_____

Automatic Detection of Exudate in Diabetic Retinopathy Using K-Clustering Algorithm

Dileep J[a], Manohar P[b]

[a] PG student, Department of EIE, RNSIT, Bengaluru, *dileepj999@gmail.com*
[b] Assistant Professor, Department of EIE, RNSIT, Bengaluru, *manoharasp@gmail.com*

Abstract— Diabetic Retinopathy is an eye disease where in which veins may swell and release liquid or new irregular veins develop on retina and piece the light touchy part, this will prompt vision misfortune. It is one of the primary drivers of visual impairment on the planet. Variety in retinal vein thickness, discharge of Exudates which is a protein spillage in the retina, Hemorrhages are a portion of the side effects of Diabetic Retinopathy. Shading fundus pictures will be utilized by ophthalmologists to study eye infections like diabetic retinopathy. Since Optic Disk shows up as a splendid spot in the retinal picture, which takes after exudates, it must be expelled from the picture. Subsequently recognition of Optic Disk is a vital parameter in retinal investigation. On the other hand, in our nation individuals experiencing this disease are all the more in number and therefore oblige more number of ophthalmologists and gigantic time to dissect and analyze the illness. In India, there are insufficient assets, regarding time and accessible master ophthalmologists.

In this paper, a programmed and proficient strategy to distinguish Optic Disk and exudates are proposed. The retinal pictures are preprocessed utilizing the method of LAB shading space picture. The preprocessed shading retinal pictures are portioned utilizing Fuzzy C Means grouping method keeping in mind the end goal to distinguish Optic Disk furthermore division is done utilizing Line Operator procedure. Among the over two techniques, best one is recognized. The exudates are removed utilizing K means bunching and finally the grouping is done utilizing SVM. With the characterization accomplished, the Exudates and Non Exudates pictures are separated.

Index Terms – Diabetic Retinopathy, Exudates, Optic disk, Fuzzy C Means clustering, K means clustering.

_____***** _____

I. INTRODUCTION

Diabetic retinopathy refers to damage to the retina caused by abnormal blood flow related to diabetes mellitus, which can potentially lead to severe loss of vision. It is an ocular manifestation of diabetes, a systemic disease, which affects up to 80 percent of all patients who have had diabetes for 10 years or more. The International Diabetes Federation(IDF)'s Diabetes Atlas reports that India has the highest number of people with diabetes in the world, and hence considered to the "Diabetes Capital of the World".
Currently, 40.9 million Indians are estimated to be suffering from diabetes. By 2025, this number will rocket to 69.9 million, and potentially 85 million by 2030. Therefore regular screening is the most efficient way of reducing the vision loss. Automatic optic disc (OD) detection from retinal images is a very important task in computer-aided diagnosis of various types of eye diseases. It is often a key step for the detection of other anatomical retinal structures, such as retinal blood vessels and macula. More importantly, it helps to establish a retinal coordinate system that can be used to determine the position of other retinal abnormalities, such as exudates, drusen, and hemorrhages. Exudates are the primary sign of diabetic retinopathy. Exudates literally means that any fluid that filters circulatory system into lesions or areas of inflammation. Exudates are the bright lesions found in the presence of retinal and systemic disease. Exudates appear as bright patterns in retinal images. They are manifested as random whitish/yellowish patches of varying sizes, shapes and location. Exudates are made up of serum lipo proteins. It takes much time for ophthalmologists to screen these exudates and hence automatic methods have been developed for screening purposes.

II. DETECTION OF OPTIC DISK

FIG.1: Block Diagram to detect Optic Disk

The retinal images are taken from the available DIARETDB1 and DIARETDB0 databases. This image is initially preprocessed in order to smoothening, correction or normalization, image registration, geometric correction and masking. Then the preprocessed images are subjected to feature extraction. Energy, Dissimilarity, Homogeneity, Mean, Standard Deviation is the features which has to be extracted in order to have the proper detection of optic disk in a retinal image. As the circular brightness has to be detected, the lightness component of a retinal image is first extracted and used within the LAB color space, where the OD detection usually performs the best. The retinal image is then smoothed to enhance the circular brightness structure associated with the OD. A bilateral smoothing filter that

2878

_____

combines geometric closeness and photometric similarity is used. Optic Disc detection is done using Line operator and Fuzzy C means Clustering. The best method is detected.
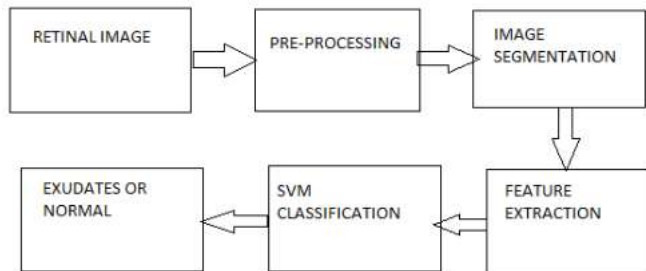
III. DETECTION OF EXUDATES

FIG.2: Block Diagram to detect Exudate

Color fundus images usually comprised with variation in lighting, poor contrast and noise. In order to reduce these imperfection and generate images which are more suitable for extracting the pixel features in the classification process, a preprocessing method comprising with the following step is applied.1)RGB to HSI conversion 2)Median Filtering 3)Contrast Limited Adaptive Histogram Equalization (CLAHE).

RGB to HSI Conversion

The RGB retinal image is converted to HSI component by normalizing the ranges of R, G and B components to the interval from 0 to 1 and then the Intensity is calculated as I=1/3*(R+G+B) with the interval from 0 to 1. The reason why we use HSI color space is because it is easier to extract each color from H value. To be independent of the intensity variance, we use the HS space. This will also help in making the processing and the computing faster.

Median Filtering

The noise reduction is a desirable process that has to be done in image processing. A Median filter is a non linear digital filtering method to remove noise. It is often used because at certain conditions it preserves edges while removing noise.

Contrast Limited Adaptive Histogram Equalization (CLAHE)

CLAHE is used to enhance the low contrast images. The CLAHE algorithm partitions the images into contextual regions and applies the histogram equalization to each one. This evens out the distribution of used grey values and thus makes hidden features of the image more visible. It overcomes the limitations of standard histogram equalization.

A. Image Segmentation based on K-Means

For applications like image recognition or compression, we cannot process the whole image directly for the reason that it is inefficient and unpractical. Image segmentation is to classify or cluster an image into several parts (regions) according to the feature of image. Segmentation partitions an image into distinct regions containing each pixels with similar attributes. In this approach, we have used color features of an image for the purpose of segmentation. The work is divided into two stages: First, enhancing the color separation is done by extracting the a*b*components from the L*a*b* color space of the preprocessed image. Then, the regions are grouped into a set of five clusters using K-means Clustering algorithm. By this two step process, we reduce the computational cost avoiding feature calculation for every pixel in the image.

The process can be summarized with the following steps:

Step 1: Read the retinal image.

Step 2: Convert the image from RGB color space to L*a*b* color space. L*a*b* color space helps us to classify the color differences. L*a*b* color space consists of a Luminosity layer L *, consists of white to black, chromaticity layer a* indicating where the color falls along the red-green axis, chromaticity layer b* indicating where the color falls along the blue-yellow axis. All of the color information is in the a* and b* layer. The difference between two colors can be measured using the Euclidean distance[7].

Step 3: Segment the colors in a*b* space using K- means clustering. Clustering is a way to separate groups of objects. The main idea is to define k centroids, one for each cluster. These centroids should be placed in such a way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early group age is done. At this point it is necessary to re-calculate k new centroids as bar centers of the clusters resulting from the previous step. After obtaining these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop, one may notice that the k centroids change their location step by step until no more changes are done. We need to specify the number of clusters to be partitioned and a distance metric to quantify how close two objects are to each others. Use K-means to cluster the objects into five clusters using the Euclidean distance metric[7].

Step 4: For every objects in the input, K-means returns an index corresponding to a cluster. Label every pixel in the image with its cluster index.

Step 5: Create images that segment the images by color.

Step 6: Since the Optic Disc and Exudates are homogenous in their color property, cluster possessing Optic Disc is localized for further processing.

B.   Feature Extraction

Feature extraction techniques are applied to get features that will be useful in classifying and recognition of images. A large set of data can be described accurately with the application of feature extraction technique by simplifying the amount of resources required. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which over fits the training sample and generalizes poorly to new samples.

An input retinal image after segmentation, has to be classified into exudates or non-exudates based on number of features which are based on color and texture, extracted using Gray Level Co-occurrence Matrix (GLCM). GLCM is a tabulation of how often
different combination of pixel brightness values occur in a pixel pair in an image. Each element (a, b) in GLCM specifies the number of times that the pixel with value 'a' occurred horizontally adjacent to a pixel with value 'b'. The resulting matrix was analyzed and based on the existing information, the feature vectors are formed.

Energy   : $E = \sum_{a,b}((P_{a,b})^2)$
                [where a,b = 0 to N-1]
Dissimilarity   : $D = \sum_{a,b} P_{a,b}|a-b|$
                [where a,b = 0 to N-1]
Homogeneity   : $H = \sum_{a,b} P_{a,b} / (1+(a-b)^2)$
                [where a,b = 0 to N-1]
Standard Deviation : $SD = \sum_{a,b} P_{a,b}((a-\mu a)^2)$
                [where a,b = 0 to N-1]

C.   Classification using Support Vector Machine(SVM)

Support Vector Machines (SVM) recently became one of the most popular classification methods. They have been used in a wide variety of applications such as text classification, facial expression recognition, gene analysis and many others. Support Vector Machines can be thought of as a method for constructing a special kind of rule, called a linear classifier, in a way that produces classifiers with theoretical guarantees of good predictive performance.

There are some properties that make SVMs noble:

l. SVMs develop a most extreme edge separator-a choice limit with the biggest conceivable separation to case focuses. This helps them sum up well [7].

2. SVMs make a direct isolating hyper plane, however they can implant the information into a higher dimensional space, utilizing the purported part trap. Utilizing this high dimensional space, the information can be made straightly divisible which is unthinkable in the first data space. The high-dimensional straight separator is really nonlinear in the first space. It demonstrates that the SVM is useful for the strategies utilizing straight representations [7].

3. SVMs are a nonparametric technique they hold the estimations of prepared samples and store all of them. Then again, by and by they frequently wind up holding just a little portion of the quantity of cases some of the time as few as a little steady time the quantity of measurements. Hence SVMs consolidate the upsides of nonparametric and parametric models: they have the adaptability to speak to complex capacities, however they are impervious to over fitting. The info focuses are mapped to a high dimensional component space, where an isolating hyper-plane can be found. The calculation is picked so as to amplify the separation from the nearest examples, an amount which is known as the edge. SVMs are learning frameworks intended to consequently exchange off exactness and many-sided quality by minimizing an upper bound on the speculation mistake. In an assortment of characterization issues, SVMs have demonstrated an execution which can decrease preparing and testing blunders, in this way acquiring higher acknowledgment precision. SVMs can be connected to high dimensional information without changing their plan [7].

IV. RESULTS AND DISCUSSION

Initially the retinal image is made on to pre-processing method. Then the preprocessed retinal image is converted to L*a*b Color space. The color component of the image is extracted. It is then applied to the K-means Clustering algorithm, which results in five clusters. Since Optic Disc and Exudates are homogenous in color, cluster containing Optic Disc is selected for feature extraction. Based on the feature extracted using GLCM, the SVM is trained for normal and abnormal images. Finally, image is classified as exudates or non-exudates using SVM.

TABLE 1: FEATURE EXTRACTION OF NORMAL AND ABNORMAL IMAGES

Parameters	Normal Image Values	Abnormal Image Values
Auto correlation	2.010e+01	1.146e+01
Contrast	1.009e-01	1.316e-01
Correlation	9.748e-01	9.612e-01
Cluster Prominence	2.492e+02	1.019e+02
Cluster Shade	-2.991e+01	-3.457e+00
Dissimilarity	8.253e-02	1.279e-01
Entropy	1.535e+00	2.019e+00
Homogeneity	9.611e-01	9.366e-01
Maximum Probability	5.714e-01	2.784e-01
Sum Variance	5.757e+01	2.556e+01
Sum Entropy	1.473e+00	1.924e+00
Difference Variance	1.009e-01	1.316e-01
Difference Entropy	2.899e-01	3.885e-01
Inverse Difference	9.910e-01	9.858e-01

TABLE 2: FEATURE EXTRACTION FROM DIARETDB0 IMAGES

Image	Mean	SD	Energy	Dissimilarity
Sample 1	0.90e+02	6.732e+01	3.792e-01	9.763e-01
Sample 2	0.746e+02	5.332e+01	3.092e-01	8.943e-01
Sample 3	0.88e+02	5.700e+01	2.992e-01	9.003e-01
Sample 4	1. 06e+02	6.123e+01	3.292e-01	8.733e-01
Sample 5	0.846e+02	5.901e+01	3.520e-01	8.363e-01

To Detect Exudate in an Abnormal Image

1. Input Image   2.HSI Image

3.Filtered Image   4.Enhanced Image

5. Cluster 1   6.Cluster 2

7. Cluster 3   8. Cluster 4

9. Cluster 5   10 .Selected Cluster
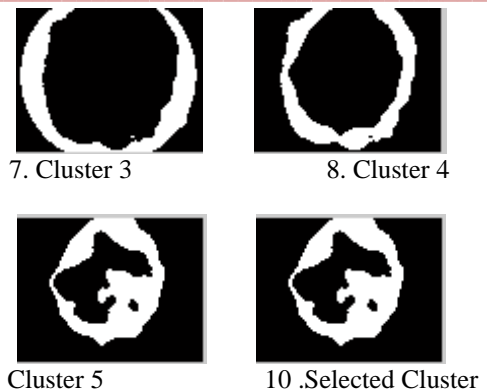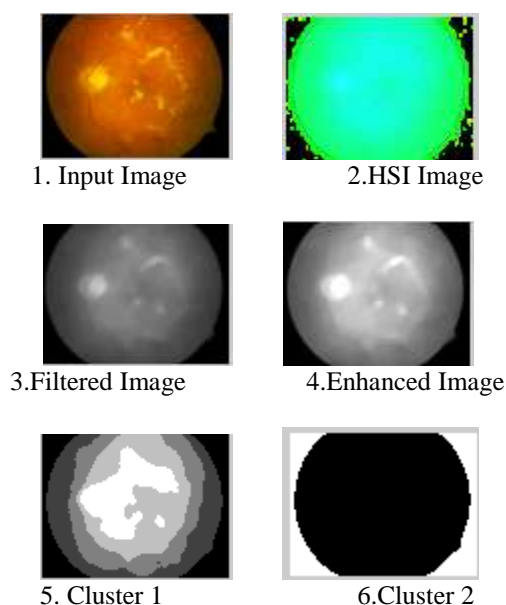
FIG.3: ABNORMAL IMAGE OUTPUTS

To detect Optic Disc

FIG.4: INPUT RETINAL IMAGE

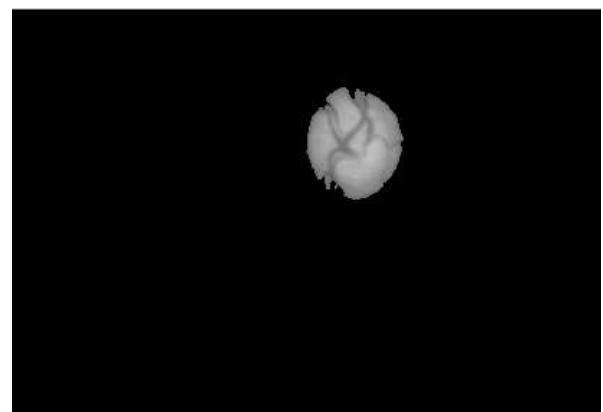**FIG.5**: OD DETECTED USING FCM CLUSTERING

FIG.6: OD DETECTED USING LINE OPERATOR

The above output shows that the Optic disc is accurately detected using Fuzzy C means Clustering compared with Line operator.

V. CONCLUSION

According to the topic of this paper, the Optic Disk has been distinguished utilizing the Line Operator and Fuzzy c-means Clustering calculation from the fundus picture inputs. Vitality, disparity, homogeneity, mean and standard deviation are a percentage of the components of retinal pictures which have been extracted. The Optic Disk is recognized with an exactness of 90% in Fuzzy C means Clustering Algorithm compared with Line operator.

Exudates are only lipo protein spillages in diabetic retinopathy. Complexity Limited Adaptive Histogram Equalization (CLAHE) is utilized to improve the low difference computerized fundus picture. The Contrast improved shading picture is divided utilizing K-means grouping, which is one of the least difficult unsupervised learning calculations for picture division. The diabetic retinopathy pictures were gathered from publically accessible DIARETDB0 and DIARETDB1 sites. Pictures are processed at quicker rate in K-means bunching contrasted with FCM. It gives more shading data from which the consequence of grouping will be progressed. To Classify these sectioned picture into exudates and non-Exudates, an arrangement of components in light of composition and shading are removed utilizing Gray Level Co-Occurrence Matrix (GLCM). The chosen elements are grouped into exudates and non-exudates utilizing Support Vector Machine (SVM) Classifiers. Utilizing this approach, the exudates are distinguished with 90% achievement rate. In future, work is done to get more achievement rate.

REFERENCES

[1] Detection of Lesions in Retinal Images" IEEE Transactions on medical imaging, vol. 30, no. 2,pp. 523-533, February 2011.

[2] Alireza Osareh, Bita Shadgar, and Richard Markham "A Computational-Intelligence-Based Approach for Detection of Exudates in Diabetic Retinopathy Images" IEEE Transactions on information Technology in Biomedicine, vol. 13, no. 4, pp. 535-545, July 2009.

[3] Carla Agurto, Eduardo Barriga, Sergio Murillo, Marios Patti chis, Herbert Davis, Stephen Russell, Michael Abrrmoff, and Peter Soliz "Multiscale AM-FM Methods for Diabetic Retinopathy Lesion Detection". iEEE Trans. Medical imaging Vol. 29, No. 2, pp. 502-512, February 2010.

[4] T. Walter, J. Klein, P. Massin, and A. Erginary,. "A contribution of image processing to the diagnosis of diabetic retinopathy, detection of exudates in colour fundus images of the human retina". iEEE Trans. Medical. imaging, Vol. 21, No. 10, pp.1236-1243, October. 2002.

[5] Akara Sopharak, Bunyarit Uyyanonvara and Sarah Barman "Automatic Exudate Detection from Non-dilated Diabetic Retinopathy Retinal Images Using Fuzzy C-means Clustering". Journal of sensors/2009. ISSN 1424-8220. www.mdpi.com/ journal/sensors.

[6] Niemeijer, B.V Ginnekan S.R, Russell. M and M.D. Abramoff "Automated detection and differentiation of drusen, exudates and Cotton wool spots in digital color fundus photographs for diabetic retinopathy diagnosis ", Invest. Ophthalmol Vis. Sci., Vol. 48, pp. 2260-2267, 2007.

[7] G.S. Annie Grace Vimala, S. Kaja Mohideen "Automatic Detection of Optic Disk and Exudate from Retinal Images Using Clustering Algorithm" 978-1-4673-4603-0/12/$31.00 ©20 12 IEEE.