Diabetic retinopathy using Machine Learning Algorithm

[1]Prof S.V Chichmalatpure ,[2] Snehal Mate ,[3] Rahul Kumar ,[4] Kushal Sharma ,[5]Yash Jaiswal

[1]Prof S.V Chichmalatpure, Professor , Information Technology , Sinhgad Institute Of Technology & Science, Maharashtra , Pune .

[2] Snehal Mate, Student , Information Technology , Sinhgad Institute Of Technology & Science, Maharashtra , Pune .

[3] Rahul Kumar, Student , Information Technology , Sinhgad Institute Of Technology & Science, Maharashtra , Pune

[4]Kushal Sharma , Student , Information Technology , Sinhgad Institute Of Technology & Science, Maharashtra , Pune .

[5]Yash Jaiswal , Student , Information Technology , Sinhgad Institute Of Technology & Science, Maharashtra , Pune .

ABSTRACT

Diabetes Retinopathy is human eye affection.It can affect to retina of eye and causes blindness. Diabetes is supposed the one of the most deadliest disease nowadays. Most of the work in this field is based on disease detection or manual extraction of features, but this paper proposed automatic analysis of this disease into different stages using machine learning. This paper presents the preprocessing technique to remove the noise reduction and hence classify high resolution image into 3 stages based on severity by using (SVM,Navie Bayes and LR) algorithm This results clearly show that the advanced technique outperforms over the usable techniques in terms of sensitivity, accuracy and error rate.There are so many datasets are available publicly such as kaggle ,stare ,Drive.

Keywords*— Diabetic retinopathy, Predictive, genetic algorithm,Machine learning,SVM,LR,Gaussian filter.*

---

1. INTRODUCTION

Diabetic retinopathy or diabetic eye disease is caused by diabetes mellitus which manifests itself in the eye retina. Diabetic eye disease is one of the most frequent causes of complete blindness in many developed countries. The detection of retinal pathologies became much easier using automated retinal image analysis whereas other methods like dilation of eye pupil is time consuming and patient has to suffer for some time. Diabetic retinopathy occurs when high blood glucose damages the small vessels that provides nutrients and oxygen to the retina. There are two types of major diabetes. First (called juvenile-onset or insulin-dependent) diabetes, In this the body completely stops producing insulin, a hormone that enables the body to use glucose found in foods for energy. Second (called adult-onset or non-insulin-dependent). This types of diabetes usually occurs in people who are over age of 40, overweight, and have a family history of diabetes. Detection of diabetes through retinal image has made the diagnoses easier. Which when compared with the other retinas helps examining the presence of diabetes. SVM,LR and Naïve Bayes algorithms are used so that the analysis of the stage of diabetes is done. Main purpose of the system is to detect diabetes. Image processing is done after that if an image is found defected then it is detected as diabetic image. Stages of diabetes is determined. The proposed system detects stages or levels of diabetes and give respective prescription for the same by using different classification techniques. Retinal images are used for the detection of diabetes stage which are compared with the other samples of diabetes and if found severe then immediate diagnoses is done and prescription is been provided for the same.

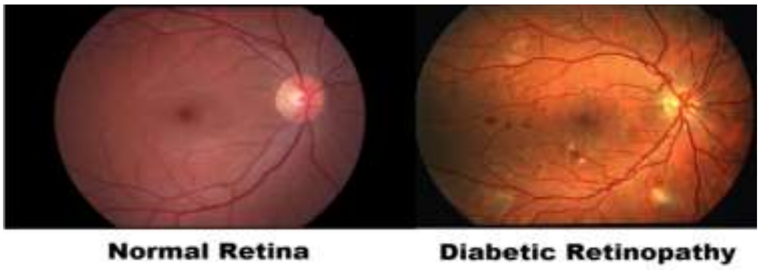

FIG.1

2. RELATED WORK

- [1]This review paper focuses on decision about the presence of disease by applying ensemble of machine learning algorithms on features extracted from output of different retinal image processing algorithms, like diameter of optic disk,lesion specific (microaneurysms, exudates), image level (prescreening,AM/FM, quality assessment). By using alternating decision tree, adaBoost, Naïve Bayes, Random Forest and SVM ,Decision making for predicting the presence of diabetic retinopathy was performed.[2] In this paper, the proposed technique firstly applied switching median filter to remove the effect of high density noise in retinal images and then genetic algorithm will come in action to locate exudates in these images.This experimental results have clearly shown that the proposed technique outperforms over the available techniques in terms of sensitivity, accuracy and error rate. [4]This paper presents the design and implementation of GPU accelerated deep convolutional neural networks. This automatically diagnose and thereby classify high-resolution retinal images into 5 stages of the disease based on severity.

3. PROPOSED METHOD

- All of existing methods are good in some measures for detection and segmentation of exudates but still raise some problems with low intensity, low accuracy, less color contrast and sensitivity, non-uniform illumination images.Therefore, our proposed algorithm and techniques have ability to solve these problems by preprocessing techniques.All process is demonstrated in Fig.2 and step-wise details are explained below.

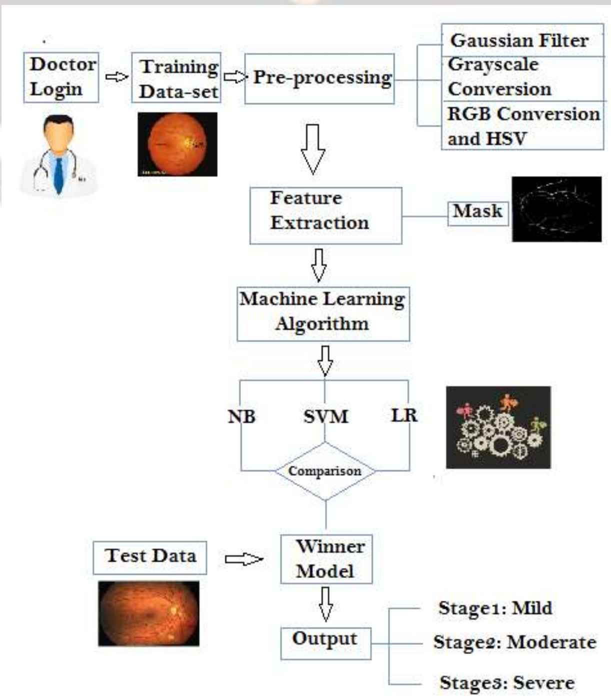

FIG 2. System Design

- Step 1: First of all, take a diabetic RGB human retinal image
- Step 2: Apply Gaussian Filter to remove noisefrom image.
- Step 3: Convert the RGB image into greyscale level
- Step 4: Apply Machine Learning Algorithm to this image.
- Step 5: Compare the Resulted image with Test data.
- Step 6: Detect the stages.
- Step 7:Finally higher values of accuracy, sensitivity and lower value of error rate are obtained.

I. PRE PROCESSING

For the detection of Diabetic Retinopathy stages the Color Fundus Images are considered as an input. These images are the color images which provides the details about retina of eye. These images are preprocessed to improve the quality of image and then it is used for the further stages.The pixel values of Color Fundus Images are permanently distorted and the superior data is used for analysis of images.This suppress undesired information and enhance required features.In Pre processing it involves brightness correction, edge detection, intensity adjustment,Histogram equalization etc.

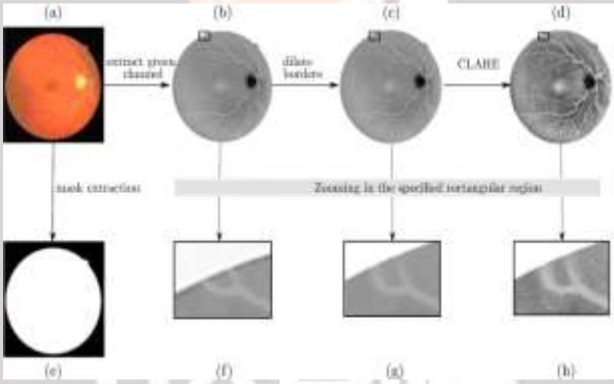

FIG.3 Feature Extraction

Gaussian filter – In this we apply method to make image smooth and reduce noise.Gaussian filter is use to blur images and remove noise.

Grey Conversion- When converting an RGB image to grayscale, we have to take the RGB values for each pixel and make as output as single value reflecting the brightness of that pixel.The purpose of doing this is to highlight the defected portion of the eye.

II. NAÏVE BAYES

Naive Bayes is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. A Naive Bayesian model is easy to build, with no muddled iterative parameter estimation which makes it particularly useful for very large datasets. Bayesian classifiers are the measurable classifiers. Bayesian classifiers can foresee class membership probabilities such as the probability that a given tuple belongs to a particular class.Bayes theorem provides a way of calculating posterior probability P(c|x) from P(c), P(x) and P(x|c). Look at the equation below:

- *P*(*c*|*x*) is the posterior probability of *class* (c, *target*) given *predictor* (x, *attributes*).
- *P*(*c*) is the earlier likelihood of *class*.
- *P*(*x*|*c*) is the likelihood which is the probability of *predictor* given *class*.
- P(x) is the earlier likelihood ofpredictor.

III. LR

Logistic regression is named for the function used at the core of the method, the logistic function.The logistic function, also called the sigmoid function was developed by statisticians to describe properties of population growth in ecology, rising rapidly and maxing out at the carrying capacity of the environment. It's an S shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits.

1 / (1 + e^-value)

Where e is the base of the natural logarithms (Euler's number or the EXP() function in your spreadsheet) and value is the actual numerical value that you want to transform. Following is a plot of the numbers between -5 and 5 transformed into the range 0 and 1 using the logistic function given below.

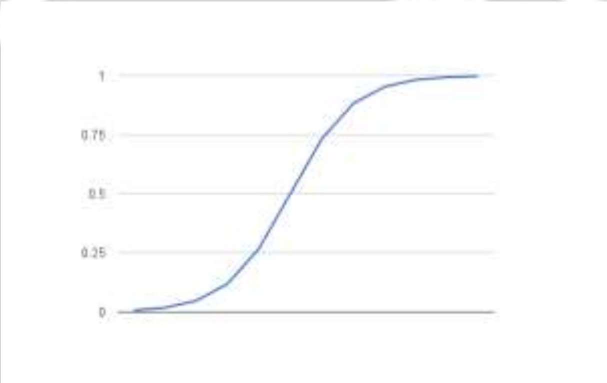

FIG.4 Logistic Function

REPRESENTATION USED FOR LOGISTIC REGRESSION

Logistic regression help an equation as the representation, very much like linear regression.Input values (x) are joined linearly using weights or coefficient values (referred to as the Greek capital letter Beta) to predict an output value (y). A key variance from linear regression is that the output value being modeled is a binary values (0 or 1) rather than a numeric value.

Below is an example logistic regression equation:

y = e^(b0 + b1*x) / (1 + e^(b0 + b1*x))

Where y is the predicted output, b0 is the bias and b1 is the coefficient for the single input value (x). Every column in your input data has an associated b coefficient (a constant real value) that must be learned from your training data.

The real representation of the model that you would store in memory or in a file are the coefficients in the equation (the beta value or b's).

IV. PERFORMANCE METRICS

To analyze the performance of the techniques, twoapproaches were used. First one is the pixel-based and second is the image-based approach. All pixels encount to a candidate that partially or totally overlaps a manually segmented shiny lesion were known as True Positive (TP).All candidate pixels outside this approach were considered as False Positives (FP). All exudate pixels that were not segmented by this method treated as False Negatives(FN).Researchers evaluate their performance in terms of sensitivity, accuracy, specificity, Error Rate and PPV. The difficulty that comes in evaluating the specificity is that if all image pixels are encountered, the number of true Negative (TN) pixels will large as compared to FP values.The performance can be evaluated and analysed on the basis of various parameters used in this research work. Various parameters are:

Sensitivity: Sensitivity is also known as true positive rate that measures the proportion of positives that are correctly identified. Higher the value of sensitivity, better results can be obtained. It can be calculated as

Sensitivity = TP/(TP+FN)

Error Rate: It is defined as the rate at which errors occur in a transmission system. Lower the value of Error Rate better will be the result. It can be calculated as

Error Rate = 1- Accuracy

Accuracy: Accuracy can be defined as the degree of closeness of measurements of a quantity to that quantity's true value. Higher the value of accuracy better will be the results. It can be calculated as

Accuracy = TP+ TN / (TP+FP+FN+ TN)

Specificity: Specificity is also known as true negative rate which measures the proportion of negatives that are correctly identified. Higher the value of specificity, better results can be obtained. It can be calculated as

Specificity = TN/(TN + FP)

processing it involves brightness correction, edge detection, intensity adjustment,Histogram equalization etc.

V. CONCLUSION

The system is designed to predict stages of blindness a diabetic patient can go through i.e. depending upon the severity of blindness. According to the stages of retina harmed medicine are suggested, And depending upon the prescription other diseases can also be predicted.

REFERENCES

[1] Diagnosis of diabetic retinopathy using machine learning classification algorithm. http://ieeexplore.ieee.org/document/7877439/

[2] An Integrated Approach for Diabetic Retinopathy Exudate Segmentation by Using Genetic Algorithm and Switching Median Filter. http://ieeexplore.ieee.org/document/7571284/

[3] Diagnosis of diabetic retinopathy using morphological process and SVM classifier . http://ieeexplore.ieee.org/document/6577181/

[4] Retinopathy Detection using Deep ConvolutionalNeuralNetworks . http://ieeexplore.ieee.org/document/7914977/

[5] http://ieeexplore.ieee.org/document/435390Segmentation of candidate dark lesions in fundus images based on local thresholding and pixeldensity http://ieeexplore.ieee.org/document/4353907/7/

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

[7] Diabetic retinopathy detectionusing deep convolutional neural networks".http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7914977.