Automatic detection of Diabetic Retinopathy using R-CNN

Athira T R[1], Athira Sivadas[2], Aleena George[3] , Amala Paul[4] , Neethu Radha Gopan[5]

[1,2,3,4]Student, Dept. of Electronics and Communication Engineering, Rajagiri School of Engineering and Technology, Cochin, Kerala, India

[5]Assistant Professor, Dept. of Electronics and Communication Engineering, Rajagiri School of Engineering and Technology, Cochin, Kerala, India

---***---

Abstract - *Diabetic Retinopathy (DR) is an eye abnormality caused as a result of long term diabetes. As the disease progresses it leads to distortion and blurred vision. The diagnosis of DR using color fundus image requires skilled clinicians to identify the presence of critical features which makes this a difficult and time consuming task. In our paper we propose a R-CNN (Regional Convolutional Neural Network) approach to diagnose DR from digital fundus images. In our research we implemented a new approach were the whole image was segmented and only the regions of interest were taken for further processing. In our method we have used 10 layers for R-CNN, trained it on 130 fundus images and tested on 110 images. All the images were classified into two groups i.e., with DR and without DR. This R-CNN (Regional CNN) approach was found to be efficient in terms of speed and accuracy. An accuracy of approximately 93.8% was obtained from R-CNN.*

Key Words: **Convolutional neural network, Regional convolutional neural network, Diabetic retinopathy**.

1. INTRODUCTION

Diabetes has now become a worldwide disease which ultimately leads to complete vision loss. Diabetic Retinopathy is a complication of type 2 diabetes. According to the IAPB (International Agency for the Prevention of Blindness) [1] report published in 2017, there were 422 million people diagnosed with diabetes. 1 in 3 people diagnosed with diabetes will have diabetic retinopathy up to a certain degree and 1 in 10 people will suffer from vision loss. DR results in the damage of blood vessels in the retinal layer of the eye. It forms microaneurysms due to the focal dilation of weakened walls [2]. The capillaries may become leaky forming yellow white flecks which are commonly referred to as exudates.

The main issue with DR is that it doesn't usually cause sight loss until it has reached the advanced stage. Due to the lack of any significant symptoms normal DR screening will only help the patients with high risk of progression. In order to identify DR we analyze the fundus images for the lesions and exudates. Normal approach for diagnosing DR is time consuming and requires experienced clinicians to identify critical features from the fundus images. An automatic method for detection of diabetic retinopathy would help people with diabetes to recognize the symptoms at its earlier stage. It can greatly reduce the clinical burden on retina specialists. This also helps to monitor the dynamics of the lesions. Countries with huge population like India, China, Indonesia and Bangladesh contributes to 45% of the global burden in diabetes [2]. Since the counts are expected to move up, an automatic clinical detection would be of much help.

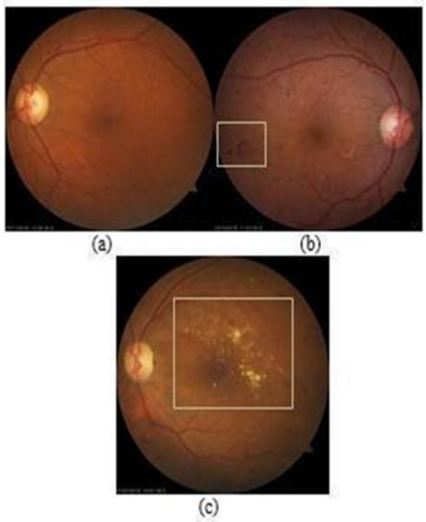

Fig-1. (a) A normal fundus image with no sign of DR, (b)Fundus image with red lesions, (c) Fundus image with exudates.

The study for automatic detection of DR becomes more and more crucial in the past few years.

In our study we are focusing on anomalies in the retina in the form of exudates and red lesions. Due to the similar color characteristics of red lesions with the retinal blood vessels it is hard to locate these lesions using normal image processing techniques. Considerable work has been done in blood vessel extraction and optic disc removal, both of which may also result in a false detection. Techniques of morphological closing and opening can be used for the removal of structures that may result in false detection, but they may still produce residues [3]. SVM classifier can be used to discriminate retinal images with bright lesions to detect DR [4]-[5].

The short comings of the above mentioned methods can be overcome through CNN and R-CNN techniques. CNN comes under the class of deep neural networks whose hidden layers include several convolution layers. CNN has the ability to learn the features during its training phase. At different layers of the network various features with varying complexities can be learned. There is no need for a manual feature extraction since it automatically extracts features while passing through each of the layers.

The hidden layer may include ReLU (Rectified linear unit), which is an activation function, convolutional layer, fully connected layer, softmax layer etc.

2. METHODOLOGY

CNN has been widely recognized for applications such as image processing, pattern recognition and video recognition. CNN in image classification takes an image as the input and classify it into the appropriate category. It has a number of hidden layers in which convolution is done to extract features and other valuable information from the image. The output is obtained from the classification layer [8].

In R-CNN, the image is segmented into many regions (or segments) and the CNN is compelled to focus on these segments. The accuracy of object detection is very high compared to that of CNN due to extraction of region of interest [12].

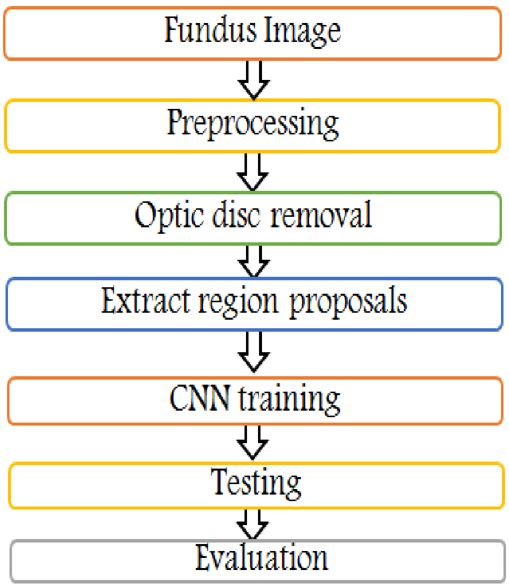

Fig-2. Block level representation of the methodology

2.1. Preprocessing

Initially the original fundus images are resized to a dimension of 336 x 448. Due to the immense information and varying contrast of images obtained from the fundus cameras preprocessing is necessary. Without preprocessing the images suffer from vignetting effects and image distortion [7]. Since the images are obtained from different fundus cameras they will be having non uniform illumination, illumination normalization techniques have to be incorporated [6]- [7].

2.1.1 Green channel extraction

Every image consists of three channels namely red, green and blue. Green channel is extracted from the fundus image. Extraction of green channel provides better contrast between maximum and minimum intensities in an image. It has less noise compared to that of red or blue channels. Blue channel is not normally considered because it provides less contrast and does not contain much information.

2.1.2. Adaptive Histogram Equalization

This image processing technique improves the contrast of the image. It is usually performed on small regions of the image called tiles. This helps in enhancing the edges of the image.

2.2. Optic disc removal

Automatic Optic Disc (OD) removal is a vital step because optic disc act as false positives as they greatly resemble exudates in intensity. Firstly Images are converted into grayscale whose intensity values ranges between 0 to 255 and then thresholding is done. The pixels that have intensities lower than the threshold were converted to 0 (black), while pixels above the threshold were converted to 255 (white). the dilate-erode combinations are used to remove unconnected small regions while preserving large unconnected regions. This provides an initial region from which the segmentation process begins. A disk structuring element of size 30 is applied over the image and an OFF pixel is set to ON if any of the structuring elements overlap ON pixels of the image. The threshold is continuously updated and this process is repeated until only one region remains.

2.3. Extraction of region proposals

Region of interest is extracted using 2 methods by means of thresholding and blocking (Pixel based) [14]. Based on intensity differences the region of interests (i.e., those segments which contain lesions) are extracted. For further processing only these segments are considered.

2.3.1 Threshold based

Regions are extracted from the preprocessed image by thresholding the intensity values of red lesions and exudates separately [13]. For thresholding, lab color space is used since it is the most accurate way of representing colors. It is a 3 axis color space in which L represents lightness, a & b represents color dimensions.

Fig-3. Thresholing

2.3.2 Pixel based

Images are divided into blocks equal segments of size 128 x 128 pixels and the blocks which are suspected to have signs of DR are extracted. Only these extracted blocks are considered for training

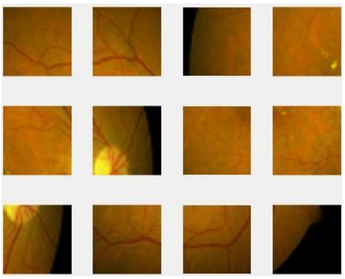

Fig-4. Pixel based blocking

2.4 CNN classification

The number of input layer neurons is equal to the number of pixels in the input image. Convolutional layer makes use of the convolutional features and computes the product between the image patches and the filter. For the activation layer ReLU(Rectified Linear Unit) can be used. ReLU layer perform a threshold operation to each element of the input where any value less than zero is set to zero. Element wise activation is applied to the output of the convolutional layer. The function of the pooling layer is to down convert the volume so as to make the computation faster and to reduce the memory requirement. The convolutional layer that is present just before the fully connected layer holds the information about the features including the edges, contrast, blobs and shapes. And the fully connected layer contains the aggregated information from the previous layers. A softmax layer is used as the final layer of CNN[9]. It assigns decimal probabilities to each class. The output layer classifies the input images into 2 classes namely normal (without DR) and abnormal (with DR). Fully connected layer takes the output of all the neurons from the former layer.

The designed layers are used for training and testing of images [10]-[11]. The images for training can be obtained from IDRiD (Indian Diabetic Retinopathy Image Dataset) which provides marked images of Diabetic lesions.

In our method we have used 10 layers for R-CNN, trained it on 130 fundus images and tested on 110 images. All the images were classified into two groups i.e., with DR and without DR.

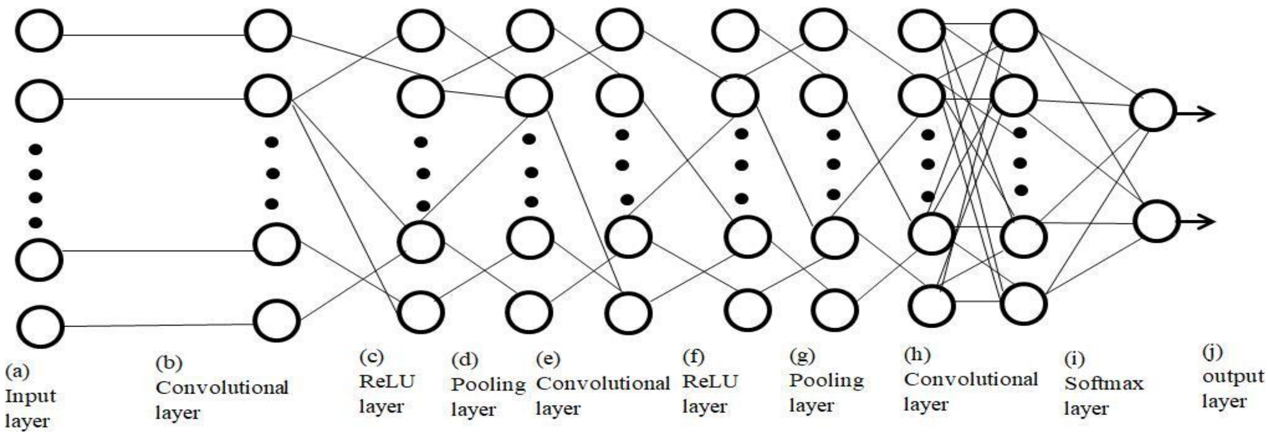

Fig-5. CNN layers

3. PERFORMANCE PARAMETERS

The parameters that are used to evaluate the performance of the proposed model are accuracy, sensitivity, specificity, False positive rate and precision.

If P is the number of positive samples (patient with DR), N is the number of negative samples (patient without DR). True Positive (TP) is the number of patients with DR which is correctly classified. False Positive (FP) is the number of patients without DR which is incorrectly classified. True Negative (TN) is the number of patients without DR which is correctly classified False Negative (FN) is the number of patients with DR which is incorrectly classified.

$$\text{Accuracy} = \frac{TP+TN}{P+N} \quad (1)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (2)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (3)$$

$$\text{False positive rate} = \frac{FP}{N} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

4. EXPERIMENTAL RESULTS

The experiment was done on the images obtained from IDRiD online database along with 50 images collected from the regional hospitals in Kerala to form a total of 240 images. After preprocessing, and candidate region extraction, the images were trained with the CNN network which consisted of 10 layers. The training was done on 130 images for 30 epochs. Testing with the trained network resulted in an accuracy of 93.8 percent..

In this section we evaluate the performance and results R-CNN techniques. The performance is evaluated by the various performance parameters like accuracy, sensitivity and speed.

Input layer takes in a whole color image in CNN and only suspected blocks in R-CNN. While thresholding the intensity of suspected blocks, the range of L component is taken to be greater than 80, a component to be less than -5 and b component to be greater than 70. First convolution layer contains a 7 x 7 kernel with a stride of 2. ReLU acts as the activation function for the previous layer's output.

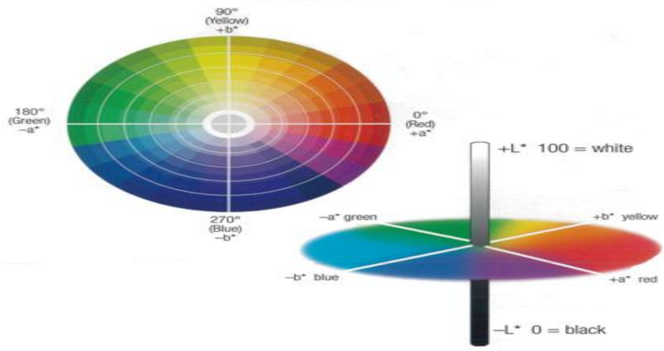

Fig-6. Lab color space

A 2 x 2 max pooling layer is used with stride [2 2] thus reducing the complexity of the computation. A second convolutional layer with 32 filters of size [5 5] is used. It also does zero paddings of size 2 along all the edges of the layer input. The bias learning rate factor is set to 2 which gives the learning rate for the bias in the layer to be twice the current bias global learning rate. A second pooling layer with pool size [3 2] is used with stride 2. Since we aimed at classifying the whole dataset into two classes namely 'with DR' and 'without DR' the fully connected layer is designed for two outputs. The softmax layer applies a softmax function to the input.

Table-1. Description of CNN layers

Layer #	Layer Type	Description
Layer 1	Input image	Input image
Layer 2	Convolution	Filter size= 7 x 7, No. of filters= 30, Stride = 2
Layer 3	RELU	Introduces non-linearity
Layer 4	Max Pooling	Pool size= 2 x 2, Stride = 2
Layer 5	Convolution	Filter size= 5 x 5, No. of filters= 32, Padding = 2 Bias linear rate factor= 2
Layer 6	RELU	Introduces non-linearity
Layer 7	Max Pooling	Pool size= 3 x 2, Stride = 2
Layer 8	Fully connected	2 nodes
Layer 9	Softmax	2 classes
Layer 10	Classification output	Output is the message box indicating the class normal or abnormal

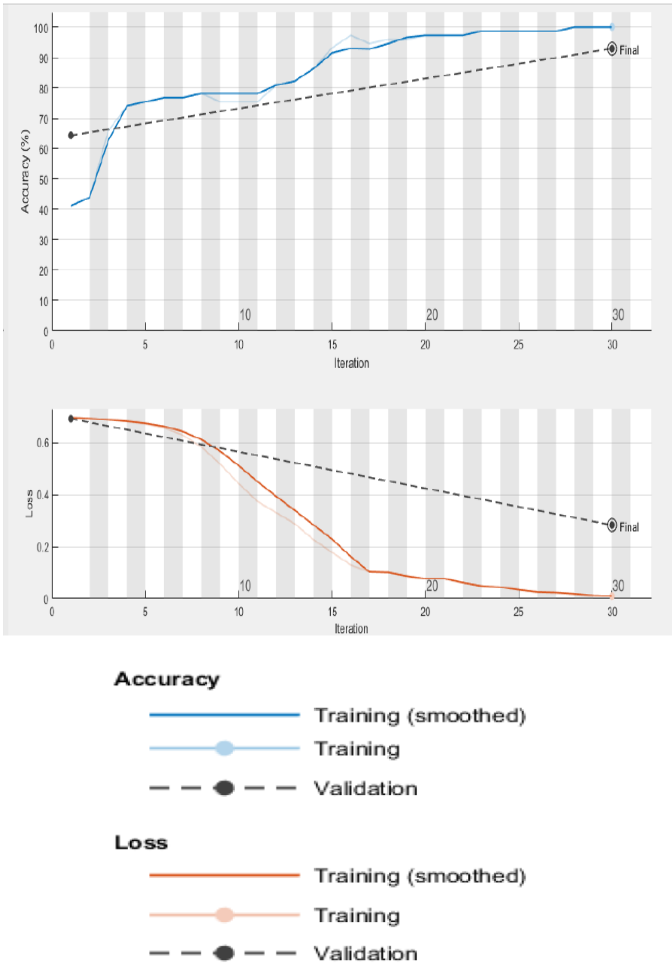

Chart-1. Dynamics of training process

Performance Parameter	R-CNN (%)
Accuracy	95.5
Sensitivity	100
Specificity	93.8
False Positive Rate	6.3
Precision	85.7
Error Rate	4.5

Table-2. Performance evaluation

4. CHALLENGES

The major challenge was about the computational complexity that the laptops that we have with us could handle. The program had to deal with several thousands of neurons. Initially while working with a laptop that was running on an intel Celeron n3050 processor which had a clock speed of 1.6 GHz, it was not possible to train more that 37 images. Matlab (software) crashed every single time we tried to add a new image for training. Then we moved to using a laptop that had a dual core intel i3 processor which had a clock frequency of 3.4GHz. With this we were able to train more images but it took approximately 20 to 25 minutes for training each new image. Finally we moved to the systems systems running on quad core intel i5 processor which took only less than two minutes to train with an image.

5. CONCLUSION

Our project proved that machine learning techniques like neural networks have very high future scope in disease detection from medical images. Researches have already proved the efficiency of R-CNN technique in the field of object detection. With this project it is proved that R-CNN can also the used for detecting very tiny features. R-CNN is found to be highly accurate and sensitive for lesion detection. An accuracy of about 93.8% is obtained for R-CNN. It was due to the manual features that increased the accuracy of R-CNN.

REFERENCES

[1] IAPB Resources “What is avoidable blindness :Diabetic Retinopathy”
[2] Donald S FongMD, MPH Fredrick, L FerrisIIMD, Mathew D DavisMD, Emily Y ChewMD, “Causes of severe visual loss in the early treatment Diabetic Retinopathy Study ETDRS Report No.24,”American journal of Ophthalmology, vol. 127, Issue 2, pp.137–141,February 1999.
[3] Sudeshna Sil Kar and Santi P. Maity, “Automatic Detection of Retinal Lesions for Screening of Diabetic Retinopathy”IEEE Transaction on Biomedical engineering , in press.
[4] Desire Sidibe, Ibrahaim Sadek,Fabrice Meriaudeau, “Discrimination of retinal images containing bright lesions using sparse coded features and SVM,” in Computers in Biology and Medicine 62(2015), pp.175-184.
[5] Muhammad faisal, Djoko Wahono, Iketut Eddy Purnama, Mochammad Hariadi, Mauridhi Hery Purnomo, “Clasification of Diabetic Rrtinopathy Patients using Support Vector Machines based on digital retinal image,”Journal

of Theoretical and Applied Information Technology,vol. 59 No.1,10th january 2014.

[6] Marco Foracchia, Enrico Grisan Alfredo Ruggeri, "Luminosity and contrast normalization in retinal images,"Medical Image Analysis, vol. 9, Issue 3, pp. 179-190, June 2005.

[7] A. Hoover and M. Goldbaum, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels." IEEE transactions on medical imaging, vol. 22, no. 8, pp. 951–8, 2003.

[8] Harry Pratt, Frans Coenen, Deborah M Broadbent, Simon P Harding, Yalin Zheng, "Convolutional Neural Networks for Diabetic Retinopathy,"International Conference On Medical Imaging Understading and Analysis 2016, MIUA 2016, 6-8 July 2016,Loughborough,UK.

[9] Rishab Gargeya, Theodore Leng,"Automated Identification of Diabetic Retinopathy Using Deep Learning,"American Academy Of Ophthalmology, in press.

[10] Arkadiusz Kwasigroch, Bartlomiej Jarzembinski, Michał Grochowski, "Deep CNN based decision support system for detection and assessing the stage of diabetic retinopathy,"2018 International Interdisciplinary PhD Workshop (IIPhDW),May 2018.

[11] Ratul Ghosh, Kuntal Ghosh, Sanjit Maitra, " Automatic Detection and Classification of Diabetic Retinopathy stages using CNN ," 2017 4th International Conference on Signal Processing and Integrated Networks (SPIN), Feb. 2017.

[12] Georgia Gkioxari, Ross Girshick, Jitendra Malik, " Contextual Action Recognition With R*CNN", The IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1080-1088.

[13] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik , " Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 580-587.

[14] Shoji Kido, Yasusi Hirano, Noriaki Hashimoto, "Detection and Classification of Lung Abnormalities by Use of Convolutional Neural Network (CNN) and Regions with CNN Features (R-CNN)," International Workshop on Advanced Image Technology , 2018.