate Professor, Department of Electronics and Communication,
St. Anne's College of Engineering and Technology

Abstract–Diabetic Retinopathy (DR) is one of the complicationswhich affects the retina and may lead to blindness in diabetes mellitus patients which can be avoided if diagnosed and detected early. Though it is often asymptotic if detected early is very much treatable. A computer vision-based algorithm can help the doctors and the patients for a faster and more precise diagnosis for treatment. Such algorithms can potentially have better accuracy in detecting different stages of the disease. However, developing such algorithms can be computationally expensive and to some extent complex in terms of extracting highly non-linear features. Applying deep learning in such scenarios increases the problem-solving capacity of the system significantly. Deep Learning algorithms have their own challenges often being dependent on corpus of labelled data. In the medical imaging field getting such large amount of labelled data can be expensive and time consuming but once completed and optimised would give a robust system for diagnosis. With image level annotation most DR gradings task are considered as Deep Learning classification problems. The model is also trained using a weakly annotated data for healthy and non-healthy retina images. Outstanding result on experimenting this model were achieved in the public dataset of Kaggle The DR related lesions have not fully been explored with Deep Learning perspectives.In this study we have used image-level and patch-level annotations to build a robust framework for the classification of DR severity grading. This framework using the feature map of AlexNet gives us promising results in terms of Accuracy 80%.

***Index Terms* – Diabetic Retinopathy, Deep Learning AlexNet,**

I. INTRODUCTION

Diabetes being a chronic disease worldwide affects one out of eleven adults globally. Around 40-45% diabetes patients have a good chance of developing the diabetic retinopathy (DR) [2]. Diabetes Mellitus is a disorder which causes high chronic concentration of glucose in the blood [1]. In an estimate, more than 370 million people worldwide have a high possibility of being affected by this disease. The estimated indicate that this number can go as high as 600 million by the year 2040[1]. If this condition is not detected in the early stages, the diabetic retinopathy could potentially cause blindness [1]. The consultation of an ophthalmologist or an optometrist is required within the 3-5 years in diabetes type 1 patients after its onset.

A blood sugar control, healthy diet and lifestyles are recommended precautionary measures to avoid DR developments [7]. DR at its early stages is usually asymptomatic and often goes undetected until patients feel vision related problems such as distortions, blurs, or floaters [7]. This makes the detection of DR in its early stages highly significant for the diagnosis as well as the treatment of the patients [7].

An automatic system with a deep learning algorithm for the detection of the grading of DR severity would help reduce the burden on the medical professional to diagnose and on the other hand the efficiency would help them to treat more patients. The model aims to classify the DR grading into class in terms of DR severity [2]. There are 5 different classes onto which we could classify the DR severity. The 5 classes being: no DR, mild, moderate, severe, proliferative. To some extent this model can be made into a binary classification by fusing categories to get non-referable which is no to mild DR or (DR and no DR) versus referable which is moderate to worse DR [2].

Considering the severity of the conditions, the healthcare systems follow a systemic rule of screening the diabetes patients with test for Diabetic Retinopathy [3]. The method used in these procedures often include telemedicine and include highly trained medical professionals in taking retinal images of the patients. The retinal images are then sent to the ophthalmologist for diagnosis [3]. However, such healthcare systems are limited in terms of equipment and dedicated specialised personnel.

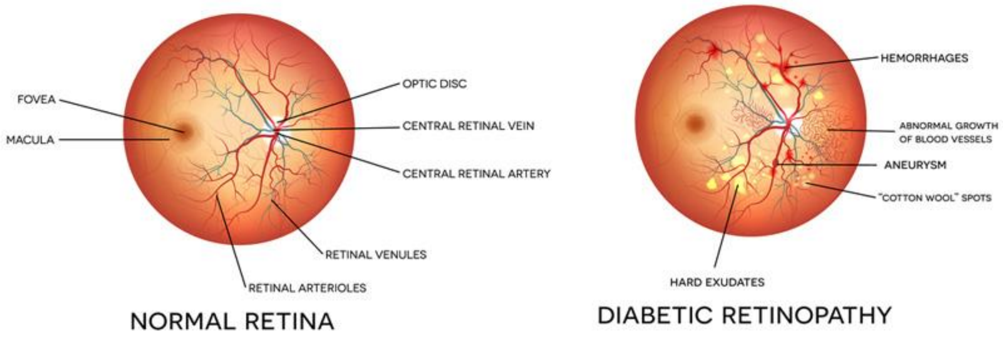


Figure 1. Normal Retina vs Retina Diagnosed with Diabetics

Conventional machine learning techniques require an expert to identify the features manually. Such conventional methods depend heavily on the expert's accuracy on the feature extraction [8]. Recent developments in deep learning have been widely appreciated and applied in the domain of medical image analysis [8]. The previously complex high-level features are increasingly more understandable in advancing deep learning algorithms [8].

The recent developments in deep learning have expanded the capacity of those algorithms beyond human capabilities [5]. A range of new application domains including retinal imaging analysis with not seen before specificity and sensitivity in detection and classification have been made possible with deep learning. The Convolution neural networks have proved to be a powerful tool and has been increasingly popular among researchers for DR grading [2]. Deep learning algorithms in various domain have shown to have an edge over the conventional techniques [1]. In retinal image analysis, several researchers and communities have developed algorithms to automate a computer aided analysis on retinal fundus images [1]. The detection of diabetic retinopathy is one of such conditions that can be detected with such algorithms[1].

AlexNet is one of the most powerful models in the object detection domain with high accuracies on challenging datasets. AlexNet has a huge potential in the domain of computer vision and artificial intelligence technology. The AlexNet by its architecture has been a leading model for object detection problems. The performance by AlexNet models have so appealing in recent times that it may replace the CNNs for image-based problems. The results from AlexNet could be record-breaking on highly challenging datasets.

The highlight of each section of this article is summarised below.In this study we describe the development and validation of AlexNet for DR Screening, Related Works, Data Sets, Proposed Architecture, Results, Conclusion and Future works

II. LITERATURE REVIEW

Yehui Yang et al. (2021) The authors of this paper collaboratively used patch and image level annotations in the classification of DR severity grading. This paper presents an optimised robust framework bilaterally exchanging the information in terms of fine level lesion and image level grade. Such a framework offers to exploit more DR grading features discriminatively. The result from this article suggests outperforming the advanced modern technologies and 3 ophthalmologistspracticing over 9 years. When tested on various distribution like labelled and camera data, the algorithms prove to be resilient in real world scenarios. The CLPI in this study on extensive experiments proves to have competing performance with SOTA algorithms and other senior ophthalmologists. The paper also shows its robustness of CLPI for classification of DR grading under real world scenarios [2].

Pengxiao Zang et al. (2021) This article proposed a DR classification framework based on Conventional Neural Network (CNNs) using OCT and OCTA.DcardNet (adaptive dropout rates) is used in this framework of continuously and densely connected neural network. To address overfitting this article also proposes a adaptive label smoothing. By the guidelines of International Clinical Diabetic Retinopathy Scale three different classifications are made. On a higher level this model classifies DR asreferable (Category 1) and non-referable (Category 2). Further, on the 2ndlevel the model can classify the eye as non-DR, NPDR (non-proliferative DR), or PDR (proliferative DR). The final level classification is done as no DR, mild to moderate NPDR, severe NPDR and PDR. The adaptive label smoothing helps in network's convergence focused more on mis predicted data. The trained model following the mentioned has better chance of handling overfitting. Such CMA generations and 3 levels of DR improves diagnosis and treatment. 95.7%, 85.0%, and 71.0 were obtained as the classification accuracy at these 3 levels respectively [4].

Zubair Khan et al. (2021) To speed up training time and convergence of model the authors have focused on classification using the lowest possible learnable parameters. A VGG-NiN model is used by stacking VGG16 as a SPP(spatial pyramid pooling layer) with NiNB(networkinnetwork) to achieve scale invariant and highly nonlinear Deep Learning model. By the virtue of SPP layerthe DR image can be processed at any scale by the VGG-NiN model. The NiN stacking helps classify better by adding a extra non linearity to the model. The results from this study suggests to have better accuracy and resource utilisation compared to state of the art technologies.

Shuqiang Wang et al. (2021) In this article themodel uses a semi supervised multichannel-based generative adversarial network or MGAN for DR grading. A series of subfundus images with respect to the scattering DR features are generated using the multichannel generative model. The MGAN minimises the dependence of labelled data by using high-resolution fundus images without any compression. The MGAN could achieve that by identifying inconspicuous lesion features. Effective results are obtained when the model is experimented with Messidor dataset. The model suggests to outperform in accuracy, AUC, sensitivity, and specificity.

Mohammad T. Al-Antary AND Yasmine Arafa et al. (2021) This article proposes MSA-Net or Multiscale attention network for the DR classification. The retianl image is embeded in the encoder network in a high-level representational space, the mid-level and high-level features are used to improve the representation. The retinal structure of different locality is incorporated by a multi-scale feature pyramid. A multi scale attention mechanism is used to enhancediscriminative power inrepresenting the features. The DR severity level classification done using a cross entropy loss method. The model is also trained using a weakly annotated data for healthy and non-healthy retina images. Outstanding result on experimenting this model were achieved in the public datasets of EyePACS and APTOS.

Mohamed M. Abdelsalam et al. used the Optical Coherence Tomography Angiography (OCTA) to detect diabetes. The authors explore the use of a simple Support Vector Machine method in detecting non-proliferative diabetes. Regardless, this approach was able to achieve 98.5% accuracy.**EmanAbdelmaksoud et al.**, in their paper from 2021, propose the use of an effective visual feature extraction step through the use of U-Net. This is followed by producing further statistical features. These are then fed into a Support Vector Machine algorithm to achieve a very high accuracy value of 95%.**Cam-Hao Hua et al.** utilized the concept of weight sharing and reverse cross-attention to develop their convolutional network. Using these two techniques, the authors were able to achieve a Quadratic Weighted Kappa rate of 90.2%. **Asra Moment Pour et al.,** in their 2020 paper, used a comparatively simpler CLAHE or Contrast Limited Adaptive Histogram Equalization method as the pre-processing step. By then leveraging the EfficientNet architecture, the authors were able to achieve an area under the curve (AUC) value of 0.936.

III. PROPOSED FRAMEWORK

Machine Learning Models learning simple feature with only a few thousand images were common a few years ago. The real-life problems are far more complex and may need huge amount of data to get a model to even be closer to the real behaviour. The boom in data collection and accessibility of large dataset like ImageNet with hundreds of millions of images with labels has facilitated the platform to develop advanced deep learning models such as AlexNet.

The Convolutional Neural Networks (CNNs) convincingly has been the popular algorithm in addressing object recognition problems. They are powerful models that are easy to use and adaptable to a range of complex problems. When the data amount is considerably big the chances of overfitting are slim. To some extent identical in terms of performance in feedforward neural networks of similar sizes. The main problem in such CNNs is their ability to address the problem when the resolution of the imagesare high. When the scale of amount of datais in the range of that of the ImageNet there comes a need to innovate in GPU optimization and performance related improvisations.

Once of such innovation is the AlexNet which gives significant improvements in the field of deep learning and computer vision applications. This has been proved by a large margin at the 2012 ImageNet LSVRC-2012 competition 15.3% VS 26.2% (second place) error rates. The architecture has stacked convolution layers with more filters per layer and deeper network when compared to the architecture of LeNet by Yann LeCun et al. The architecture consists of the following

- 5×5,3×3, convolutions,
- max pooling,
- dropout,
- data augmentation,
- ReLU activation function,
- SGD with momentum

After every convolutional and fully connected layers the architecture has a RelU activations. The network is split into two pipelines as the AlexNet was trained on 2 Nvidia Geforce GTX 580 GPUs for 6 days simultaneously.

Key Highlights

1. Reluas aactivation function is used to give nonlinearity and it helps in accelerating the speed by 6 times when compared to tanh activation function with the same accuracy.
2. To deal with overfitting dropout is used instead of regularisation. With dropout rate of 0.5 the training time is doubled.
3. To reduce the size of the network overlapping pooling is used. The Error rates are reduced by 0.4% and 0.3% in the top-1 and top-5 respectively.

Data

ImageNet is a huge dataset with over 15 million labelled images of highresolution in over 22000 categories. The images were internet sourced and were labelled using Amazon’s Mechanical Truck crowd sourcing tool by human labellers. The ImageNet Large-Scale visual recognition challenges (ILSVC) conducted since 2010 was an annual challenge as part of the Pascal Visual object challenge. Around 1000 images from 1000 category were obtained from the ImageNet for the ILSVC. Around 1.2 million images were used for training, 50 thousand images for validation and 150 thousand images were used for testing. All the images at ImageNet are down sampled to about 256x256 fixed resolution to account for the variable resolution nature of ImageNet dataset. Hence the image is rescaled and then cropped out of the central 256x256 patch from the resulting image.

Parameters

The architectural design of AlexNet has eight layers and learnable parameters. The model constitutes max pooling layer, 3 fully connected layers and reluas activation function (excluding theoutput layer).The use of Relu has improved performance about 6 times. To avoid overfitting dropouts are used.

AlexNet.

The innovations in AlexNet that are as follows :

- **ReLU Nonlinearity** – advantage in training time with over 6 times faster than CNNs with Tanh. 25% error on the CIFAR-10 dataset.
- **Multiple GPUs.** Half of the total neurons in one GPU and other half on the other. Cuts down training time and improves efficiency.
- **Overlapping Pooling.** It helps to reduce the error rate by 0.5% and the chances of overfitting reduced.

The Overfitting Problem: With over 60 million parameters the following were adapted to reduce the overfitting.

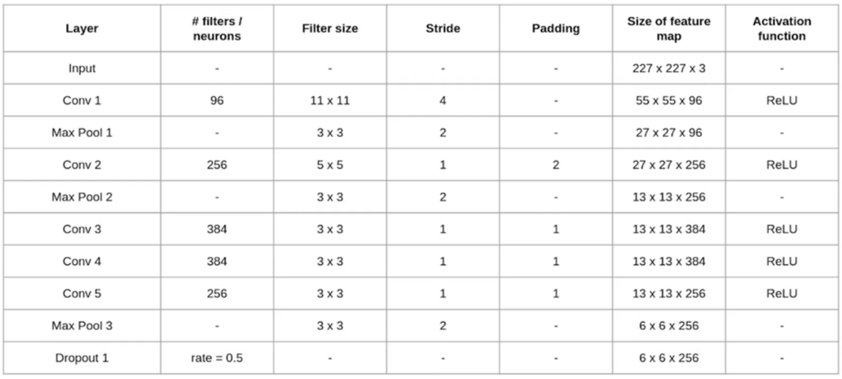
Dropout. The model improves the robustness in each neuron for retaining the essential details by turning off the neurons with a fixed probability (say 0.5). But this increases the time for training for the convergence of the model.

Data Augmentation. The data was made more varies with label preserving transformations like image translations and horizontal reflections increasing the count ofthe data for training by 2048 folds. PCA or principle component analysis were done on RGB pixels values which resulted in reducing the error rate on the top1 by around one percent.

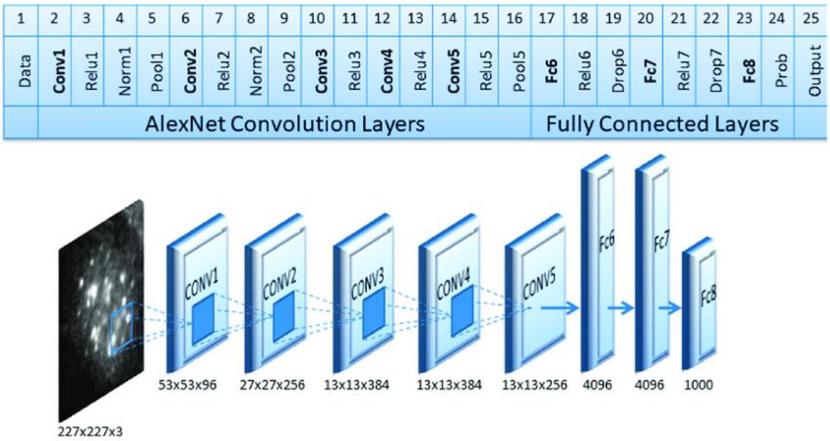
Result. AlexNet overperformed at the ImageNet competition (2010) achieving 37.5% top-1 error and a 17.0% top-5 error. It could recognize theoffcentre objects and competitively most of the top five classes.

Alexnet Architecture

The architecture of AlexNet in deep so puddings were introduced to address avoid reducing size of feature maps.The 227X227X3 is the input size of image for the model.



Layer	# filters / neurons	Filter size	Stride	Padding	Size of feature map	Activation function
Input	-	-	-	-	227 x 227 x 3	-
Conv 1	96	11 x 11	4	-	55 x 55 x 96	ReLU
Max Pool 1	-	3 x 3	2	-	27 x 27 x 96	-
Conv 2	256	5 x 5	1	2	27 x 27 x 256	ReLU
Max Pool 2	-	3 x 3	2	-	13 x 13 x 256	-
Conv 3	384	3 x 3	1	1	13 x 13 x 384	ReLU
Conv 4	384	3 x 3	1	1	13 x 13 x 384	ReLU
Conv 5	256	3 x 3	1	1	13 x 13 x 256	ReLU
Max Pool 3	-	3 x 3	2	-	6 x 6 x 256	-
Dropout 1	rate = 0.5	-	-	-	6 x 6 x 256	-



1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Data	**Conv1**	Relu1	Norm1	Pool1	**Conv2**	Relu2	Norm2	Pool2	**Conv3**	Relu3	**Conv4**	Relu4	**Conv5**	Relu5	Pool5	**Fc6**	Relu6	Drop6	**Fc7**	Relu7	Drop7	**Fc8**	Prob	Output
	AlexNet Convolution Layers															Fully Connected Layers								

Figure 2. Layers in AlexNet Architecture

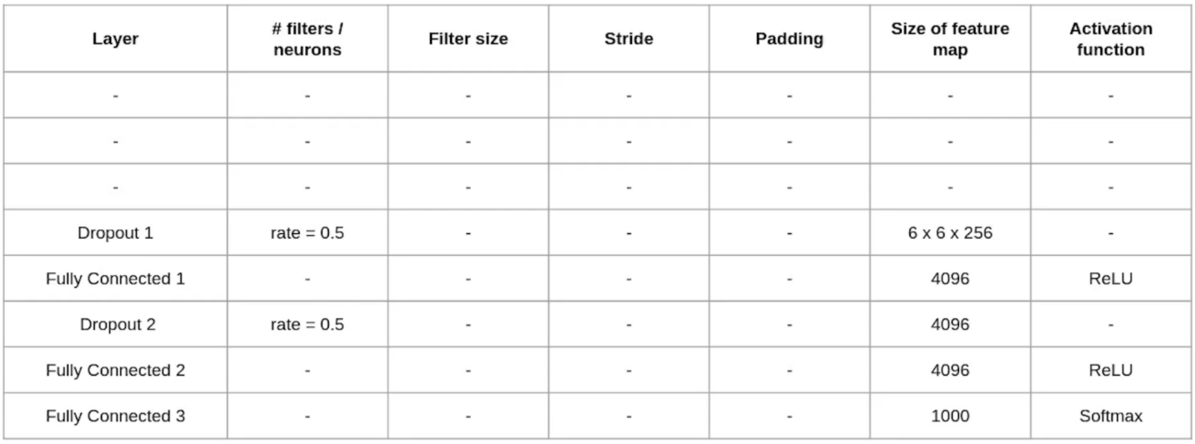
Convolution and Maxpooling Layers

The 1st convolution layer is applied on 11x11 size with 96 filters, stride 4, Relu activation function and output feature map as 55X55X96.

Output= (input size of the filter/ stride)+1

The number of filters is the output feature map.This is followed by 3x3 Maxpooling layer with stride as 2 resulting the feature map as 27X27X96.In second convolution operation we have 256 filter with the filter size reduced to 5x5, stride 1 ,padding 2, relu activation function and output size as 27X27X256. Further 3X3 max pooling layer with stride 2 is applied which resultes in a 13X13X256 feature map.The third convolution is done with 384 3x3 filters, stride 1, padding 1, relu activation function resulting in a 13X13X384 feature map. The fourth convolution has 384 3x3 filters, stride 1, padding 1, relu activation function resulting in again 13X13X384 feature map.The final convolution layer has 256 3x3 filters, stride 1, padding 1, relu activation function resulting in 13X13X256 feature map.Hence the number of filters increases as the network goes deep implying its capacity to extract more features. Also the reducing filter size reduces the feature map size as we go deeper.Further we apply 3X3 maxpooling layers with stride 2 resulting in 6X6X256 feature map.

Fully Connected and Dropout Layers



Layer	# filters / neurons	Filter size	Stride	Padding	Size of feature map	Activation function
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
Dropout 1	rate = 0.5	-	-	-	6 x 6 x 256	-
Fully Connected 1	-	-	-	-	4096	ReLU
Dropout 2	rate = 0.5	-	-	-	4096	-
Fully Connected 2	-	-	-	-	4096	ReLU
Fully Connected 3	-	-	-	-	1000	Softmax

Table 1. Fully Connected and Dropout Layers in AlexNet

Now dropout layers are introduced with dropout rate of 0.5. Further there is a fully connected layer, relu as activation function, output of size 4096 and dropout layer of0.5 fixed dropout rate.A2nd fully connected layer is followed with relu as activation and 4096 neurons.The last layer with Softmax as the activation function we have the output full connected layer with 1000 neurons for 1000 classes.Hence we get the AlexNet architecture with 62.3 million learnable parameters.

The following summarizes the architecture.

- Eight layers along with learnable parameters
- RGB image inputs
- A combination of max-pooling layers for five convolution layers.
- Followed by three fully connected layers.
- Relu as activation function in all layers excluding output
- Two Dropout layers
- Softmax as output layer activation function
- 62.3 million learnable parameters in the architecture.

IV. EXPERIMENTAL RESULTS & DISCUSSION

To generate the results for this research, we have used a dataset of 100 retinal scans. The number of healthy patient scans is 30, and the number of diabetic patient scans is 70. This gives us a total of 100 scans, with a skew towards diabetic scans. This class imbalance will affect model performance, but we will observe later in the document that our model is able to handle this imbalance reasonably well.

Let us take a look at a few sample images from the dataset to get an overview of the type of retinopathy images present.

Images of Healthy patients' scans:

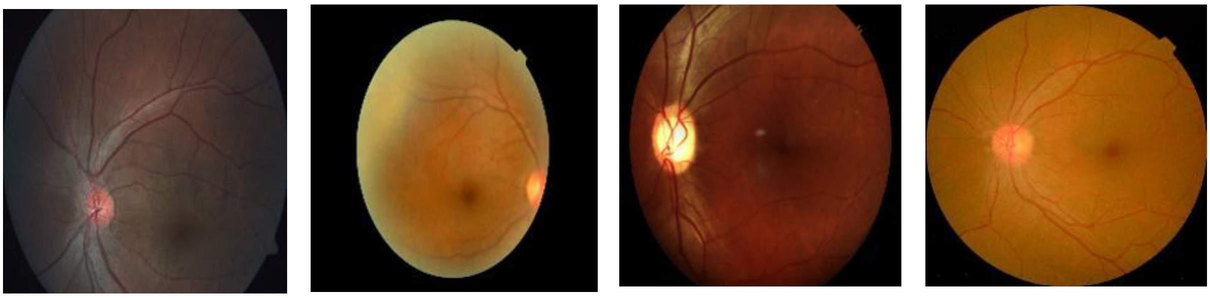


Figure 3. Sample Input Images taken for Experimental Results

As we can see, the images of various types are present. This is important to note because our deep learning model has to account of such variations. This observation is even more significant considering the small dataset size.

Images of Diabetic patients' scans:

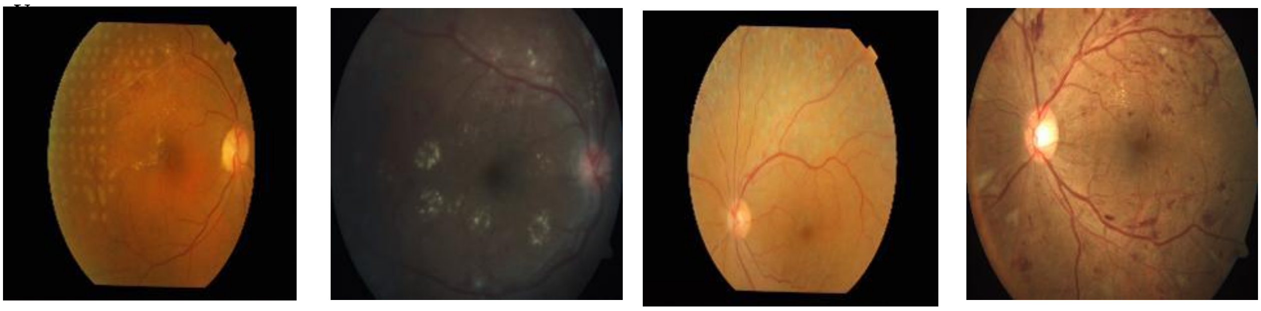


Figure 4. Images Detected with Diabetic Retinopathy after Experimental Results

Again, we can a lot of variations. Apart from the shape of the retina, we can also see the presence of veins and the patterns present in the retina are localized differently.

Architecture:

We have used the AlexNet architecture for this project as it is a simple architecture, but is also able to achieve good results. Below is the output we obtain during the training cycle.

```
Total params: 21,853,985
Trainable params: 51,201
Non-trainable params: 21,802,784
Found 100 images belonging to 2 classes.
Epoch 1/10
4/4 [==============================] - 63s 15s/step - loss: 3.5623 - accuracy: 0.7531 - val_loss:
0.9839 - val_accuracy: 0.6000
Epoch 2/10
4/4 [==============================] - 22s 6s/step - loss: 2.9423 - accuracy: 0.6814 - val_loss:
0.1695 - val_accuracy: 0.5000
Epoch 3/10
4/4 [==============================] - 22s 6s/step - loss: 0.6266 - accuracy: 0.8342 - val_loss:
0.9744 - val_accuracy: 0.7000
Epoch 4/10
4/4 [==============================] - 21s 6s/step - loss: 2.3673 - accuracy: 0.8936 - val_loss:
0.6271 - val_accuracy: 0.8000
Epoch 5/10
4/4 [==============================] - 21s 7s/step - loss: 0.5440 - accuracy: 0.8700 - val_loss:
0.3473 - val_accuracy: 0.9000
Epoch 6/10
```

```
4/4 [==============================] - 22s 6s/step - loss: 0.9990 - accuracy: 0.8179 - val_loss:
0.4064 - val_accuracy: 0.8000
Epoch 7/10
4/4 [==============================] - 22s 6s/step - loss: 0.4390 - accuracy: 0.9943 - val_loss:
0.1176 - val_accuracy: 0.9000
Epoch 8/10
4/4 [==============================] - 21s 7s/step - loss: 0.4440 - accuracy: 0.8125 - val_loss:
0.4843 - val_accuracy: 0.8000
Epoch 9/10
4/4 [==============================] - 21s 6s/step - loss: 0.4811 - accuracy: 0.8958 - val_loss:
0.7737 - val_accuracy: 0.8022
Epoch 10/10
4/4 [==============================] - 21s 6s/step - loss: 0.4450 - accuracy: 0.8035 - val_loss:
0.1627 - val_accuracy: 0.8055
```

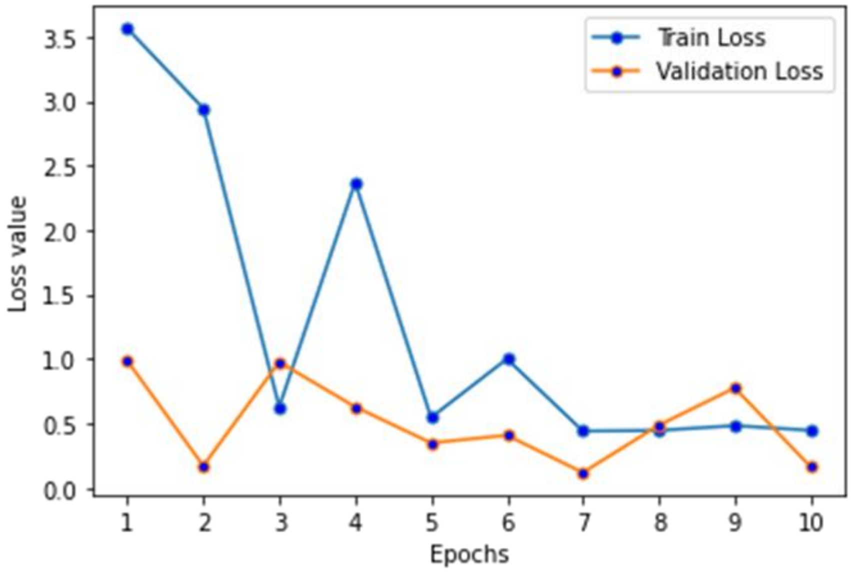


Figure 5. Training Loss vs Validation Loss

The above graph shows the convergence that the train and validation losses achieve as the epochs progress.

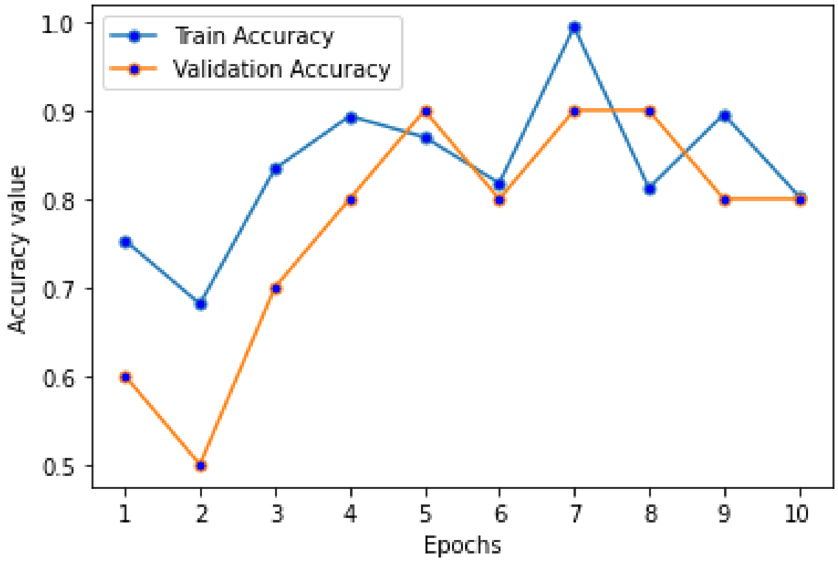


Figure 6. Training Accuracy Vs Validation Accuracy

The above graph shows the training and validation accuracies as the epochs progress. We can particularly see how similar these two accuracies get in the final epoch showing very high generalization.

By taking a look at the network performance as the epochs progressed, we can make two important observations. Though the neural network starts with disparate train and validation losses, it converges towards a train accuracy of 80.35% and validation accuracy of 80.00%. These values are very similar, and this informs us that the model has achieved good generalization. Secondly, we also observe that the validation loss is lesser than the train loss.

The above two observations mean that our validation set is well-chosen and that the model has not overfitted to the train set. Together, these two observations mean that our model is robust even with very little epochs. This is significant considering the previously stated skewness present in our dataset.

Analysis of Results

The confusion matrix and associated metrics presented below allow us to get a better understanding of our model's performance.

	True Positive	True Negative
Predicted Positive	56	14
Predicted Negative	6	24

Table 2. Confusion Matrix for the Sample Dataset

Note: Here, Positive label denotes a diabetic patient, and, Negative label denotes a healthy patient.

Based on the above matrix, we can calculate various metrics, as shown below

Metric	Value	Formula
Accuracy	0. 8055	(TP + TN) / (P + N)
False Discovery Rate	0.2000	FP / (FP + TP)
Precision	0.8000	TP / (TP + FP)
Recall	0.9032	TP / (TP + FN)
F1 Score	0.8485	2TP / (2TP + FP + FN)
Specificity	0.6316	TN / (FP + TN)
False Positive Rate	0.3684	FP / (FP + TN)
False Negative Rate	0.0968	FN / (FN + TP)

Table 3. Parametric Observations from the Predictions using AlexNet Architecture

From the table, we can see that our model's accuracy is 80.00%. Complementary to this is the False Discovery Rate of 20.00%. Together, they effectively mean that there is a 20% chance our model could mislabel an image as the incorrect class. Our precision score tells that 87.00% of those instances we classified as diabetic were actually diabetic. The recall score tells that, of all the people who were diabetic, we correctly predicted 90.32% of them. Using the precision and recall values, the F1 Score of 84.85% we obtain is the average of the precision and recall values.

So far, the metrics we have displayed show that our model is performing well. But using more metrics allows us to get a deeper understanding of the limitations in the performance of the model. The specificity score we get has significance since it tells that of all the healthy people, we correctly labelled only 63.16% of them. Also, the False Positive Rate tells that there is a 36.84% chance that our model could label as healthy patient as having diabetes. This number is not trivial considering that the model is being used in a medical setting. Finally, our False Negative Rate tells that there is 9.68% chance that our model could label a diabetic patient as healthy.

Therefore, the initial few metrics allowed us to get a broad understanding of the model's performance. The last few metrics provide an intricate explanation for the model's behavior. Though we cannot consider this a drawback, since we only trained it on a crude dataset for a very short number of epochs, the metrics discussed in the previous paragraphs allow us to see where the model can be improved.

From these helpful metrics, there are some important points to be noted. Generally, given the small training dataset size of just 100 images and also the fact the we only trained it for 10 epochs, our model has given good results. With a more balanced and larger dataset, paired with a better architecture, we can drastically improve our model performance. Even with our current model, we can observe that the specificity is relatively low. This is due to the small number of images we have of healthy people's retinal scans. We can also see that our False Positive Rate is very high. Therefore, when using this model, a secondary check should be done to actually verify if the person has diabetes. Otherwise, there is good chance a healthy individual would receive treatment for diabetes. Also, the False Negative Rate is not very good. There is an almost 10% chance that our model could miss out on diabetic patients who actually need treatment. Again, we can ascertain these results to the very little number of healthy patients scans we have.

V. CONCLUSION

In this paper, we have developed a very simple computer vision model for diabetic retinopathy classification using the AlexNet architecture. The fact that hundreds of millions of people around the world are at risk of being affected by diabetic retinopathy makes this problem a fertile ground for applying deep learning methods to tackle this. The choice of the model to solve this task is simplistic by choice.We aim to show how even such a simple model can perform well despite small dataset sizes. The dataset used is a simple set of 70 diabetic patient scans and 30 healthy patient scans. As is evident, the dataset is skewed. Regardless, the model performs well and is able to achieve a prediction accuracy of 80%.

In this paper, we have primarily made three contributions to the simple AlexNet model. 1) We have opted for the use of a ReLU activation to speed up training time. 2) We have used multi-GPU training to improve efficiency in training the model. 3) We have used overlapping pooling to reduce the chance of overfitting. Of particular note to us is the fact that the model, within just 10 epochs was able to produce a generalizable solution. We have noted down the results and have made detailed inferences about the generalizability and performance of the model. We have also suggested cautionary measures to using this model at this current stage.

The model developed shows promise, and with better quality data and improved training, it can give far superior performance. We therefore conclude that this proposed solution opens ground for further research work to be conducted in the direction of using simple deep learning models to achieve good generalizable solutions.

REFERENCES

[1] G. G. S, P. N, A. Yaji, A. M, A. M. Dsilva and C. S R, "Review on Text and Speech Conversion Techniques based on Hand Gesture," *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2021, pp. 1682-1689, doi: 10.1109/ICICCS51141.2021.9432277.

[2] J. J. Raval and R. Gajjar, "Real-time Sign Language Recognition using Computer Vision," *2021 3rd International Conference on Signal Processing and Communication (ICPSC)*, 2021, pp. 542-546, doi: 10.1109/ICSPC51351.2021.9451709.

[3] H. S. Anupama, B. A. Usha, S. Madhushankar, V. Vivek and Y. Kulkarni, "Automated Sign Language Interpreter Using Data Gloves," *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, 2021, pp. 472-476, doi: 10.1109/ICAIS50930.2021.9395749.

[4] P. Sonawane, K. Shah, P. Patel, S. Shah and J. Shah, "Speech To Indian Sign Language (ISL) Translation System," *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, 2021, pp. 92-96, doi: 10.1109/ICCCIS51004.2021.9397097.

[5] M. Fora, B. Ben Atitallah, K. Lweesy and O. Kanoun, "Hand Gesture Recognition Based on Force Myography Measurements using KNN Classifier," *2021 18th International Multi-Conference on Systems, Signals & Devices (SSD)*, 2021, pp. 960-964, doi: 10.1109/SSD52085.2021.9429514

[6] D. Tasmere and B. Ahmed, "Hand Gesture Recognition for Bangla Sign Language Using Deep Convolution Neural Network," *2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI)*, 2020, pp. 1-5, doi: 10.1109/STI50764.2020.9350484.

[7] D. Gangadia, V. Chamaria, V. Doshi and J. Gandhi, "Indian Sign Language Interpretation and Sentence Formation," *2020 IEEE Pune Section International Conference (PuneCon)*, 2020, pp. 71-76, doi: 10.1109/PuneCon50868.2020.9362383.

[8] V. Kumar, S. Singh, R. Pal and A. Tiwari, "Ad-Libbed Hand Motion Acknowledgment Framework Utilizing PCA," *2020 International Conference on Advances in Computing, Communication & Materials (ICACCM)*, 2020, pp. 76-81, doi: 10.1109/ICACCM50413.2020.9212940.

[9] K. Bhat and C. L. Chayalakshmi, "Advanced Glove for Deaf and Dumb with Speech and Text Message on Android Cell Phone," *2020 IEEE International Conference for Innovation in Technology (INOCON)*, 2020, pp. 1-7, doi: 10.1109/INOCON50539.2020.9298283.

[10] M. Agrawal, R. Ainapure, S. Agrawal, S. Bhosale and S. Desai, "Models for Hand Gesture Recognition using Deep Learning," *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*, 2020, pp. 589-594, doi: 10.1109/ICCCA49541.2020.9250846.

[11] P. Das, T. Ahmed and M. F. Ali, "Static Hand Gesture Recognition for American Sign Language using Deep Convolutional Neural Network," *2020 IEEE Region 10 Symposium (TENSYMP)*, 2020, pp. 1762-1765, doi: 10.1109/TENSYMP50017.2020.9230772.

[12] R. R. Koli and T. I. Bagban, "Human Action Recognition Using Deep Neural Networks," *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, 2020, pp. 376-380, doi: 10.1109/WorldS450073.2020.9210345.

[13] Y. Suresh, J. Vaishnavi, M. Vindhya, M. S. A. Meeran and S. Vemala, "MUDRAKSHARA - A Voice for Deaf/Dumb People," *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2020, pp. 1-8, doi: 10.1109/ICCCNT49239.2020.9225656.

[14] S. Hossain, D. Sarma, T. Mittra, M. N. Alam, I. Saha and F. T. Johora, "Bengali Hand Sign Gestures Recognition using Convolutional Neural Network," *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2020, pp. 636-641, doi: 10.1109/ICIRCA48905.2020.9183357.

[15] P. K. Datta, A. Biswas, A. Ghosh and N. Chaudhury, "Creation of Image Segmentation Classifiers for Sign Language Processing for Deaf and Dumb," *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 2020, pp. 772-775, doi: 10.1109/ICRITO48877.2020.9197978.

[16] K. Zhao, K. Zhang, Y. Zhai, D. Wang and J. Su, "Real-Time Sign Language Recognition Based on Video Stream," *2020 39th Chinese Control Conference (CCC)*, 2020, pp. 7469-7474, doi: 10.23919/CCC50068.2020.9188508.

[17] K. Nimisha and A. Jacob, "A Brief Review of the Recent Trends in Sign Language Recognition," *2020 International Conference on Communication and Signal Processing (ICCSP)*, 2020, pp. 186-190, doi: 10.1109/ICCSP48568.2020.9182351.

[18] A. Zanzarukiya, B. Jethwa, M. Panchasara and R. Parekh, "Assistive Hand Gesture Glove for Hearing and Speech Impaired," *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*, 2020, pp. 837-841, doi: 10.1109/ICOEI48184.2020.9143031.

[19] G. Jayadeep, N. V. Vishnupriya, V. Venugopal, S. Vishnu and M. Geetha, "Mudra: Convolutional Neural Network based Indian Sign Language Translator for Banks," *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2020, pp. 1228-1232, doi: 10.1109/ICICCS48265.2020.9121144.

[20] D. Hatibaruah, A. K. Talukdar and K. Kumar Sarma, "A Static Hand Gesture Based Sign Language Recognition System using Convolutional Neural Networks," *2020 IEEE 17th India Council International Conference (INDICON)*, 2020, pp. 1-6, doi: 10.1109/INDICON49873.2020.9342405.

[21] S. Kumuda and P. K. Mane, "Smart Assistant for Deaf and Dumb Using Flexible Resistive Sensor : Implemented on LabVIEW Platform," *2020 International Conference on Inventive Computation Technologies (ICICT)*, 2020, pp. 994-1000, doi: 10.1109/ICICT48043.2020.9112553.

[22] D. Gupta, J. P. Mohanty, A. K. Swain and K. Mahapatra, "AutoGstr: Relatively Accurate Sign Language Interpreter," *2019 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS)*, 2019, pp. 322-323, doi: 10.1109/iSES47678.2019.00080.

[23] L. Boppana, R. Ahamed, H. Rane and R. K. Kodali, "Assistive Sign Language Converter for Deaf and Dumb," *2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, 2019, pp. 302-307, doi: 10.1109/iThings/GreenCom/CPSCom/SmartData.2019.00071.

[24] S. Vigneshwaran, M. Shifa Fathima, V. Vijay Sagar and R. SreeArshika, "Hand Gesture Recognition and Voice Conversion System for Dump People," *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, 2019, pp. 762-765, doi: 10.1109/ICACCS.2019.8728538.

[25] P. S.G., J. J., S. R., S. Y., P. G. S. Hiremath and N. T. Pendari, "Dynamic Tool for American Sign Language Finger Spelling Interpreter," *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, 2018, pp. 596-600, doi: 10.1109/ICACCCN.2018.8748859.

[26] S. Chaman, D. D'souza, B. D'mello, K. Bhavsar and J. D'souza, "Real-Time Hand Gesture Communication System in Hindi for Speech and Hearing Impaired," *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2018, pp. 1954-1958, doi: 10.1109/ICCONS.2018.8663015.

[27] H. N. Saha, S. Tapadar, S. Ray, S. K. Chatterjee and S. Saha, "A Machine Learning Based Approach for Hand Gesture Recognition using Distinctive Feature Extraction," *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, 2018, pp. 91-98, doi: 10.1109/CCWC.2018.8301631.

[28] S. Chattoraj, K. Vishwakarma and T. Paul, "Assistive system for physically disabled people using gesture recognition," *2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP)*, 2017, pp. 60-65, doi: 10.1109/SIPROCESS.2017.8124506.

[29] R. A. Elsayed, M. S. Sayed and M. I. Abdalla, "Hand gesture recognition based on dimensionality reduction of histogram of oriented gradients," *2017 Japan-Africa Conference on Electronics, Communications and Computers (JAC-ECC)*, 2017, pp. 119-122, doi: 10.1109/JEC-ECC.2017.8305792.

[30] R. A. Elsayed, M. I. Abdalla and M. S. Sayed, "Hybrid method based on multi-feature descriptor for static sign language recognition," *2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*, 2017, pp. 98-105, doi: 10.1109/INTELCIS.2017.8260039.

[31] Z. A. Memon, M. U. Ahmed, S. T. Hussain, Z. A. Baig and U. Aziz, "Real Time Translator for Sign Languages," *2017 International Conference on Frontiers of Information Technology (FIT)*, 2017, pp. 144-148, doi: 10.1109/FIT.2017.00033.

[32] S. S. Kakkoth and S. Gharge, "Survey on Real Time Hand Gesture Recognition," *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*, 2017, pp. 948-954, doi: 10.1109/CTCEEC.2017.8455041.

[33] U. Patel and A. G. Ambekar, "Moment Based Sign Language Recognition for Indian Languages," *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, 2017, pp. 1-6, doi: 10.1109/ICCUBEA.2017.8463901.

[34] S. Rathi and U. Gawande, "Development of full duplex intelligent communication system for deaf and dumb people," *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence*, 2017, pp. 733-738, doi: 10.1109/CONFLUENCE.2017.7943247.

[35] G. G. Nath and V. S. Anu, "Embedded sign language interpreter system for deaf and dumb people," *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, 2017, pp. 1-5, doi: 10.1109/ICIIECS.2017.8275907