Detection of Diabetic Retinopathy Using CNN

U. SATISH[1], S K. ABBDDULARAFATH[2], V. RISHINDRA[3], Y. MUKHESH[4], V. SAI HARSHA[5]
[1, 2, 3, 4, 5] *Student, Vasireddy Venkatadri Institute of Technology, A.P, India*

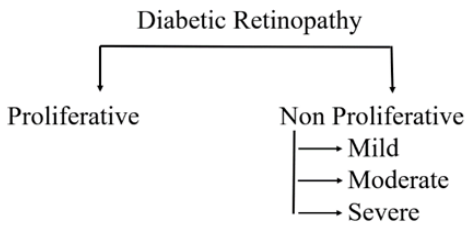
Abstract- ***The diagnosis the diagnosis of diabetic retinopathy (DR) through color fundus images requires experienced clinicians to identify the presence and significance of many small features which, along with a complex grading system, makes this a difficult and time-consuming task. In this paper, we propose a CNN approach to diagnosing DR from digital fundus images and accurately classifying its severity. We develop a network with CNN architecture and data augmentation which can identify the intricate features involved in the classification task such as micro-aneurysms, exudate and hemorrhage on the retina and consequently provide a diagnosis automatically and without user input. We train this network using a high-end graphics processor unit (GPU) on the publicly available Kaggle dataset and demonstrate impressive results, particularly for a high-level classification task. On the data set of 80,000 images used our proposed CNN achieves a sensitivity of 95% and an accuracy of 75% on 5,000 validation images.***

Indexed Terms- ***Deep Learning; Convolution Neural Networks; Diabetic Retinopathy; Image Classification; Diabetes.***

I. INTRODUCTION

A. Diabetic Retinopathy

Diabetic retinopathy (DR), also known as diabetic eye disease, is when damage occurs to the retina due to diabetes. It can eventually lead to blindness. It is an ocular manifestation of diabetes. Despite these intimidating statistics, research indicates that at least 90% of these new cases could be reduced if there were proper and vigilant treatment and monitoring of the eyes.



The longer a person has diabetes, the higher his or her chances of developing diabetic retinopathy. Diabetic retinopathy can be diagnosed into 5 stages: mild, moderate, severe, proliferative or no disease. The various signs and markers of diabetic retinopathy include micro aneurysms, leaking blood vessels, retinal swellings, growth of abnormal new blood vessels and damaged nerve tissues [7]. DR detection is challenging because by the time human readers submit their reviews, often a day or two later, the delayed results lead to lost follow up, miscommunication, and delayed treatment. Clinicians can identify DR by the presence of lesions associated with the vascular abnormalities caused by the disease. While this approach is effective, its resource demands are high. The expertise and equipment required are often lacking in areas where the rate of diabetes in local populations is high and DR detection is most needed. The need for a comprehensive and automated method of DR screening has long been recognized, and previous efforts have made good progress using image classification, pattern recognition, and machine learning [7]. The current research in diagnosing diabetic retinopathy has been based on explicit extraction of features like micro aneurysms and lesions through which the classification is performed. There has also been research in using machine-learning techniques to classify the image as normal or diseased [14, 15, 16]. This paper aims at proposing a diabetic retinopathy diagnosis model that automatically learns features, which are pivotal in diagnosing the stage of the disease without explicit or manual feature extraction.

Fig 1: Normal Vision

B. Convolutional Neural Networks

Convolutional networks have recently enjoyed a great success in large-scale image and video recognition which has become possible due to the large public image repositories, such as ImageNet (Deng et al., 2009), and high-performance computing systems, such as GPUs or large-scale distributed clusters. In particular, an important role in the advance of deep visual recognition architectures has been played by the ImageNet Large-Scale Visual

Fig 2: Vision with diabetic retinopathy

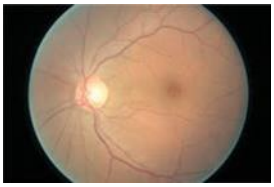
Recognition Challenge (ILSVRC) which has served as a testbed for a few generations of large-scale image classification systems, from high-dimensional shallow feature encodings (Pyronin et al., 2010) (the winner of ILSVRC-2011) to deep ConvNets. (Krizhevsky et al., 2012) (the winner of ILSVRC-2012). The performance of convolutional neural networks in these competitions was the motivation behind adopting CNN for this research. A Convolutional Neural Network (CNN) is comprised of one or more convolutional layers (often with subsampling step) and then followed by one or more fully connected layers as in a standard multilayer neural network. The architecture of a CNN is designed to take advantage of the 2D hierarchical structure of an input image (or other 2D input such as a speech signal). This is achieved with local connections and tied weights followed by some form of pooling which results in translation invariant features. Another benefit of CNNs is that they are easier to train and have many fewer parameters than fully connected networks with the same number of hidden units. CNNs also consider the hierarchical representation of images while training by stacking multiple trainable stages on each other.

II. RELATED WORK

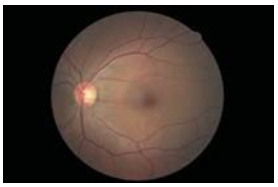
Extensive research has been carried out on methods for a binary classification of DR with encouraging results. Gardner et al used Neural Networks and pixel intensity values to achieve sensitivity and specificity results of 88.4% and 83.5% respectively for yes or no classification of DR 16. They used a small dataset of around 200 images and split each image in to patches and then required a clinician to classify the patches for features before SVM implementation.

Neural Networks have also been used in three-class classification of DR. Nayak et al 17 used features such as the area of exudates and the area of blood vessels together with texture parameters. Features are entered into the neural network to classify images into normal, non-proliferative retinopathy and proliferative retinopathy. The neural net- work used these features as input for classification. The detection results were validated by comparing with grading from expert ophthalmologists. They demonstrated a classification accuracy of 93%, sensitivity of 90% and specificity of 100%. This was carried out on a dataset of 140 images and feature extraction was required on all images in both training and testing which can be time consuming.

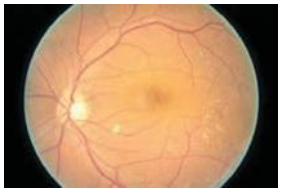
The vast majority of research on the five-class classification that has been carried out has used support vector ma- chines (SVMs). Acharya et al 18 have created an automated method for identifying the five classes. Features, which are extracted from the raw data using a higher-order spectra method, are fed in to the SVM classifier and capture the variation in the shapes and contours in the images.



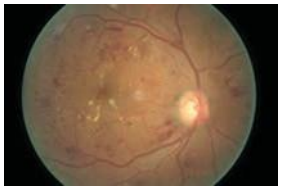
(a) No DR



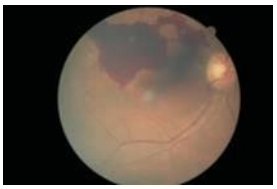
(b) Mild DR



(c) Moderate DR



(d) Severe DR



(e) Proliferative DR

This SVM method reported an average accuracy of 82%, sensitivity of 82% and specificity of 88%. Acharya et al. 19 also created a five-class classification method by calculating the areas of several features such as hemorrhage's, micro-aneurysms, exudate and blood vessel. The features determined to be the most crucial; blood vessels, micro-aneurysms exudates, and hemorrhage's, were extracted from the raw images using image processing techniques. These were then fed to the SVM for classification. A sensitivity of 82 %, specificity of 86% and accuracy of 85.9% was achieved using this system. These methods were performed on relatively small datasets and the drop in sensitivity and specificity was likely due to the complex nature of the five class problem.

Adarsh et al 20 also used image processing techniques to produce an automated diagnosis for DR through the detection of retinal blood vessels, exudate, micro-aneurysms and texture features. The area of lesions and texture features were used to construct the feature vector for the multi- class SVM. This achieved accuracies of 96% and 94.6% on the public 89 and 130 image databases DIARETDB0 and DIARETDB1 respectively.

Each of the previous five class methods required feature extraction from the images before being input to an SVM classifier and have only been validated on small test sets of approximately 100 images. These methods are less real-time applicable than a CNN.

III. THE DATASET, HARDWARE, SOFTWARE

The dataset used for testing was provided by the Kaggle coding website and contains over 80,000 images, of approximately 6M pixels per image and scales of retinopathy. Resizing these I'm ages and running our CNN on a high-end GPU, the NVIDIA K40c, meant we were able to train on the whole dataset. The NVIDIA K40c contains 2880 CUDA cores and comes with the NVIDIA CUDA Deep Neural Network library (cuDNN) for GPU learning. Through using his package around 15,000 images were uploaded on the GPU memory at any one time. The deep learning package Kera's was used with machine learning back end. This was chosen due to good documentation and short calculation time. An image can be classified in 0.04 seconds meaning real-time feedback for the patient is possible.

In this we used two fundoscopic image datasets to train an automated classifier for this study. Diabetic retinopathy images were acquired from a Kaggle dataset of 35,000 images with 5-class labels normal, mild, moderate, severe, end stage. Messidor-1 dataset of 1,200 colour fundus images with 4-class labels normal, mild, moderate, severe. Both datasets consist of colour 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 labelling and poor quality. 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 labelled in a consistent manner.

TABLE 1: Class Distribution in Original Dataset

Class	Name	No. of images	Percentage
0	Mild DR	2443	6.96%
1	Moderate DR	5292	15.07%
2	Severe DR	873	2.48%

IV. DATA PRE-PROCESSING

Pre-processing involved several steps: Converting the input image to RGB model and resizing the image as per CNN Grayscale or Colour Images were converted into RGB model. Images of any size are resized to the specified size of CNN.

Due to non-standard image resolutions, the training images could not be utilized directly for training. The images were scaled down to a fixed resolution size of 512x512 pixels to form a standardized dataset. Training images of resolution 512x512 pixels on all three-color channels demanded high memory requirements. Due to this limitation, the images were converted to a single channel. After several experiments, it was found that green channel images retained information better than the other channel images. In order to enhance the contrast of the image evenly across pixels, histogram equalization technique was applied on the images. In order to prevent the convolutional neural network from learning the inherent background noise in the image, each image was normalized using Min-Max normalization.

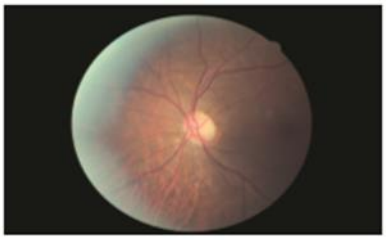


Fig 4: Original Image

V. CNN ARCHITECTURE

The name “convolutional neural network” indicates that the network employs a mathematical operation called convolution. Convolutional networks are simply neural networks that use convolution in place of general matrix multiplication in at least one of their layers. A convolutional neural network consists of an input and an output layer, as well as multiple hidden layers. The hidden layers are Convolutional layers, ReLU layers, Pooling layers, a fully connected layer.

A. Convolutional layer:

Convolutional layer applies a convolution operation to the input, passing the result to the next layer. The convolution emulates the response of an individual neuron to visual stimuli. Each convolutional neuron processes data only for its receptive field.

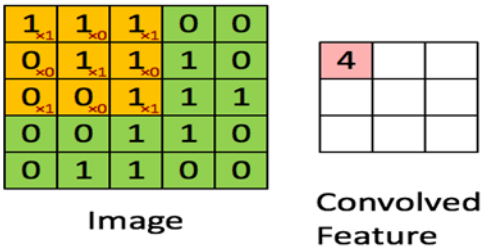


Fig 5: Convolution layer

B. ReLU Layer

ReLU layer applies an activation function onto your feature maps to increase non-linearity in the network. This is because images themselves are highly non-linear! It removes negative values from an activation map by setting them to zero.

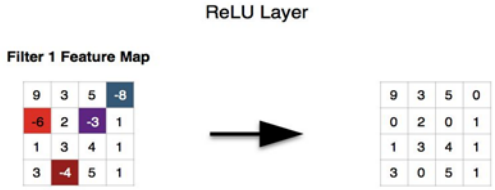


Fig 6: ReLU layer

C. Pooling Layer

Convolutional networks may include local or global pooling layers, which combine the outputs of neuron clusters at one layer into a single neuron in the next layer for minimizing the risk.

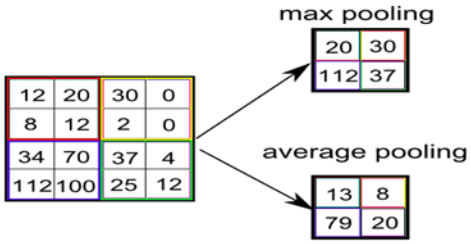
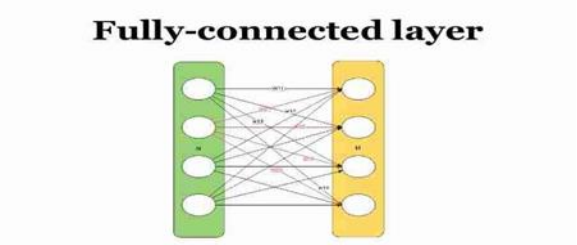


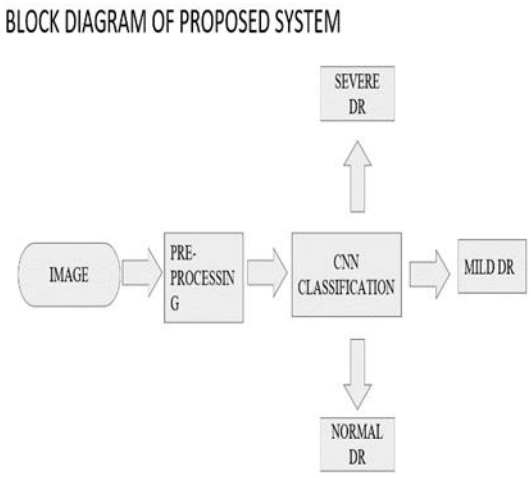
Fig 7: Pooling Layer

D. Fully Connected Layer

Fully connected layers connect every neuron in one layer to every neuron in another layer.



VI. PROPOSED SYSTEM



A. Training:

The CNN was initially pre-trained on 10,290 images until it reached a significant level. This was needed to achieve a relatively quick classification result without wasting substantial training time. After 120 epochs of training on the initial images the network was then trained on the full 78,000 training images for a further 20 epochs. Neural networks suffer from severe over-fitting, especially in a dataset such as ours in which the majority of the images in the dataset are classified in one class, that showing no signs of retinopathy. To solve this issue, we implemented real-time class weights in the network. For every batch loaded for back-propagation, the class-weights were updated with a ratio respective to how many images in the training batch were classified as having no signs of DR. This reduced the risk of over-fitting to a certain class to be greatly reduced.

The network was trained using stochastic gradient descent with Nestrov momentum. A low learning rate of 0.0001 was used for 5 epochs to stabilize the weights. This was then increased to 0.0003 for the substantial 120 epochs of training on the initial 10,290 images, taking the accuracy of the model to over 60%, this took circa 350 hours of training. The network was then trained on the full training set of images with a low learning rate. Within a couple of large epochs of the full dataset the accuracy of the network had increased to over 70%. The learning rate was then lowered by a factor of 10 every time training loss and accuracy saturated.

B. Augmentation:

The original pre-processed images were only used for training the network once. Afterwards, real-time data-augmentation was used throughout training to improve the localization ability of the network. During every epoch each image was randomly augmented with: random rotation 0-90 degrees, random yes or no horizontal and vertical flips and random horizontal and vertical shifts. The result of an image augmentation can be seen in Fig 3 (c).

CONCLUSION

In order to prevent DR for a person who is suffering from diabetics from 5 to 10 years he need to have control on his diet and have frequent eye checkups. So, that DR is identified at early stage.

Our study has shown that the five-class problem for national screening of DR can be approached using a

CNN method. Our network has shown promising signs of being able to learn the features required to classify the fundus images, accurately classifying the majority of proliferative cases and cases with no DR. As in other studies using large datasets high specificity has come with a trade off of lower sensitivity 8. Our method produces comparable results to these previous methods without any feature-specific detection and using a much more general dataset.

The potential benefit of using our trained CNN is that it can classify thousands of images every minute allowing it to be used in real-time whenever a new image is acquired. In practice images are sent to clinicians for grading and not accurately graded when the patient is in for screening. The trained CNN makes a quick diagnosis and instant response to a patient possible. The network also achieved these results with only one image per eye.

The network has no issue learning to detect an image of a healthy eye. This is likely due to the large number of healthy eyes within the dataset. In training the learning required to classify the images at the extreme ends of the scale was significantly less. The issues came in making the network to distinguish between the mild, moderate and severe cases of DR. The low sensitivity, mainly from the mild and moderate classes suggests the network struggled to learn deep enough features to detect some of the more intricate aspects of DR. An associated issue identified, which was certified by a clinician, was that by national UK standards around over 10% of the images in our dataset are deemed ungradable. These images were defined a class on the basis of having at least a certain level of DR. This could have severely hindered our results as the images are misclassified for both training and validation.

In future, we have plans to collect a much cleaner dataset from real UK screening settings. The ongoing developments in CNNs allow much deeper networks which could learn better the intricate features that this network struggled to learn. The results from our network are very promising from an orthodox network topology. Unlike in previous methods, nothing specifically related to the features of our fundus images have been used such as vessels, exudate etc. This makes the CNN results impressive but in future we have ideas to cater our network towards this specific task, in order to learn the more subtle classification features. We will also look to compare these networks to five class SVM methods trained on the same datasets.

To conclude, we have shown that CNNs have the potential to be trained to identify the features of Diabetic Retinopathy in fundus images. CNNs have the potential to be incredibly useful to DR clinicians in the future as the networks and the datasets continue improving and they will offer real-time classifications

RESULT

5,000 images from the dataset were saved for validation purposes. Running the validation images on the network took 188 seconds. For this five class problem we define specificity as the number of patients correctly identified as not having DR out of the true total amount not having DR and sensitivity as the number of patients correctly identified as having DR out of the true total amount with DR. We define accuracy as the amount of patients with a correct classification. The final trained network achieved, 95% specificity, 75% accuracy and 30% sensitivity. The classifications in the network were de- fined numerically as: 0 - No DR and 1 – for other types of DR.

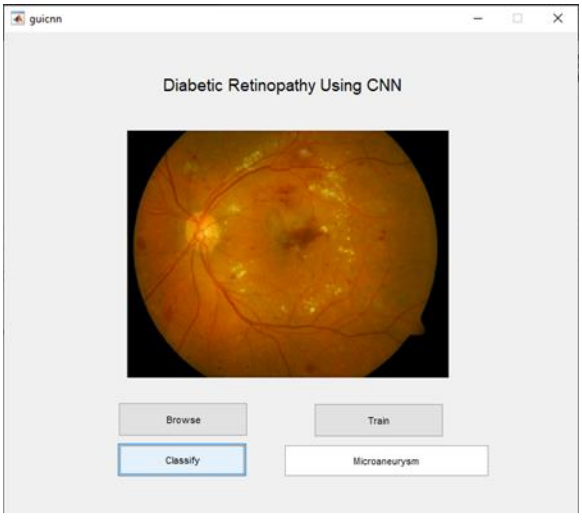


Fig: Moderate DR

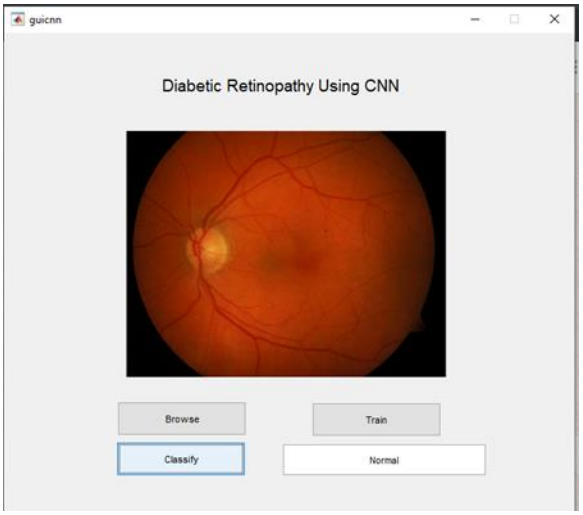


Fig: Severe DR

REFERENCES

[1] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. ImageNet Classi- fication with Deep Convolutional Neural Network. Advances in Neural Information Processing System 25, 2012

[2] Y. LeCun, K. Kavukcuoglu, and C. Farabet. Convolutional networks and applications in vision. In Circuits and Systems (ISCAS), Proceed- ings of 2010 IEEE International Symposium on, pages 253256. IEEE, 2010.

[3] Karen Simonyan, Andrew Zisserman. Very Deep Convolutional Net- works for Large-Scale Image Recognition. ICLR 2015.

[4] D. Nouri(2014, Dec 17). *Using Convolutional Neural Nets To Detect Facial Keypoints Tutorial* [Online]. Available : http://danielnouri.org/notes/2014/12/17/using-convolutional-neural-nets-to-detect-facial-keypoints-tutorial/

[5] *Diabetic retinopathy image dataset* [Online]. Available: https://www.kaggle.com/c/diabetic-retinopathy-detection/data

[6] *Unsupervised Feature Learning and Deep Learning Tutorial by Stan- ford University* [Online]. Available: http://ufldl.stanford.edu/tutorial/

[7] *Diabetic retinopathy* [Online]. Available : https://en.wikipedia.org/wiki/Diabetic retinopathy

[8] *Quadratic Weighted Kappa Metric* [Online]. Avail- able : https://www.kaggle.com/c/diabetic-retinopathy- detection/details/evaluation

[9] Ian J. Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, Yoshua Bengio. Maxout Networks arXiv:1302.4389v4 [stat.ML] 20 Sep 2013

[10] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research 15 (2014).