A diabetic retinopathy detection method using an improved pillar K-means algorithm

Susmitha valli Gogula[1]*, CH Divakar[2], CH Satyanarayana[3] & Allam Appa Rao[4]

[1]Department of IT, GITAM University, Patancheru, Medak Dist-502329, India; [2]Pydah College of Engineering And Technology, Gambheeram, Anandapuram, Visakhapatnam-531163, India; [3]Jawaharlal Nehru Technological University Kakinada, Kakinada, Andhra Pradesh-533003, India; [4]C R Rao Advanced Institute Of Mathematics Statistics And Computer Science, University Of Hyderabad Campus, Gachibowli, Hyderabad-500046, India; Susmitha valli Gogula - Email: susmitagv@gmail.com; *Corresponding author

Abstract:
The paper presents a new approach for medical image segmentation. Exudates are a visible sign of diabetic retinopathy that is the major reason of vision loss in patients with diabetes. If the exudates extend into the macular area, blindness may occur. Automated detection of exudates will assist ophthalmologists in early diagnosis. This segmentation process includes a new mechanism for clustering the elements of high-resolution images in order to improve precision and reduce computation time. The system applies K-means clustering to the image segmentation after getting optimized by Pillar algorithm; pillars are constructed in such a way that they can withstand the pressure. Improved pillar algorithm can optimize the K-means clustering for image segmentation in aspects of precision and computation time. This evaluates the proposed approach for image segmentation by comparing with K-means and Fuzzy C-means in a medical image. Using this method, identification of dark spot in the retina becomes easier and the proposed algorithm is applied on diabetic retinal images of all stages to identify hard and soft exudates, where the existing pillar K-means is more appropriate for brain MRI images. This proposed system help the doctors to identify the problem in the early stage and can suggest a better drug for preventing further retinal damage.

Keywords: Diabetic Retinopathy, K-Means, Fuzzy C-means, Pillar k-Means, Dark Spots, Hard exudates, Soft exudates.

Background:
Medical image segmentation plays an important role in clinical diagnosis .An ideal scheme should posses some properties such as minimum user interaction, fast computing, accurate and robust segmentation result [1]. Segmentation achieves its best results with semi automated algorithms with human operator. Image segmentation often allows doctors and surgeons to analyze the patient data prior to determining the location of disease [2]. Computer vision the area is characterized by the extraction of information from image data for the purpose of a medical diagnosis to a patient. Image data is in the form of ophthalmology images. An example of information which can be extracted from such image data is detection of tumors, hardexudats, soft exudates, hemorrhages or other malign changes. Diagnosis needs segmentation of the image into pixels the goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier and simple to analyze the result of image segmentation is a set of pixels that collectively cover the entire image. Diabetic retinopathy is damage to the retina caused by long time diabetes, which leads to blindness it is an ocular manifestation of diabetes, a disease, which affects up to 80 percent of all patients who have had disease for 10 years or more [3] application of pillar algorithm is observed over 150 images from different databases for checking with the better results.
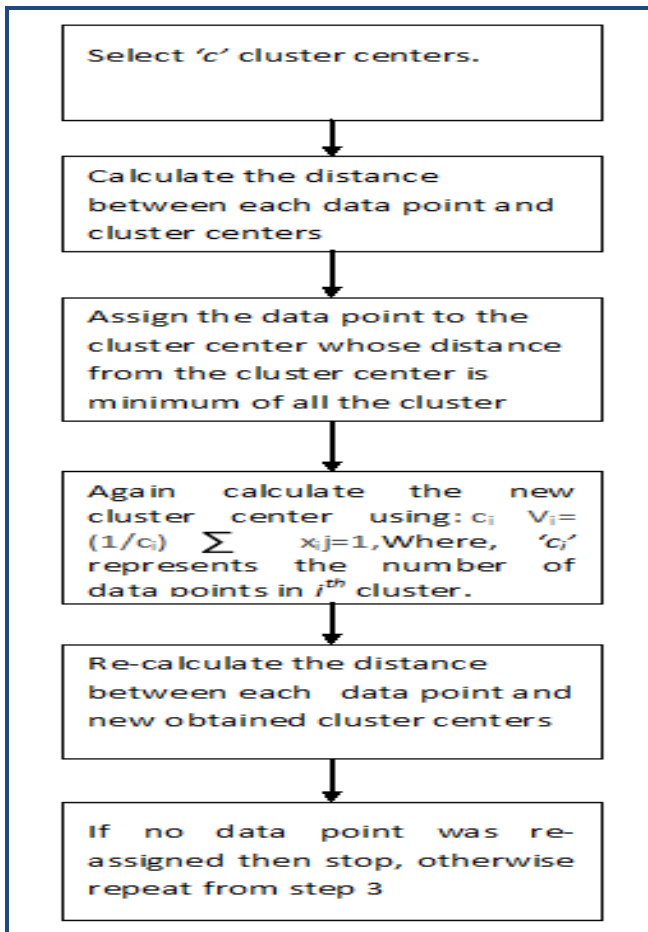
BIOINFORMATION

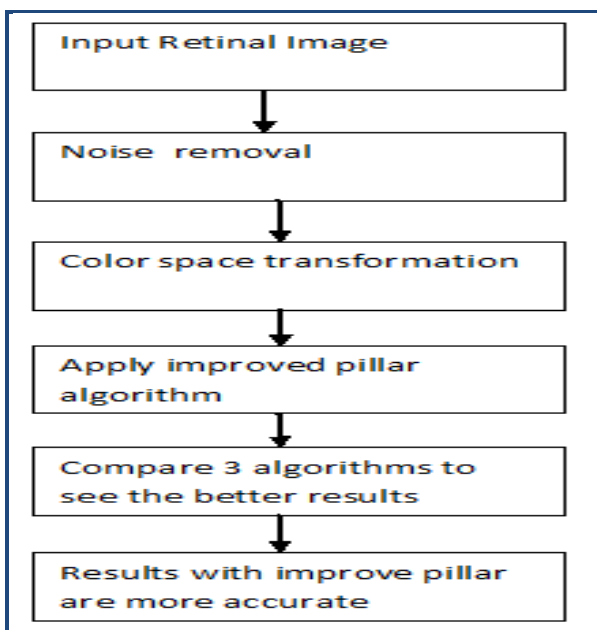**Figure 1:** K-Means algorithm applied before improved pillar k-means algorithm.

Figure 2: Flow chart to represent the data flow

Methodology:

Basics of K-Means and Fuzzy C-Means
This section briefly explains about the algorithms related to k-means and fuzzy c-means clustering techniques [4, 5].

Algorithmic for K-Means Clustering
K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean

Let $X = \{x1, x2, x3, \ldots, xn\}$ be the set of data points and $V = \{v1, v2, \ldots, vc\}$ be the set of centers **(Figure 1)**.

Algorithm for Fuzzy C Means
The fuzzy c-means algorithm is as same as the k-means algorithm. The algorithm minimizes intra-cluster variance, but has the same problems as k-means, the minimum is a local minimum, and the results are based on the initial choice of weights [6]. The expectation-maximization algorithm is a more statistically formalized method which includes some of these ideas: partial membership in classes. It has better convergence properties and is in general preferred to fuzzy-c-means.

Medical imaging is the special method and process used to create images of the human body for clinical purposes (medical procedures seeking to reveal, diagnose, or examine disease) or medical science (including the study of normal anatomy and physiology).

Measurement and recording techniques which are not primarily designed to produce images, such as electroencephalography (EEG), etc., but which produce data susceptible to be represented as maps (i.e., containing positional information), can be seen as forms of medical imaging [7].

Image Segmentation Approach
Image segmentation approach follows steps as
Reading the image, removing the noise from the image, color space transformation, data normalization, comparative analysis of algorithms (KM, FCM, IPKM) **(Figure 2)**.
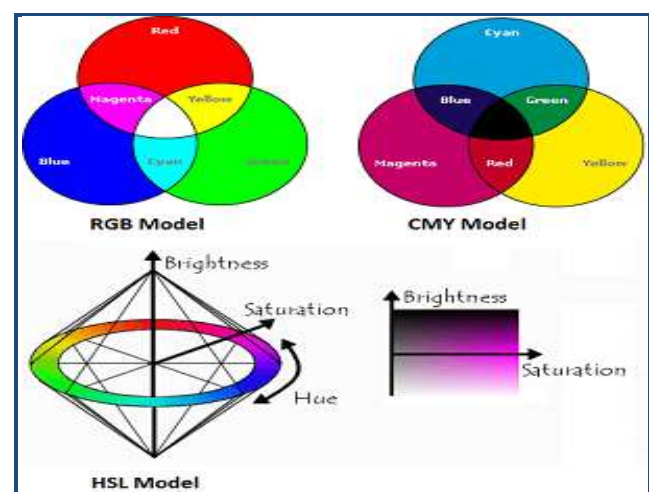

Figure 3: Different color models to represent color space transformation.

Discussion:

Noise Removal
An adaptive noise removal filtering using the median filter is applied for noise removal of images. This filter is widely used for solution of image restoration problems this is often used in processing to reduce salt pepper noise from image [8].

Color Space Transformation

Our image segmentation system pre-proceeds the image by transforming the color space from RGB to grayscale. Rgb2gray converts RGB images to grayscale by eliminating the hue and saturation Information while retaining the luminance **(Figure 3)**. RGB (red, green, blue); CMY (cyan, magenta, yellow); HSI (hue, saturation, intensity)

Our image segmentation system pre proceeds image by transforming the color space from RGB to HSL, RGB to grayscale and CIELAB color systems. HSL is the improved color space of HSV because it represents brightness much better than saturation since the hue component in the HSL color space integrates all chromatic information, it is more powerful and successful for segmentation of color images than the primary colors **[9]**. The CIELAB color system has the advantage of being approximately perceptually uniform, and it is better than the RGB color system based on the assumption of three statistically independent colors attributes. The CIELAB color space is also widely-used for image restoration and segmentation. Considering the advantages of each color system of HSL and CIELAB, in our approach we use both of them as hybrid color systems for image segmentation.

Data Normalization

Our image is processed and put in the form of rows and columns to avoid redundancy and duplication of data. Here Normalization is a process which changes the range of pixel intensity value. In our system, Soft max algorithm is used for the data normalization the output range is between 0 and 1.

Improved Pillar K Means

The system uses the real size of the image in order to perform high quality of the image segmentation. It causes high-resolution image data points to be clustered. Therefore we use the K-means algorithm for clustering image data considering that its ability to cluster huge data, quickly, effectively and efficiently. However, Because of initial starting points generated randomly, K-means algorithm is difficult to reach global optimum, but only to one of local minima which it will lead to incorrect clustering results **[10, 11]**. To avoid this phenomenon, we use our previous work regarding initial clusters optimization for K-means using Pillar algorithm. The Pillar algorithm is very robust and superior for initial centroids optimization for K-means by positioning all centroids far separately among them in the data distribution **[11]**.

This algorithm is inspired by the thought of determining a set of pillars, locations in order to make a stable building. Illustrates the locating of two, three, and four pillars, in order to withstand the pressure of several different roof structures composed of discrete points. It is inspiring that by distributing the pillars as far as possible from each other within the pressure distribution of a roof, the pillars can withstand the roof's pressure and stabilize a building, as number of centroids among the gravity weight of data distribution in the vector space. Therefore, this algorithm designates positions of initial centroids in the farthest accumulated distance between them in the data distribution **(Figure 4)** below is the illustration of locating a set of pillars (white points) withstanding against different pressure distribution of roofs **(Figure 4)**.
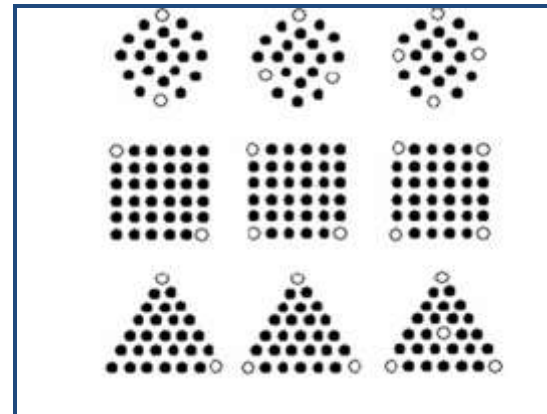

Figure 4: With standing against different pressure distribution of roofs white spots represent the pillars
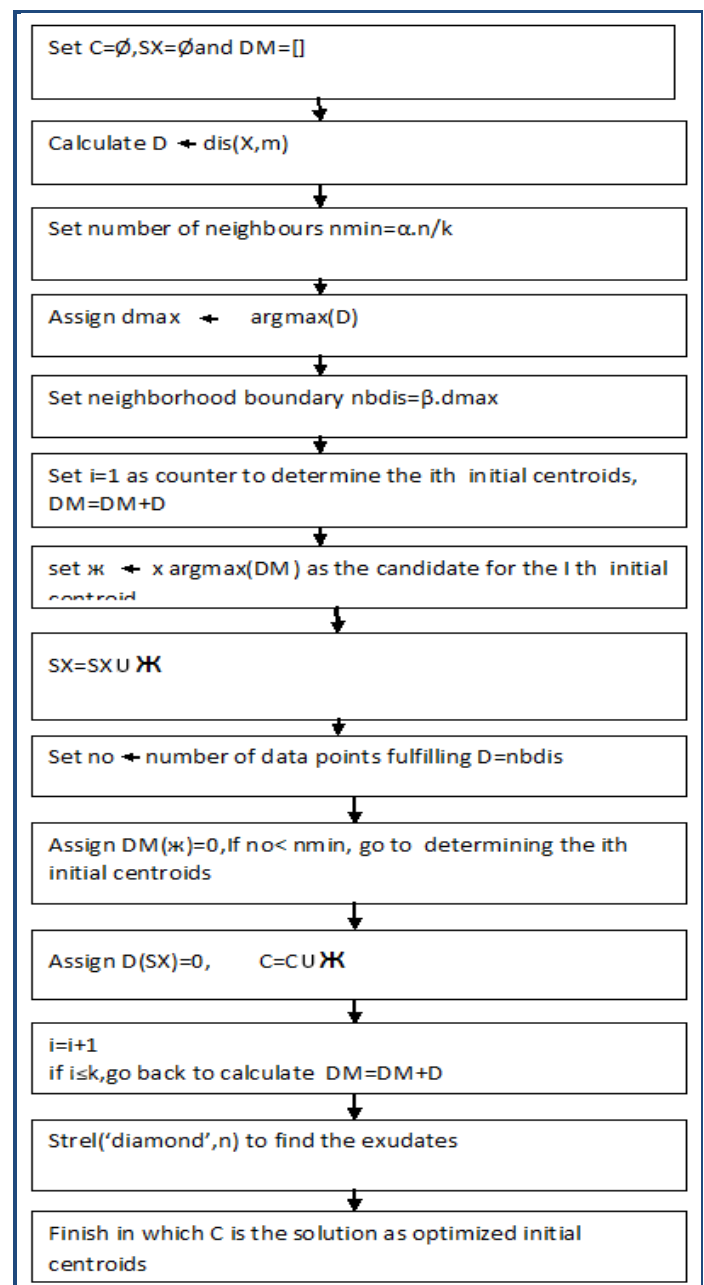

Figure 5: Improved pillar k-means algorithm

BIOINFORMATION

The Improved Pillar k-means algorithm is as follows:
Let X={xi | i=1,…,n} be data, k be number of clusters,

C={ci | i=1,…,k} be initial centroids, SX ⊆ X be identification for X which are already selected in the sequence of process,

DM={xi |i=1,…,n}

Be accumulated distance metric,

D={xi | i=1,…,n} be

distance metric for each iteration, and m be the grand mean of X.

The following execution steps of the proposed algorithm are described in **(Figure 5).** However, the computation may take long time if we apply the Pillar algorithm directly for all elements of high resolution image data points. In order to solve this problem, we reduce the image size to 5%, and then we apply the Pillar algorithm we apply clustering using the K-means algorithm and then obtain the position of final centroids. We use these final centroids as the initial centroids for the real size of the image as and then apply the image data point clustering using K-means. This mechanism is able to improve segmentation results and make faster computation for the image segmentation.
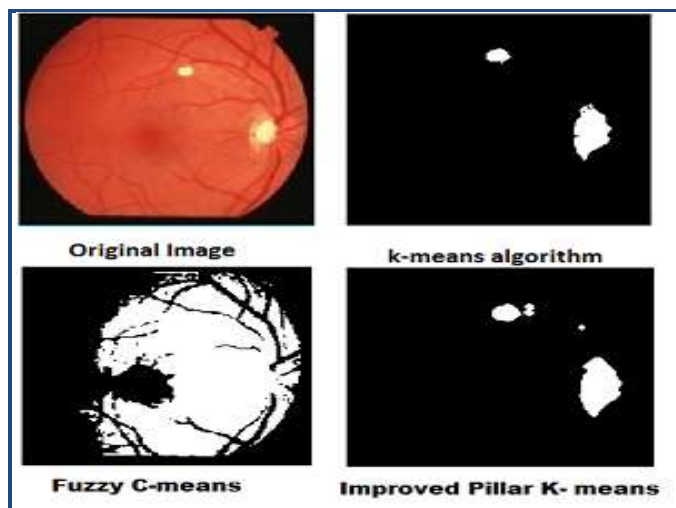

Figure 6: Results of different algorithms applied on diabetic retinopathy

Benefits of the Proposed System:
Computation time is less; Efficiency is more; Accuracy in detecting dark spots in the retinal images of initial stages.

Conclusion:
We presented a technique for segmentation of retinal images using improved pillar k means algorithm to find particular dark spots in the diabetic retina. Here in results **(Figure 6)**, **Table 1 (see supplementary material)** we find the dark spots in the retina with less computation time and more clearly compared with k-means and fuzzy c-means. Separation of macular is difficult. In further work detection and separation of macular can be done.

Acknowledgment:

References:
[1] Sravya K *et al. Medical image segmentation by pillar k means algorithm International journals of Advanced Engineering Technologies* 2013

[2] Neelapala AnilKumar *et al. Automatic detection of Adenocarcinoma using active counters International journal of advanced computer research* 2013

[3] Swetha V *et al. Diabetes care* 2008 **31:** 1905

[4] Manish Verma *et al. International journal of engineering Research and applications* Vol 2 issue3, Jun 2012

[5] Tou JT & Gonzalez RC, *Pattern Recognition Principles, Massachusetts: Addison-Wesley* 1974

[6] Pethalakshmi A *et al. International journal of science and research* 2013.

[7] Manimurugan S *et al. IOSR Journal of computer engineering* 2012 **5:** 01

[8] Debananda Padhi *et al. International journal of Advanced research in computer science and software engineering* 2012 **2:** 225

[9] Vivek Barwat *et al. IJAIR.* 2013

[10] Ali Ridho Barakbah *et al. International Journal of Information and Communication Engineering* 2010 **6:** 2

[11] Pavani M & Balaji S, *International Journal of Computer Trends and Technology* 2013 **4:** 636

BIOINFORMATION

Supplementary material:

Table 1: Computation time for different algorithms in first stage of Diabetic retinopathy Images

k-means	Fuzzy c-means	improved pillar K-algorithm
0.3125 s	4.0625 s	0.04682 s