A Fuzzy C-Means Based Feature Selection to Process Medical Data

C. Swathi
Dept of CSE,
JNTUCEA, ANANTAPURAMU.
EMAIL ID: swathi.chilakapati@gmail.com

S. Vasundra
Dept of CSE,
JNTUCEA, ANANTAPURAMU.
EMAIID:vasundras.cse@jntua.ac.in

Abstract- Classification and Feature selection are crucial steps in medical image data mining to diagnose the problems effectively. Micro Aneurysm, a retinal disease that mostly appears in diabetic patients is identified with a careful examination only. A classifier that recognises the micro aneurysm with the help of fuzzy technique has been proposed here. A Comparison has been made between the proposed approach and with the annealing and boosting approaches.

Keywords: *Feature selection, Fuzzy logic, Classification*

*****

I. INTRODUCTION

Feature Selection , the process of selecting useful and relevant features is the most essential step in classification problems. Medical imaging problems contains huge features and more observations. Hence a classifier that selects only necessary features is required. Such a classifier will encounter great challenges for feature selection.

Numerous ad-hoc procedures are available for feature selection problems. Recently, penalized methods and boosting algorithms have gained popularity in classifying high dimensional data like medical images. But these are not scalable and are time consuming. Moreover, these methods are not capturing nonlinearity adaptively. Simulated Annealing algorithm [1], a combination of regularization and sequential algorithm was also proposed for selecting features. The prototype of FSA is very simple to implement and complexity of problem is also reduced. But the irrelevant features are identified only when we are close to the optimal solution which in turn increases cost.

Eventhough , many methods are motivating there is a need to develop a technique that can simple to implement and easily adaptable.In this paper, a method for feature selection that makes use of Quantized Fuzzy C-means algorithm for classifying medical Images such that system will be trained and tested to diagnose problems in human eye, has been proposed.

Rest of the paper is organised as follows: Section II describes related work on fuzzy classifier, implementation of quantized fuzzy to classify infected eye is presented in Section III, the next Section IV gives information of the experiments being conducted and the appropriate results and Section V concludes the paper.

II. RELATED WORK

Following are the related works to Fuzzy C-Means for processing medical image data. While processing medical images, there is a high possibility to meet imprecise conditions[9], hence use of fuzzy theory is recommended as it deals with non-crisp sets and ambiguous classes which are very helpful in diagnosis.Fuzzy set theory [2] plays a key in formalizing uncertainties for medical diagnosis as observed by many researchers.

Zadeh *etal* (1965),Bezdek *etal* (1981),Adlassing *etal* (1986), Sterimann *etal* (1997), Kuncheva *etal*. (1999), Steimann *etal* (2001). Segmentation is the process of dividing an image into different regions which is the very crucial step in medical image diagnosis. Padmavathi implemented a fuzzy c means clustering method with thresholding for segmenting the medical images. Different modifications performed by researchers on conventional Fuzzy c-means were lead to the introduction of different algorithms using fuzzy logic. Chen and Zhang, 2004; Zhang and Chen, 2003 proposed KFCM,[3] a kernel based FCM which takes kernel distance in place of Euclidean distance to obtain effective results in segmentation of medical images. Brain MRI segmentation was done by Huynh Van Lung and Jong-Myon Kim who developed generalized Spatial Fuzzy C-Means Clustering algorithm (GSFCM) Comparing of spatial information and pixel attributes to neighbours is done in GSFCM.

It would be interesting to compare FCM with other traditional approaches like Annealing, Boosting and Penalized classifiers.

Boosting technique tried to take all the fourteen harliack features of an image. As its main goal is to consider all features, it doesn't maintain good accuracy on BigData, which result in identifying an image as micro aneurysm defected eye or an blotted eye even though it is not. To

overcome this Penalized techniques[4] were used that applied lossy functions which turns the inaccuracy into losses and tries to give the result.

Considering all features of an image belongs to large dataset is time consuming and leads to over fitting problems. Hence Regression method with Annealing technique was introduced such that it takes only four most useful features of image like Entropy, Contrast, Information Measures regarding to Correlation with respect to entropy and variance. Instead of taking unuseful features it takes only relevant ones that forms the completion of feature selection step. By making use of these features FSA classifier classifies given input image into one of the three classes labelled like N, Bt, Ma which refers to Normal, Blotted, Micro aneurysm images.

But the classifier is not accurately recognizing an Ma class images since it marking the areas as defected eventhough the regions are not actually effected with that disease.

To provide solution for this problem a Fuzzy classifier[5][7] will be used. In other words, to improve the accuracy of recognising an infected eye a quantized fuzzy means algorithm has been introduced. This methodology gives classification of given image as well as degree of belonginingness to that class. This provides level of confidence of classification decision. This is the lag in other classification approaches that are unable to predict non-crisp cases in given dataset.

III. PROPOSED SYSTEM

In order to identify whether an eye image is normal image or blotted one or Micro aneurysm effected image a Quantized Fuzzy C Means classification approach has been proposed.

Feature Selection is performed in QFCM by taking RGB Estimation, Contrast and Entropy among the fourteen features. We can show the good performance of QFCM in terms of accuracy such that how many micro aneurysms present in an image and in terms of reduction in time when compared with FSA and Boosting techniques.

Making use of fuzzy membership functions FCM assigns pixels to each class. Lets suppose X=(x1, x2, x3…xN) denotes an image with N pixels to be categorized into C clusters. Minimizing the below objective function will yield desired clusters of pixels.

$$J= \sum_{j=1}^{N} \sum_{i=1}^{C} a U_{ij}^{m} \| X_j - V_i \|^2 \text{ --- (1)}$$

where u_{ij} is the membership of pixel x_j in i^{th} cluster, v_i is the i^{th} cluster centre.

Let us take x classes with m attributes ,in our problem it is 3 classes Normal image, Blotted Image, and Micro aneurysm effected image with 3 attributes entropy, contrast and RGB estimates[8].

Firstly we have to calculate relative frequencies along each pixel nothing but one relative frequency for single feature before sorted into decreasing order. Next, the membership degrees μ(k) are computed per each feature using μ(k) =*kf*(*k*)*f* (*k* 1)*f* (*k* 2)..*f* (*n*) Lastly, the fuzzy set *FS*1, *FS* 2, *FS*3 having membership μ (*k*) are determined that is one fuzzy set for each attribute (for each dimension).The same will be repeated with the remaining classes. Following is the Algorithm has been used in the proposed system.

Algorithm Quantized Fuzzy C-Means(QFCM)

1: **procedure** SEGMENTATION (Image I, No. Of clusters c, No.of bins q)
2: Pre-process the image I.
3: initialize cluster centers V using the Ordering –split procedure(Algorithm 1).
4: **repeat**
5: Update partition matrix U.
6: Update prototypes matrix V.
7: **until** $\| V-V_{old} \| < \epsilon >$ $\| . \|$ is a matrix norm.
8: Regularize the partition U.
9: **return**(U,V).
10: **end procedure.**

IV. EXPERIMENTS AND RESULTS

A total of 100 Eye images have been collected from various sources of internet, among them 50 images were used for training and the remaining for testing phases respectively. These are a collection of Normal eye, Blotted eye and Micro aneurysm defected eye.

Let us look at the working of QFCM on these images.

1. Choose the QFCM classifier to train 50 images.

2. Note down the time taken for training to compare with the previous approaches.

3. Select an image for testing with QFCM.

It undergoes for quantization phase where compression of values to single quantum value has been achieved here. The result will is shown in the below Fig1.

 ISSN: 2321-8169
609 – 612

Fig1:Quantized output of a Normal eye

The output of QFCM for the above quantized image is shown in below Fig2.

Fig2: A normal eye detected by system

Hence the QFCM correctly recognised given image is Normal Eye image as the square boxes indicates that.
Now let us give Micro aneurysm effected eye image as input. The Quantized output of MA effected image is shown in Fig3:

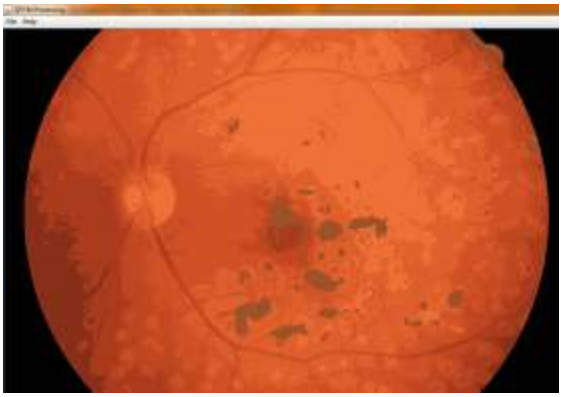

Fig 3: Quantized output of Ma effected eye.

The above image is clearly saying that it has micro aneurysms. The Output given by QFCM for the affected eye is shown in Fig4

Fig 4: Output of MA affected eye.

Thus QFCM spotted the micro aneurysms with black circles around the effected regions correctly.
By checking the accuracy in terms of identifying correct number of MAs, training time and testing time through QFCM, Annealing and Boosting techniques, implies that QFCM has best performance among them.
The following results will give a clear information about that.

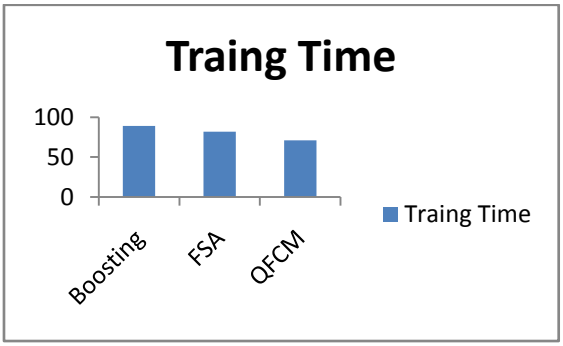

Fig5: Training Time taken by Boosting,FSA,QFCM

Hence training of 50 images has taken more time in annealing approach and even more in boosting technique when compared with quantized Fuzzy c-means approach.

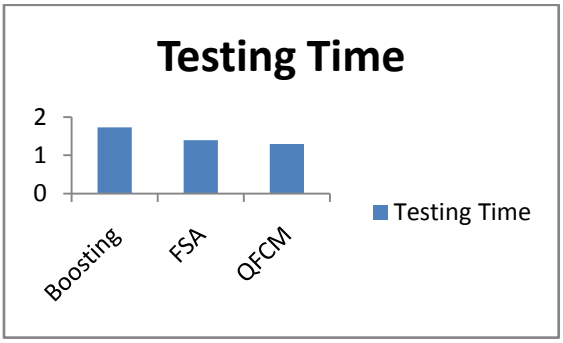

Fig 6: Testing time taken by Boosting, FSA,QFCM

Testing of Single image has also taken more time in Boosting and FSA techniques when compared with that of QFCM.

 ISSN: 2321-8169
609 – 612

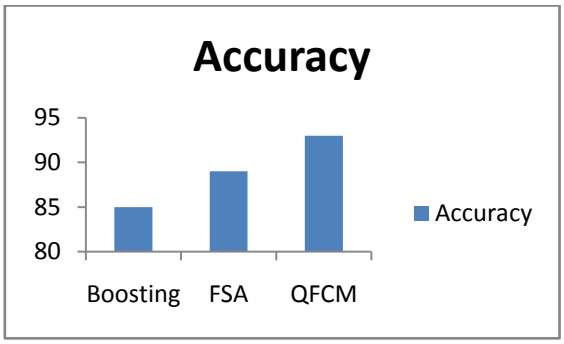

Fig7: Accuracy of Boosting, FSA, QFCM

Thus finding number of micro aneurysms in an image is correctly identified by QFCM greatly when compared with FSA and Boosting

V. CONCLUSION:

As feature selection is the most important step in classification, One should take care in selecting the features for accurate results. QFCM has performed well when compared with approaches like Selection with Annealing and Boosting techniques. It has given high accuracy and less testing and training time when compared with that of discussed techniques.

REFERENCES:

[1] Barbu, A., She, Y., Ding, L., & Gramajo, G. (2017). Feature selection with annealing for computer vision and big data learning. IEEE transactions on pattern analysis and machine intelligence, 39(2), 272-286.

[2] Ahmed Mohamed N., Yamany Sameh M., Mohamed Nevin, Farag Aly A. and Moriarty Thomas (2002). A Modified Fuzzy *c*-Means Algorithm for Bias Field Estimation and Segmentation of MRI Data. IEEE Transaction on Medical Imaging, Vol. 21, No. 3, March

[3] Seker, H., et al., "A fuzzy logic based-method for prognostic decision making in breast and prostate cancers". IEEE Trans Inf Technol Biomed, 2003. **7**(2): p. 114-22.

[4] Göbl, C. S., Bozkurt, L., Tura, A., Pacini, G., Kautzky-Willer, A., & Mittlböck, M. (2015). Application of penalized regression techniques in modelling insulin sensitivity by correlated metabolic parameters. *PloS one*, *10*(11), e0141524.

[5] Visa, S., Ralescu, A., "Data-Driven Fuzzy Sets ForClassification". International Journal of Advanced Intelligence Paradigms 2008 - Vol. 1, No.1 pp. 3 - 30, 2008.

[6] Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms.Plenum Press, New York.

[7] Ren, J., Jiang, X., & Yuan, J. (2015, April). Quantized fuzzy LBP for face recognition. In Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on (pp. 1503-1507). IEEE.

[8] Engelhard, N., Endres, F., Hess, J., Sturm, J., & Burgard, W. (2011, April). Real-time 3D visual SLAM with a hand-held RGB-D camera. In *Proc. of the RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum, Vasteras, Sweden* (Vol. 180, pp. 1-15).

[9] S.Vasundra,K.NagaDivya,A.Madhuri, T.Srilakshmi (2012 september). Cancer Detection by advanced technology. IJCST Vol. 3, Issue 3, July - Sept 2012