# Automated Detection of Diabetic Retinopathy using Deep Learning

## Shivashish Naithani[1], Shivam Bharadwaj[2], Dhiraj Kumar[3]

*[1,2,3]Student, Dep.t of CS Engineering, ABESIT Ghaziabad, India*

---------------------------------------------------------------------***---------------------------------------------------------------------

**Abstract -** *Approximately four hundred and twenty million people worldwide have been diagnosed with diabetes mellitus. The prevalence of this disease has doubled in the past 30 years and is only expected to increase, particularly in Asia. Of those with diabetes, approximately one-third are expected to be diagnosed with diabetic retinopathy (DR), a chronic eye disease that can progress to irreversible vision loss. Early detection, which is critical for good prognosis, relies on skilled readers and is both labor and time-intensive. This poses a challenge in areas that traditionally lack access to skilled clinical facilities. Moreover, the manual nature of DR screening methods promotes widespread inconsistency among readers. Finally, given an increase in prevalence of both diabetes and associated retinal complications throughout the world, manual methods of diagnosis may be unable to keep apace with demand for screening services. In this paper we introduce an automatic DR grading system capable of classifying images based on disease pathologies from four severity levels. A convolutional neural network (CNN) convolves an input image with a defined weight matrix to extract specific image features without losing spatial arrangement information.*

## 1. INTRODUCTION

Automated techniques for diabetic retinopathy diagnoses are essential to solving these problems. While deep learning for binary classification in general has achieved high validation accuracies, multi-stage classification results are less impressive, particularly for early-stage disease. We initially evaluate different architectures to determine the best performing CNN for the binary classification task and aim to achieve literature reported performance levels. We then seek to train multi-class models that enhance sensitivities for the mild or early stage classes, including various methods of data pre-processing and data augmentation to both improve test accuracy as well as increase our effective dataset sample size.

We address concerns of data fidelity and quality by collating a set of ophthalmologist verified images. Finally, we address the issue of insufficient sample size using a deep layered CNN with transfer learning on discriminant colour space for the recognition task.

We then trained and tested two CNN architectures, AlexNet and GoogLeNet, as 2-ary, 3-ary and 4- ary classification models. They are tuned to perform optimally on a training dataset using several techniques including batch normalization, L2 regularization, dropout, learning rate policies and gradient descent update rules. Experimental studies were conducted using two primary data sources, the publicly available Kaggle dataset of 35,000 retinal images with 5-class labels (normal, mild, moderate, severe, end stage) and a physician-verified Messidor-1 dataset of 1,200 colour fundus images with 4-class labels.

Throughout this study we aim to elucidate a more effective means of classifying early stage diabetic retinopathy for potential clinical benefits.

### 1.1. Background And Related Works

Diagnosis of pathological findings in fundoscopy, a medical technique to visualize the retina, depends on a complex range of features and localizations within the image. The diagnosis is particularly difficult for patients with early stage diabetic retinopathy as this relies on discerning the presence of microaneursyms, small saccular outpouching of capillaries, retinal hemorrhages, ruptured blood vessels— among other features—on the fundoscopic images.

Computer-aided diagnosis of diabetic retinopathy has been explored in the past to reduce the burden on opthamologists and mitigate diagnostic inconsistencies between manual readers16. Automated methods to detect microaneursyms and reliably grade fundoscopic images of diabetic retinopathy patients have been active areas of research in computer vision19. The first artificial neural networks explored the ability to classify patches of normal retina without blood vessels, normal retinas with blood vessels, pathologic retinas with exudates, and pathologic retinas with microaneurysms. The accuracy of being able to detect microaneursyms compared to normal patches of retina was reported at 74%.

Past studies using various high bias, low variance digital image processing techniques have performed well at identifying one specific feature used in the detection of subtle disease such as the use of top-hat algorithm for microaneurysm detection. However, a variety of other features besides microaneurysms are efficacious for disease detection.

## 2. DATA

We used two fundoscope image datasets to train an automated classifier for this study. Rapid prototyping was facilitated by pretrained models obtained from the ImageNet visual object recognition database.  Diabetic retinopathy images were acquired from a Kaggle dataset of 35,000 images with 5-class labels (normal, mild, moderate, severe, end stage) and Messidor-1 dataset of 1,200 color fundus images with 4-class labels (normal, mild, moderate, severe). Both datasets consist of color photographs that vary in height and width between the low hundreds to low thousands.

Compared to Messidor-1, the Kaggle dataset consists of a larger proportion of uninterpretable images due to artifact preponderance, faulty labeling and poor quality.  After training on the larger Kaggle datasets and identifying limitations of the conventional approach to retinal image classification, we performed experiments on higher fidelity datasets with improved image quality and reliable labeling.

In the interest of efficient model building, we progressed to a smaller but more ideal dataset for learning difficult features. The Messidor dataset was supplemented with a Kaggle partition (MildDR) consisting of 550 images that was verified for its efficacy by direct physician interpretation. The dataset contains images from a disparate patient population with extremely varied levels of fundus photography lighting and is labeled in a consistent manner.

The lighting affects pixel intensity values within the images and creates variation unrelated to classification pathology. Our study only uses the retinopathy grade as a reference, a description of which is provided in Table 1 together with the number of images for each category.

To assess the suitability of transfer learning, we considered several frameworks to design, train and deploy a modified version of GoogLeNet as a baseline 2-ary, 3-ary and 4-ary classification model.  Our final model was implemented in Tensorflow and was influenced by results from a deep learning GPU interactive training system (DIGITS) that enabled rapid neural network prototype training, performance monitoring and real-time visualizations. You are provided with a large set of high-resolution retina images taken under a variety of imaging conditions.

## 3. METHODS

In order to assess the strengths and limitations of CNNs, several architectures were trained and tested with particular focus on a 50 layers deep model called ResNet50.  This very efficient network achieves state-of-the-art accuracy using a mixture of low-dimensional embeddings and heterogeneous sized spatial filters.

Increased convolution layers and improved utilization of internal network computing resources allow the network to learn deeper features. For example, the first layer might learn edges while the deepest layer learns to interpret hard exudate, a DR classification feature.

The network contains convolution blocks with activation on the top layer that defines complex functional mappings between inputs and response variable, followed by batch normalization after each convolution layer. As the number of feature maps increase, one batch normalization per block is introduced in succession.
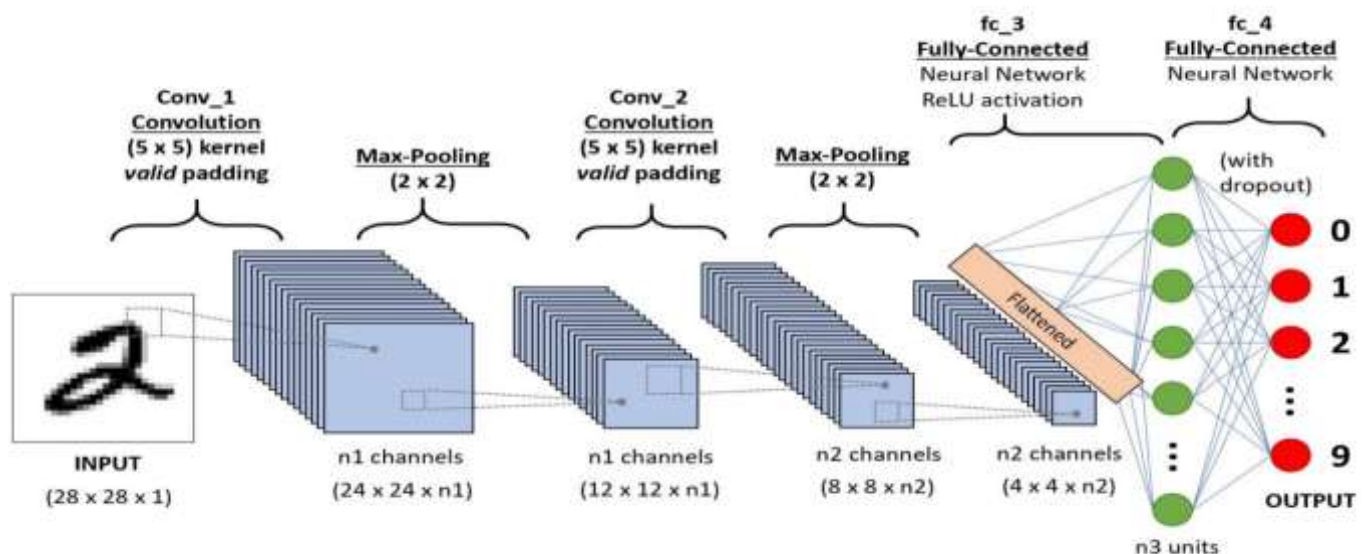


**Fig -1**: A CNN Network

## 4. TRAINING AND TESTING MODELS

A Deep Learning GPU Training System (DIGITS) with prebuilt convolutional neural networks for image classification facilitated data management, model prototyping and real-time performance monitoring. DIGITS is an interactive system and was first used to build a classification dataset by splitting the Messidor and MildDR fundus folder into training and validation subsets of 1077 and 269 images respectively.

The images were cropped to area size 256x256 and used as input data by Imagenet models previously trained for generic classification tasks. The test subset folder contained 400 images from the Lariboisiere hospital Messidor partition and was disjoint from training data. This training system, which offered extensive hyperparameter selections, was then used to build model prototypes over 100 epochs requiring approximately 20 minutes each to complete.

A NVIDIA GTX 1060 GPU hardware device powered the training and a form of early stopping determined the optimal test set model epoch.

## 5. CONCLUSIONS

Automated detection and screening offers a unique opportunity to prevent a significant proportion of vision loss in our population. In recent years, researchers have added CNNs into the set of algorithms used to screen for diabetic disease. CNNs promise to leverage the large amounts of images that have been amassed for physician interpreted screening and learn from raw pixels. The high variance and low bias of these models could allow CNNs to diagnose a wider range of nondiabetic diseases as well.

However, while we achieve state-of-the-art performance with CNNs using binary classifiers, the model performance degrades with increasing number of classes. Though it is tempting to surmise that more data may be better, previous work in the field has corroborated that CNN ability to tolerate scale variations is restricted and others have suggested that in the case of retinal images, more data cannot supplement for this inherent limitation. Gulshan et al. reported a 93-96% recall for binary classification of disease but reports that recall is not improved when training with 60,000 samples vs 120,000 samples of a private dataset.

We then trained and tested two CNN architectures, AlexNet and GoogLeNet, as 2-ary, 3-ary and 4- ary classification models. They are tuned to perform optimally on a training dataset using several techniques including batch normalization, L2 regularization, dropout, learning rate policies and gradient descent update rules. Experimental studies were conducted using two primary data sources, the publicly available Kaggle dataset of 35,000 retinal images with 5-class labels (normal, mild, moderate, severe, end stage) and a physician-verified Messidor-1 dataset of 1,200 color fundus images with 5-class labels.

## ACKNOWLEDGEMENT

## REFERENCES

1) Hinton, G.: A practical guide to training restricted boltzmann machines. Momentum 9(1), 926 (2010)
2) Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.: Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580 (2012)
3) Ji, S., Xu, W., Yang, M., Yu, K.: 3d convolutional neural networks for human action recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on 35(1), 221–231 (2013)
4) Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L.: Largescale video classification with convolutional neural networks. In: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. pp. 1725–1732. IEEE (2014)
5) Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
6) LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Backpropagation applied to handwritten zip code recognition. Neural computation 1(4), 541–551 (1989)
7) LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE 86(11), 2278–2324 (1998)
8) Nebauer, C.: Evaluation of convolutional neural networks for visual recognition. Neural Networks, IEEE Transactions on 9(4), 685–696 (1998)