Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Diabetic Retinopathy Disease Detection Using Machine Learning

Anusha Shetty, Charithra Kini T, Deepti Viola Barboza, Ashmitha Kalkur, Siju V Soman

"CSE Department, Shri Madhwa Vadiraja Institute Of Technology And Management Udupi"

Email:anusha.17cs009@sode-edu.in, charitha.17cs016@sode-edu.in, deepti.17cs024@sode-edu.in, ashmitha.17cs010@sode-edu.in, sijuvsoman.cs@sode-edu.in

Abstract:

Diabetic Retinopathy (DR) is a human disease among individuals with diabetes that causes damage to tissue layer of eye and if not detected at earlier stage, will eventually result in complete visual defect. Detection of DR at earlier stage is necessary to avoid visual defect and we ought to distinguish them ahead of schedule for viable treatment. As the disease advances, the sight of an affected person might weaken and lead to Diabetic Retinopathy. Subsequently, two gatherings are present, to be specific, (PDR) Proliferative Diabetic Retinopathy and (NPDR) Non Proliferative Diabetic Retinopathy. Different techniques are used to detect the disease.The measures for the infection caused in the eye could be distinguished by removing the features or highlights of the blood vessels. The highlights like veins, hemorrhages of NPDR picture and exudates of PDR picture are removed from the crude picture utilizing the picture processing procedures. After that the pictures are fed t o classifier forgrouping.

Keywords:

Diabetic Retinopathy,Retina,SVM,RF,Retinal Fundus Image

I. INTRODUCTION

The person related to diabetics mellitus (or diabetics) are influenced by assortment of eye conditions, for example DME or Diabetic Macular edema, DR or Diabetic Retinopathy and glaucoma. These diabetic eye infections can cause visual lack or vision loss. Diabetic Retinopathy is the disease that causes visual harm between working - age adults. In this examination, Diabetic Retinopathy is included and it includes changes to retinal veins that harms or cause the veins hole and loss of vision. It is shown that early assurance and helpful treatment could gainfully stay away from this vision loss. Among the working age grown-ups, the vision misfortune can cause the visual deficiency and debilitation among theindividuals

Diabetic Retinopathy is created by varieties in veins of the retina. Vision misfortune or vision is impeded because of harm in nerve cells. These vessels may spill blood and new vessels may be created as a result of veins harm. Diabetic Retinopathy (DR) may advance in following sorts. NPDR or Non Proliferative Diabetic Retinopathy specifically mellow, moderate and serious PDR or Proliferative Diabetic Retinopathy. In mallow or mild NPDR, micro aneurysms are nothing but swelling in sides of retinal blood vessels that causes leak of blood. In Moderate PDR, we can see the growth of new blood vessels and in severe PDR blocking of some blood vessels takes place.

In Preprocessing step the problems like non clarity images, blurred images, etc are treated. Preprocessing involves increasing or decreasing of the pixel sizes.

. Then the input image is converted into HSI, followed by increasing of the contrast of the image. That is done by stretching out the intensity range of images. Then the image clarity is improved by adjusting the contrast, followed by mapping the maximum intensity to 1 and the minimum intensity to 0. Later the salt and pepper noise in the image is removed.

Optic plate end: In this prepare the optic plate is veiled and evacuated from picture. Blood vessels extraction and expulsion: Blood vessels are identified and evacuated by utilizing Edge discovery strategies. Discovery of exudates and micro-aneurysms: Exudates and micro-aneurysms are identified by morphological operations.

SVM classification is utilized to classify the pictures based on seriousness. The pictures are classified as NPDR and PDR.

Our project aims in predicting the severity of diabetic retinopathy in patients. Support vector machine ,an machine learning algorithm can be used for training by pro- viding dataset so that it can predict the severity of the disease by comparing it with the trained data set. Our main objective here is to train our model by providing proper data set in order to detect the level of severity in patients.

Following are the steps of methodology for diabetic retinopathy disease detection-

1) Collection of retinal images and storing it in a particular file format(.jpg).
2) Resizing all the images to a particular size.
3) Input the image from retinal dataset.
4) Preprocessing techniques are applied on input image to remove noise and to enhance the quality of theimages.
5) After pre processing, optic disc and blood vessels are removed from the pre processed images.
6) Morphological operation are performed on the images to Identify the exudates and microaneurysms .
7) Finally SVM & RF classifier is used to classify theimages Based on the severity of the disease.

II .LITERATURE REVIEW

S Guptha and Karandikar in their paper "Diagnosis of Diabetic retinopathy" have found out micro-aneurysms and exudates of retina from the fundus image. The next step performed was processing the image, here morphological operations were done to detect the features which were used for classification. Their accuracy ranged between 87 percent to 100 percent with average accuracy being 86 percentage [1].

"Machine learning approach to automatic exudates detection in retinal images from diabetic patients" is a paper by Akara S, Mathew N Dailey. They introduced a series of investigations on highlight determination and grouping the exudates utilizing SVM(Support Vector Machine) and naive Bayes model. The Bayes model was fed with features or highlights taken from the dataset containing images which had equal amounts of diabetic and non diabetic images. Highlights were extracted until the performance of the model stopped showing improvements. For SVM, the highlight set of Bayes model was fed for extraction of the features [2].

The calculations or algorithms for the mechanized identification of DR that is Diabetic Retinopathy utilizing advanced fundus pictures was presented by Rajendra Acharya U, C Y K Ng, Kwan Hoong Ng and Jasjit S Suri. The algorithm utilized for extraction of certain highlights were improved for obtaining better accuracy. They reviewed the algorithms from the digital fundus pictures and the efficiency of different DR systems was discussed [3].

Deep learning approach for Diabetic Retinopathy identification was proposed by Varun G and Lily P. CNN or deep Convolutional Neural Network was trained with dataset containing infected pictures of the retina. The neural network classified the pictures based on severity to worst and moderate. The neural network was evaluated at two sets, one for sensitivity and one for specificity [4].

In "Machine learning on the Diabetic Retinopathy Dirbeseean Dataset" research paper, Tiagro T.G has utilized R language for anticipating diabetic retinopathy. He had utilized a dataset in which features where removed from the patient's retinal images.His work has 78 percent of accuracy[5].

Winder, R. John, et AL study based on calculations for programmed discovery of retinopathy whereas considering advanced color retinal pictures. The calculations considered for think about were categorized into 5 stages (preprocessing, localization and division of the retinal vasculature, division of the optic disk,, localization of the macula and fovea localization and division of retinopathy) [6].

III. RESEARCH METHODOLOGIES

The investigate strategy indicates how the inquire about is proposed to upgrade to realize the investigate objective. As a result, the research plans are set to gather data from different sources that are accessible to conclude how the past inquires about conducted when they assessed the look engines.

A. Data Collection

In our project, we have used a set of 100 images for the evaluation of performance of our system. All the images are stored in a particular file format that is JPG. An Image is taken as input from the dataset and the image is resized to a particular size. "clearborder" function is used to remove the names and dates that are unwanted. "bwareaopen" function is used to remove the spots,pixels that are unwanted based on the area. "floodfill" method is used to fill the pixels to complete image given as dataset. After that pre-processing techniques are applied, followed by segmentaion and feature extraction. Finally SVM classifier is used for classifying the images based on the level of severity of the disease.

IV. EXPERIMENT

A. Segmentation

Segmentation process is the process where blood vessel is detected from retinal fundus image. Detected blood vessel will be segmented for further classification. The segmentation method used are optical disk elimination, exudates detection, micro-aneurysms detection.

Sequalize is the pre-processed image that we consider here for the optical disk elimination and feature extraction initially we split the image into three channels that is into red, green and blue channels. Then we find the mean and standard deviation for each channel so that we can find the threshold value for the channels. Before detection of the optical disk we increase the contrast of the image using guassian filter and histogram equalization. We convert the image into array/matrix format.

B. Optic disc elimination

The optical disk in retinal image is the bright yellowish region in the colour and it is the normal feature of an eye it is oval or elliptical in shape.

In fundus image optical disk appears to be bright white region . We need to remove optical disk because exudates has similar colour and bright pattern and similar intensity like optical disk. So the optical disk should be masked and removed from the retinal image. After pre-processing we perform edge detection algorithm for detection of the optical disk and blood vessels.Once the image is converted to array format we use the for loop to remove optical disk and optical cups, where we check each value of array with the threshold value if the array value is greater than the threshold value then we assign the value to the highest value that is 255 else if the array value is lesser than the threshold value then we assign to the lowest value that is 0 that is how we find the boundary and remove the optical disk.

C. Exudates and microaneurysms detection

Exudate is the fluid that leaks from the blood vessels and spreads to nearby tissues. An aneurysm is a swelling of a blood vessel in eyes, A retinal microaneurysm is a tiny area in the back of eye where blood protrude from an artery or vein. After removal of optical disk and blood vessels from the retinal image we need to detect the exudates and micro-aneurysms features of retinal image.

After removal of optical disk feature extraction of image is done, now segmented image is considered as test image and we need to extract features that is contrast, homogeneity, correlation, energy, dissimilarity and ASM(average square mean) after we extract all the features we store it in single array.

D. Feature Extraction

Feature Extraction is a process of collecting images for indexing and retrival. Gray Level Co-Occurance Matrix (GLCM) functions are used to find the texture of the image by finding how repeatedly pairs of pixel occur with given value for an image. Echoview provides GLCM texture operator, which gives a virtual variable which represents a texture calculations.The features which GLCM considers are:

1. **Energy :**Vitality returns the complete of squared components within the GLCM. Essentialness is 1 for consistent picture.This property is something else called consistency,which returns a esteem that will discover the closeness of the conveyance of components within the GLCM to the GLCM inclining. At the point when picture isn't texturally uniform, various GLCM components have outstandingly small qualities,which propose that entropy is greatly colossal. Hence,entropy is on the other hand relative to GLCM energy.

2. **Contrast :** Contrast returns a measure of intensity contrast between it and its neighbour over the entire picture. Contrast is 0 for a consistent picture.The property contrast is otherwise called difference and inactivity. It returns a proportion of how associated a pixel is, to its neighbour over the entire picture.

3. **Entrophy :** Entropy qualifies the disorderness of a picture and it fulfills its greatest esteem when all the components in P lattice are rise to.

4. **Homogeneity :** Homogeneity normalizes the GLCM such that the sum of its feature is 1. Each element is assigned to 1.

5. **Correlation :** Correlation is 1 or -1 for a splendidly positive or a negative picture. It returns the whole of squared components in the GLCM. It returns a esteem that measures the closeness of the conveyance of components within the GLCM to the GLCM corner to corner.

E. Classification

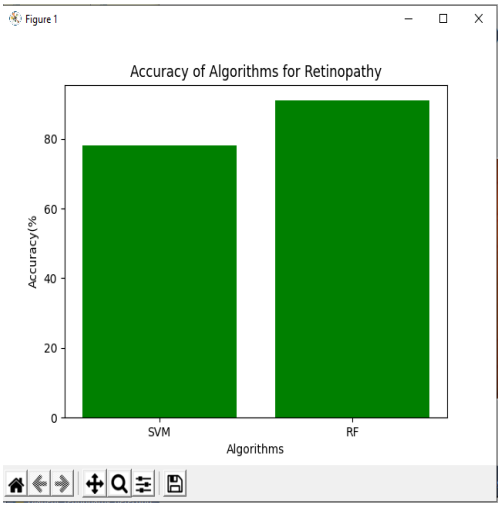
SVM classifier is one of classification algorithm used to solve the problems which involves two groups that has to be separated. The machine is first trained and it is then used for testing. N-dimensional hyper planes are then constructed where N number of features are extracted to train the machine and then that is used to divide the testing data into two -groups. This is how hyper plane will classify the data based on the training. The data above the hyper planes will belong to one group and data below the hyperplane belongs to another group.

RF builds multiple decision trees based on the given dataset and then merges them together to get more accurate prediction. We can random records. For example, we have picked 90 images for training and the decision tree is built on these records. We can choose the number of trees we want. Each tree within the timberland predicts to which category the modern record has a place. At that point based on lion's share voting the modern record is alloted to the given category.

III. ANALYSIS

In our project, we have used the SVM and RF classifier to classify the retinal images into either affected from diabetes or to normal, based on the extracted features. The features are extracted from image and provided as input to the machine and then those features are used to train the machine. The testing is done to classify the image. We compare the accuracy obtained by both SVM and RF.

The input image will be converted into black and white image or we say gray image then we get the threshold image followed by applying morphological operations and finally we get the segmented colored output and accuracy with respect to SVM and RF classification algorithm. For non-diabetic retinal image we get a graph showing accuracy with respect to SVM and RF.



VI .CONCLUSION

The Project has effectively developed the GUI(Graphical User Interface).It is able to include a test picture for processing, extract highlights for the picture, show these qualities and classify the images based on the level of severity. It gives the comprehensive perspective of Retinal state to the Doctors. It extracts the GLCM features from the images. It detects micro aneurysms and exudates even from a low quality shading fundus picture. It uses SVM and RF classification for better accuracy and classifies the images based on the seriousness of the disease. It can also be used for mass screening of images

REFERENCES

[1] S. Gupta and Karandikar, “Diagnosis Of Diabetic Retinopathy using Machine Learning ”, journal of research and development, 2015

[2] Akara S, Buntarit U., Sarah B., Tom W., Khine T.(2009) “ Machine Learning approach to automatic exudate detection in Retina images from diabetic patients” Volume 57-issue 2.

[3] Rajendra Acharya U.,E.Y.K.Ng, Kwan-Hoong Ng, Jasrit S. Suri(2012) “Algorithm for the automated detection of diabetic retinopathy using digital fundas images” volume 36, Issue 1, pp 145-157

[4] Trong-Varun G., Lily P., Mark-C., “Developmentand validation of a deep learning Algorithm for detection of Diabetic Retinopathy”, December 2016

[5] Laura Tiago T.G “Machine Learning on the diabetic Retinopathy Debrecen Dataset”, knowledge based system60, 20-27. Published on June 25,2016

[6] Winder, R.John, “Algorithms for digital image processing in diabetic retinopathy”, Computerized Imaging and Graphics,vol 33, no 8 (2009)

