American Journal of Applied Sciences 8 (11): 1159-1168, 2011
ISSN 1546-9239
© 2011 Science Publications

Dermoscopic Image Segmentation using Machine Learning Algorithm

[1]L. Padma Suresh, [2]K.L. Shunmuganathan and [3]S.H. Krishna Veni
[1] Research Scholar, CSE Department, Manonmaniam Sundaranar University, Tvl, India
[2]Computer Science Engineering Department, RMK Engineering College, India
[3]IT Department, Noorul Islam Centre for Higher Education, India

Abstract: Problem statement: Malignant melanoma is the most frequent type of skin cancer. Its incidence has been rapidly increasing over the last few decades. Medical image segmentation is the most essential and crucial process in order to facilitate the characterization and visualization of the structure of interest in medical images. **Approach:** This study explains the task of segmenting skin lesions in Dermoscopy images based on intelligent systems such as Fuzzy and Neural Networks clustering techniques for the early diagnosis of Malignant Melanoma. The various intelligent systems based clustering techniques used were Fuzzy C Means Algorithm (FCM), Possibilistic C Means Algorithm (PCM), Hierarchical C Means Algorithm (HCM); C-mean based Fuzzy Hopfield Neural Network, Adaline Neural Network and Regression Neural Network. **Results:** The segmented images were compared with the ground truth image using various parameters such as False Positive Error (FPE), False Negative Error (FNE) Coefficient of similarity, spatial overlap and their performance was evaluated. **Conclusion:** The experimental results show that Hierarchical C Means algorithm(Fuzzy) provides better segmentation than other (Fuzzy C Means, Possibilistic C Means, Adaline Neural Network, FHNN and GRNN) clustering algorithms. Hierarchical C Means approach can handle uncertainties that exist in the data efficiently and useful for the lesion segmentation in a computer aided diagnosis system to assist the clinical diagnosis of dermatologists.

Key words: Fuzzy C Means (FCM), Hierarchical C Means (HCM), Possibilistic C Means (PCM), hopfield neural network, regression neural network, adaline neural network, false positive, negative error, spatial overlap

INTRODUCTION

Dermoscopy, also known as epiluminescence microscopy, is a non-invasive skin imaging technique that uses optical magnification and either liquid immersion or cross-polarized lighting, making subsurface structures more easily visible when compared to conventional clinical images. Dermoscopy allows the identification of dozens of morphological features such as pigment networks, dots/globules, streaks, blue-white areas and blotches. This reduces screening errors and provides greater differentiation between difficult lesions such as pigmented Spitz nevi and small clinically equivocal lesions. The standard approach in automatic dermoscopic image analysis has usually three stages: (1) image segmentation; (2) feature extraction and feature selection; (3) lesion classification. The segmentation stage is one of the most important since it affects the accuracy of the subsequent steps. The segmentation technique subdivides an image into its constituent regions or objects. The segmentation should stop when the objects or regions of interest in an application have been detected. It is used in image analysis and recognition. For instance, for the automated detection of cancerous cells from mammographic images, segmentation followed by recognition or classification is required. Most of the segmentation algorithms are based on one of two basic properties of intensity values: discontinuity and similarity. In the first category, the approach is to partition an image based on abrupt changes in intensity, such as edges. The principal approaches in the second category are based on partitioning an image into regions that are similar according to a set of predefined criteria. A large numbers of algorithms have been proposed in previous years. One of the Conventional image segmentation algorithms is clustering by which homogeneous properties around a given pixel are enlarged.

The crisp segmentation methods such as thresholding (Ganster *et al*., 2001), region growing, Edge-based approaches (Kapur *et al*., 1996), k-means and split and merge methods are generally used for image segmentation. . Besides this, soft segmentation methods were also seen effective for segmentation.

Corresponding Author: L. Padma Suresh, Research Scholar, CSE Department, Manonmaniam Sundaranar University, Tvl, India

Segmentation methods are applied from the artificial intelligence field, especially using fuzzy and neural networks approaches. The present survey is intended to be a more comprehensive study of the existing Fuzzy and Neural-network-based segmentation techniques. In this study we propose and evaluate several Fuzzy and Neural Network based clustering techniques: Fuzzy C Means Algorithm (FCM), Possibilistic C Means Algorithm, Hierarchical C Means Algorithm, C-mean based Fuzzy Hopfield Neural Network, Adaline Neural Network and Regression Neural Network. These algorithms are applied to the dermoscopic image and are compared with the expected lesion segmentation (ground truth). The evaluation is based on different parameters and quality metrics that take into account different types of error.

MATERIALS AND METHODS

Intelligent fuzzy and neural network based clustering techniques: The Intelligent system is a branch of computer science concerned with making computers behave like humans. Cluster analysis is a technique for classifying data, i.e., to divide a given dataset into a set of classes or clusters. In classical cluster analysis each datum must be assigned to exactly one cluster. The intelligent system cluster analysis relaxes this requirement by allowing gradual memberships, thus offering the opportunity to deal with data that belong to more than one cluster at the same time. Crisp clustering assigns each data to a single cluster but in fuzzy the membership function measures the degree of belonging of each feature in a cluster. Most fuzzy clustering algorithms are objective function based: They determine an optimal classification by minimizing an objective function.

The degrees of membership to which a given data point belongs to the different clusters are computed from the distances of the data point to the cluster centers. Several fuzzy clustering algorithms can be distinguished depending on the additional size and shape information contained in the cluster prototypes, the way in which the distances are determined and the restrictions that are placed on the membership degrees (Silveira *et al*., 2009; Lin *et al*., 1996). Here we focus on the fuzzy c-means algorithm (Zhang, 2006), which uses only cluster centers and a Euclidean distance function and the Gustafson{Kessel algorithm, which uses cluster centers, covariance matrices and a Mahalanobis distance function. Hopfield network is a recurrent network in which all neurons are connected to each other, with the exception that no neuron has any connection to itself (Bezdek and Pal, 1992). Computer Aided Diagnosis (CAD) system for the detection of brain tumor by using parallel implementation of ACO system for medical image segmentation applications due to the rapid execution for obtaining and extracting the Region of Interest (ROI) from the images for diagnostic purposes in medical field (Jaya and Thanushkodi, 2011). A New Modified Gaussian Mixture Model for Color-Texture segmentation presents a new, simple and efficient modified Gaussian mixture model based clustering algorithm for color-texture segmentation. The proposed mixture model introduces a new component density function which incorporates spatial information and the weighting factor for neighborhood effect is fully adaptive to the Image conten (Sujaritha and Annadurai, 2011)

A genetic algorithm based segmentation is implemented in the process of computer vision and object classification. The objective of this study was to develop a robust technique for the automatic segmentation and classification of touching objects(Scavino *et al*., 2009). This survey describes the current state of the ongoing the BC automated diagnosis research program and describes a software system that provides expert diagnosis of breast cancer based on three step of cytological image analysis (Sebri *et al*., 2007)

This work propose and compare various intelligent fuzzy (Bezdek and Pal, 1992) and neural network based clustering techniques. The various intelligent system based clustering methods are:

- Fuzzy C Means Algorithm (FCM)
- Possibilistic C Means Algorithm (PCM)
- Hierarchical C Means Algorithm (HCM)
- Fuzzy Hopfield Neural Network (FHNN)
- Regression neural network
- Adaline neural network

The intelligent clustering technique differs from the conventional hard computing in that, unlike the later, it is tolerant of imprecision, uncertainty, partial truth and approximation.

Fuzzy C Means algorithm (FCM): The most prominent algorithm is the FCM or Fuzzy C Means algorithm. The FCM algorithm receives the data or sample space in matrix format. The number of clusters, the assumption partitioning matrix, the convergence value all must be given to the algorithm. The FCM algorithm assigns pixels to each category by using fuzzy memberships N. The algorithm is an iterative optimization that minimizes the cost function. The number of clusters c, the assumption partition matrix U, convergence value E all must be given

to the algorithm. The first step is to calculate the cluster centers. The second step is to calculate the distance matrix d. The distance matrix constitutes the Euclidean distance between every pixel and every cluster center. If the difference between the initial partition matrix and calculated partition matrix is greater than the convergence value then the entire process from calculating the cluster centers to the final partition matrix. The final partition matrix is taken and used to reconstruct the image. The cluster centroid V_i for each cluster Eq. 1:

$$V_i = \frac{\sum_{j=1}^{N} (u_{ij})^m X_j}{\sum_{j=1}^{N} (u_{ij})^m} \quad i = 1, 2....C \tag{1}$$

The objective function is minimized when pixels close to the centroids are assigned high membership values and low membership values assigned to pixel far from centroid. The standard FCM objective function is given by Eq. 2:

$$J_m(U, V) = \sum_{j=1}^{N} \sum_{i=1}^{C} u_{ij}^{\ m} d^2(X_j, V_i) \tag{2}$$

where, X = {X1, X_2 ,……X_j ,…..X_N } is a p×N input data matrix, where p represents the dimension of each feature vector and N represents the number of feature vectors. C is the number of clusters, U_{ij} represents the membership function of the jth data in ith cluster C_i, d is the distance between input and centroid, V_i is the ith cluster center and m is a constant. The membership functions and cluster centers are updated by the following Eq. 3:

$$Uij = \frac{1}{\sum_{K=1}^{C} \left(d(x_j, v_{ij}) \,|\, d(x_j, v_k)\right)^{2/m-1}} \tag{3}$$

where, m is a weighting factor which controls the degree of fuzziness. A measure of similarity between X_j and V_i is given as Eq. 4:

$$d^2(X_j, V_i) = \| X_j - V_i \| \tag{4}$$

Convergence can be detected by comparing the changes in the membership function or the cluster center at two successive iteration steps.

Possibilistic C Means algorithm (PCM): In possibilistic fuzzy (Zhang and Jiang, 2009) clustering one tries to achieve a more intuitive assignment of degrees of membership by dropping the probability constraint of FCM, which is responsible for the undesirable effect. However, this leads to the mathematical problem that the objective function is now minimized by assigning U_{ij} = 0 for all i € {1,……, c} and j € {1,….., n}. In order to avoid this trivial solution, a penalty term is introduced, which forces the membership degrees away from zero. That is, the objective function J is modified to Eq. 5:

$$J_m(x, \mu, c) = \sum_{i=1}^{c} \sum_{j=1}^{N} \mu_{ij}^m d_{ij}^2 + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{N} (1 - \mu_{ij})^m \tag{5}$$

where, d_{ij} is the distance between the jth data and the ith cluster center, μ_{ij} is the degree of belonging of the jth data to the ith cluster, m is the degree of fuzziness, η_i is a suitable positive number, c is the number of the clusters and N is the number of the data. μ_{ij} can be obtained as Eq. 6:

$$\mu = \frac{1}{1 + \left(\frac{d_{ij}^2}{\eta_i}\right)^{\frac{1}{m-1}}} \tag{6}$$

where, d_{ij} is the distance between the jth data and the ith cluster center, μ_{ij} is the degree of belonging of the jth data to the ith cluster, m is the degree of fuzziness, η_i is a suitable positive numbers. The value of η_i determines the distance at which the membership value of a point in a cluster becomes 0.5. The value of η_i is obtained as Eq. 7:

$$\eta_i = \frac{\sum_{j=1}^{N} \mu_{ij}^m d_{ij}^2}{\sum_{j=1}^{N} \mu_{ij}^m} \tag{7}$$

The value of η_i can be fixed or changed in each iteration by changing the values of μ_{ij} and d_{ij}. This method is more robust in the presence of noise, in finding valid clusters and in giving a robust estimate of the centers. At first sight this approach looks very promising. However, if we take a closer look, we discover that the objective function J defined above is, in general, truly minimized only if all cluster centers are identical. The reason is that formula for the membership degree of a datum to a cluster depends only on the distance of the datum to that cluster, but not on its distance to other clusters. Hence, if there is a single optimal point for a cluster center (as it will usually be the case, since multiple optimal points would require a high symmetry in the data), all cluster centers will converge to this point. More formally, consider two cluster centers β_1 and β_2 which are not identical and let Eq. 8:

$$z_i = \sum_{j=1}^{n} u_{ij}^m d^2(\beta_i, x_j) + \eta_i \sum_{j=1}^{n} (1 - u_{ij})^m \tag{8}$$

That is, let z_i be the amount that clusters β_i contributes to the value of the objective function. Except in very rare cases of high data symmetry, it will then either be $z_1 > z_2$ or $z_2 > z_1$. That is, we can improve the value of the objective function by setting both cluster centers to the same value, namely the one which yields the smaller z-value, because the two z-values do not interact. In the probabilistic approach the cluster centers are driven apart, because a cluster, in a way, consumes part of the weight of a datum and thus leaves less that may attract other cluster centers. Hence sharing a datum between clusters is disadvantageous. In the possibilistic approach there is nothing equivalent to this effect. Nevertheless, possibilistic fuzzy clustering (Steck and Balakrishnan, 1994) usually leads to acceptable results, although it suffers from stability problems if it is not initialized with the corresponding probabilistic algorithm. We assume that other results than all cluster centers being identical are achieved only, because the algorithm gets stuck in a local minimum of the objective function.

Hierarchical C Means algorithm (HCM): Given a set of elements X, a mixed approach is applied to build a fuzzy hierarchical structure. The process starts building a fuzzy partition of X applying fuzzy c-means. This results into a set of fuzzy membership functions μ_i, each one built on the centroid v_i. This fuzzy partition bootstraps the process. Then, the iterative process is applied to build the hierarchical clustering following a bottom-up strategy. Such set of membership functions is partitioned using a partitive clustering method for fuzzy sets. Such partitive clustering method returns a new fuzzy partition $\mu_i^{/}$ that is used as the starting point of the new step.

In this algorithm, the fuzzy c-means algorithm is used for building the initial fuzzy partition. Such fuzzy partition is obtained by applying the fuzzy c-means algorithm to X. In this case, the algorithm is applied with a large number of clusters (i.e., c is large). This selection of c is to have a large number of leaves in the fuzzy hierarchy. In the iterative process, fuzzy c-means based clustering method is used. Differences consist on the way the distance $\|x_k-v_i\|$ is computed. Here, x_k and v_i represent fuzzy sets. More specifically, x_k stands for the k-th fuzzy set to be partitioned and v_i is one of the fuzzy sets in the new partition. Accordingly, $\|x_k-v_i\|$ is a distance between fuzzy sets. Following the standard approach in fuzzy c-means, the fuzzy membership of a fuzzy set with centroid v is defined considering all other centroids v_i. In our case, the membership of the fuzzy set with centroid x_k is computed for all x taking into account all other centroids x_j as follows Eq. 9:

$$\mu_{x_k}(x) = \left(\sum_{j=1}^{C} \left(\frac{d(x_k, x)^2}{d(x_k, x_j)^2} \right)^{\frac{1}{m-1}} \right)^{-1} \tag{9}$$

Similarly, the membership of the fuzzy set with centroid v_i is computed for all x taking into account all other centroids v_j as follows Eq. 10:

$$\mu_{v_k}(x) = \left(\sum_{j=1}^{C} \left(\frac{d(v_k, x)^2}{d(v_k, v_j)^2} \right)^{\frac{1}{m-1}} \right)^{-1} \tag{10}$$

Note that here, x_j are the centroids of the fuzzy sets being clustered and v_j is the centroids of the clusters we are constructing with the fuzzy c means. Similarly, c is the number of centroids xj_j and c is the number of centroids in v_j .Then, the distance between a fuzzy set with centroid x_k and another with centroid v_i will be computed. This is, how to determine the new v_i Eq. 11:

$$v_i = \frac{\sum_{k=1}^{n} (u_{ik})^m x_k}{\sum_{k=1}^{n} (u_{ik})^m} \tag{11}$$

Note that this approach leads to different membership values. Thus this method builds hierarchies of clusters where membership to clusters is fuzzy.

Fuzzy Hopfield Neural Network method (FHNN method): This method use the fuzzy c-means algorithm to eliminate the need for finding weighting factors in the Lyapunov energy function. The number of neurons used to construct the Network depends on the image size; the larger the image size, the more neurons that are required. These neurons are fully interconnected. The total input of neuron (i, k) denoted as Neti,k can be formulated as Eq. 12:

$$Net_{i,k} = \sum_{q=1}^{N} \sum_{j=1}^{c} W_{i,k;j,q} V_{j,q} + I_{i,k} \tag{12}$$

where, N is the number of data points, c is number of clusters, $V_{j,q}$ denotes the binary state of neuron (j,q), $W_{i,j,k,q}$ is interconnection weight between neuron (i,k) and neuron (j,q), $I_{i,k}$ is external bias vector for neuron (i,k). The Hopfield neural network consists of N x c neurons that can be conceived as a 2-D array for the image-segmentation problem and the Lyapunov energy function is given (Roozbahani *et al*., 2001) as Eq. 13:

$$E = \frac{1}{2}\sum_{k=1}^{N}\sum_{q=1}^{N}\sum_{i=1}^{c}\sum_{j=1}^{c} V_{i,k} W_{i,k;j,q} V_{j,q} - \sum_{k=1}^{N}\sum_{i=1}^{c} I_{i,k} V_{i,k} \quad (13)$$

When the Lyapunov energy function is minimized, the neural network reaches a stable state. The optimization problem can be mapped into a 2D fully interconnected Hopfield neural network with the fuzzy c-means algorithm. The total input for neuron (i, k) can be modified (Ganster *et al*., 2001) as Eq. 14:

$$Net_{i,k} = \left[x_k - \sum_{q=1}^{N} W_{i,k;i,q} \mu_{i,q}^{m} \right]^2 + I_{i,k} \quad (14)$$

Lyapunov energy can be changed (Roozbahani *et al*., 2001), m is the fuzzification parameter, $\sum_{a}^{N} = 1 \; W_{i,k;i,q}\mu_{i,q^M}$ is the total weighed input received from the neuron (i,q), x_k is x pixel value of image and membership value $\mu_{i,k}$ is the output state at neuron (i, k). A neuron (i, k) in a maximum membership state indicates that x_k pixel belongs to class i. The 2D Hopfield neural network represents cluster centroids in columns and image pixel in rows. In order to generate an adequate classification with the constraints, we define Lyapunov energy function as follows (Roozbahani *et al*., 2001) Eq. 15:

$$E = \frac{1}{2}\sum_{k=1}^{N}\sum_{i=1}^{c}\mu_{i,k}^{m}\left[x_k - \sum_{q=1}^{N}\frac{1}{\sum_{h=1}^{N}\mu_{i,h}^{m}} x_q \mu_{i,q}^{m} \right]^2 - \frac{B}{2}\left[\left(\sum_{k=1}^{N}\sum_{i=1}^{c}\mu_{i,k}^{m}\right) - N\right] \quad (15)$$

E is the total intra-class scatter energy that accounts for the scattered energies distributed by all pixels in same class. More specifically, the first term within-class scatter energy, minimizes the intra-class Euclidean distance from a sample to the cluster center in any given cluster and the second term which guarantees those number of data point N in image can only be distributed among these c classes, imposes constraints on the objective function (Roozbahani *et al*., 2001). The quality of classification result is very sensitive to the weighting factors and to search optimal values for these weighting factors is expected to be time-consuming and laborious. To alleviate this problem, a Hopfield neural network with a fuzzy c-means clustering method, called FHNN, is proposed. Because each image pixel can only be occupied by one class, the summation of states in the same row equals 1. This also ensures that only N data points will be classified into these c clusters . That is, the network must match the following constraints Eq. 16:

$$\sum_{i=1}^{c}\mu_{i,k} = 1 \quad \sum_{k=1}^{N}\sum_{i=1}^{c}\mu_{i,k} = N \quad (16)$$

Therefore, the energy function can be further simplified as Eq. 17:

$$E = \frac{1}{2}\sum_{k=1}^{N}\sum_{i=1}^{c}\mu_{i,k}^{m}\left[x_k - \sum_{q=1}^{N}\frac{1}{\sum_{h=1}^{N}\mu_{i,h}^{m}} x_q \mu_{i,q}^{m} \right]^2 \quad (17)$$

The normalization operation guarantees that each image pixel will be absorbed on several classes with certain probability degrees so there will be N data points assigned among c clusters. The minimization of energy E is greatly simplified because it contains only one term and hence the requirement of having to determine the weighting factors A and B vanishes. The synaptic interconnection weights and the bias input can be obtained as Eq. 18:

$$W_{i,k;j,q} = \frac{x_q}{\sum_{h=1}^{N}\mu_{i,h}^{m}} I_{i,k} = 0 \quad (18)$$

The input to neuron (i,k) can be expressed as Eq. 19:

$$Net_{i,k} = \left[x_k - \sum_{q=1}^{N}\frac{1}{\sum_{h=1}^{N}\mu_{i,h}^{m}} x_q \mu_{i,q}^{m} \right]^2 = \left[x_k - v_i\right]^2 = D_{i,k}^{2} \quad (19)$$

The membership function for k-th pixel is given as Eq. 20:

$$\mu_{i,k} = \left[\sum_{j=1}^{c}\left(\frac{Net_{i,k}}{Net_{j,k}}\right)^{\frac{2}{m-1}} \right]^{-1} \quad (20)$$

This membership function is effective to minimize new objective function in iteration. New objective function consists of average distance between image pixels and cluster centroids for separate and compact clustering. New objective function is given as Eq. 21:

$$J = \frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{c}\mu_{i,k}^{m} \; D_{i,k}^{2} \tag{21}$$

The FHNN Algorithm steps are given below:

- Given the data set X, choose the number of clusters 1<c<N, the weighting exponent m>1 (Membership functions for large value m are fuzzier than those for small value m, but the interconnection weights are updated slowly), the termination tolerance ε>0 (is used as a criterion to determine the performance of the objective function. The larger the threshold value ε, the less the numbers of iterations will, however, be the optimal membership function cannot be found) and the norm-inducing matrix A.
- Normalization, (gray levels of image)
- Calculate of primary centroids v0
- Compute the distances

$D^2_{ikA} = (x_k - v_i)^T A\,(x_k - v_i)$, $1 \le i \le c$, $1 \le k \le N$.

Compute the initial membership value:

$$U^{(0)} = \mu_{i,k}^{(0)} = \left[\sum_{j=1}^{c}\left(\frac{D_{ikA}}{D_{jkA}}\right)^{\frac{2}{m-1}}\right]^{-1}$$

Compute new cluster centroid:

$$v_i = \sum_{q=1}^{N}\frac{1}{\sum_{h=1}^{N}\mu_{i,h}^{m}}x_q\mu_{i,q}^{m}$$

Calculate the input to each neuron (i, k) Eq. 22:

$$Net_{i,k} = \left[x_k - \sum_{q=1}^{N}\frac{1}{\sum_{h=1}^{N}\mu_{i,h}^{m}}x_q\mu_{i,q}^{m}\right]^2 \tag{22}$$

Compute new membership value (Fuzzy c-means) Eq. 23:

$$\mu_{i,k} = \left[\sum_{j=1}^{c}\left(\frac{Net_{i,k}}{Net_{j,k}}\right)^{\frac{2}{m-1}}\right]^{-1} \tag{23}$$

Compute J^t Eq. 24:

$$J^t = \frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{c}\mu_{i,k}^{m} \; D_{i,k}^{2} \tag{24}$$

If $|J^{t+1} - J^{t}| > \in$ go to step 6, otherwise stop

Adaline neural network: ADALINE (Adaptive Linear Neuron or later Adaptive Linear Element) is based on the McCulloch-Pitts neuron. It consists of a weight, a bias and a summation function. It was first developed to recognize binary patterns so that if it was reading streaming bits from a phone line, it could predict the next bit. The difference between Adaline and the standard perceptron is that in the learning phase the weights are adjusted according to the weighted sum of the inputs (the net). In the standard perceptron, the net is passed to the activation (transfer) function and the function's output is used for adjusting the weights. Adaline is a neural network with multiple nodes where each node accepts multiple inputs and generates one output. Given the following variables:

- x is the input vector
- w is the weight vector
- n is the number of inputs
- θ some constant
- y is the output

Then we find that the output is Eq. 25:

$$y = \sum_{j=1}^{n} x_j w_j + \theta \tag{25}$$

If we further assume that $x_{n+1} = 1$ and $w_{n+1} = \theta$ then the output reduces to the dot product of x and w, $y = x_j . w_j$. Adaline Network is a simple Neural Network with two Neuron Layers-one input Neuron Layer and one output Neuron Layer. The output layer has only one Neuron node. Adaline Network is also the first Neural Network we built that "learns". The learning rule is simple: We give it some input values, fire the Network and compare the output value with the desired value. If there is any discrepancy, the Links in the Link Layer will adjust their weights until the rate of error is smaller than our tolerance. The weight vector w can be obtained by minimizing the least-squares-error criterion. The delta learning rule adopted in ADALINE is a data-adaptive technique for deriving a least-squares-error solution. Let us assume:

- η is the learning rate (some constant)
- d is the desired output
- is the actual output

Then the weights are updated as follows $w \leftarrow w+\eta(d-o)x$. The ADALINE converges to the least squares error which is $E = (d^*-o)^2$.

Regression neural network: Generalized regression neural networks are a kind of radial basis network that is often used for function approximation. Radial basis transfer function calculates a layer's output from its net input. This takes one input, N-S×Q matrix of net input (column) vectors and returns each element of N passed through a radial basis function. The probability density function used in GRNN is the normal distribution Eq. 26:

$$Y(X) = \frac{\sum_{i=1}^{n} Y_i \exp\left(\frac{-D_i^2}{2\sigma^2}\right)}{\sum_{i=1}^{n} \exp\left(\frac{-D_i^2}{2\sigma^2}\right)} \quad (26)$$

$$D_i^2 = (X - X_i)^T.(X - X_i)$$

Each training sample, X_j, is used as the mean of a Normal Distribution. The distance, D_j, between the training sample and the point of prediction, is used as a measure of how well the each training sample can represent the position of prediction, X. If the Distance, D_j, between the training sample and the point of prediction is small, exp $(-D_j^2/2\sigma^2)$, becomes big. For D_j =0, exp $(-D_j^2/2\sigma^2)$, becomes one and the point of evaluation is represented best by this training sample. The distance to all the other training samples is bigger. A bigger distance, D_j, causes the term exp $(-D_j^2/2\sigma^2)$ to become smaller and therefore the contribution of the other training samples to the prediction is relatively small. The term Yj* exp $(-D_j^2/2\sigma^2)$ for the jth training sample is the biggest one and contributes very much to the prediction. The standard deviation or the smoothness parameter is subject to a search. For a bigger smoothness parameter, the possible representation of the point of evaluation by the training sample is possible for a wider range of X. For a small value of the smoothness parameter the representation is limited to a narrow range of X, respectively. With it is possible to:

- Predict behavior of systems based on few training samples
- Predict smooth multi-dimensional curves
- Interpolate between training samples

Performance measures: Different parameters were used to analyze the performance of various fuzzy clustering algorithms. They are False Positive Error (FPE), False Negative Error (FNE) Coefficient of similarity and spatial overlap. To define the first two types of quality metrics let SR denote the result of an automatic segmentation method and GT denote the ground truth segmentation obtained by the medical expert. Both SR and GT are binary images such that all the pixels inside the curve hove label 1 and all others have label 0. The metrics are calculated as follows:

False Positive Error (FPE): These metric measures the rate of pixels classified as lesions by the automatic segmentation that were not classified as lesion by the medical expert Eq. 27:

$$FPE(SR,GT) = \frac{\#(SR \cap \overline{GT})}{\#(GT)} \quad (27)$$

False Negative Error (FNE): The FNR measures the rate of pixels classified as lesions by the medical expert that were not classified as lesion by the automatic segmentation Eq. 28:

$$FNE(SR,GT) = 1 - \frac{\#(SR \cap GT)}{\#(GT} \quad (28)$$

Clinically, this is worse of two types of error.

Coefficient of Similarity: The coefficient of similarity is the measure of relatedness between the automatic and manual segmentation (12) and is given by Eq. 29:

$$\in = 1 - \frac{|v_{manual} - v_{automatic}|}{v_{manual}} \quad (29)$$

The value of 1 represents perfect overlap and 0 represents no overlap.

Spatial overlap: The measure of spatial overlap between the automatic (algorithmic) and the manual segmentation is given as Eq. 30:

$$\in_s = \frac{2 * v_{intersection}}{v_{manual} + v_{algorithm}} \quad (30)$$

Spatial Overlap is more accurate measure of agreement than the coefficient of similarity, because the approach takes into account the spatial properties of the segmented region. It is more sensitive to small unmatched errors.

RESULTS AND DISCUSSION

Six different segmentation methods were simulated for melanoma diagnosis. The evaluation was based on the performances measures calculated and tested using ground truth image which is manually segmented. The

proposed Fuzzy clustering algorithms is simulated using MATLAB and tested with ground truth image to explore the segmentation accuracy of the various fuzzy clustering techniques. The effectiveness of the proposed approach is experimentally determined using the ground truth image.

The input malignant melanoma image is as shown in Fig. (1) was selected randomly from the database. The size of the image is 256×1200. Fig. 2-7 shows the segmentation results of malignant melanoma image using various intelligent systems based clustering techniques.

In the segmented output , the white region indicates the infected region (portions of skin affected with malignant melanoma) and the black trace or spots indicates the non-infected region (portions of the skin free from the malignant melanoma).

Fig.1: Input Image (MM image)

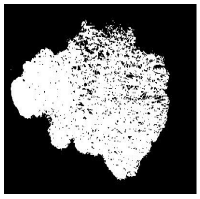

Fig.2: Segmented image using FCM

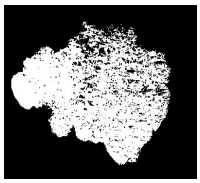

Fig. 3: Segmented image using PCM

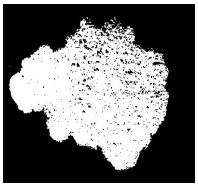

Fig. 4: Segmented image using HCM

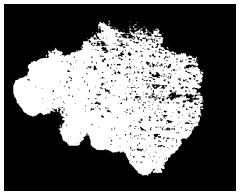

Fig. 5: Segmented image using ANN

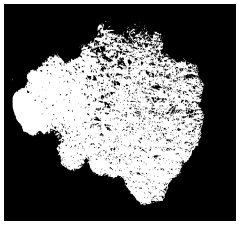

Fig.6: Segmented image using FHNN

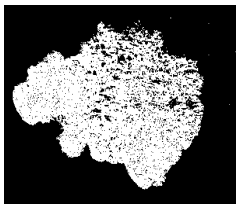

Fig.7: Segmented image using GRNN

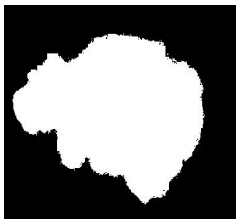

Fig.8: Ground truth image

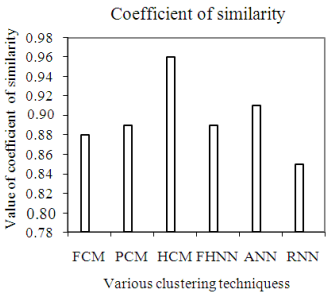

Fig. 9: Coefficient of Similarity for different clustering methods

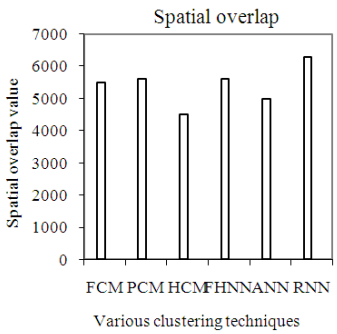

Fig.10: Spatial overlap for different clustering methods

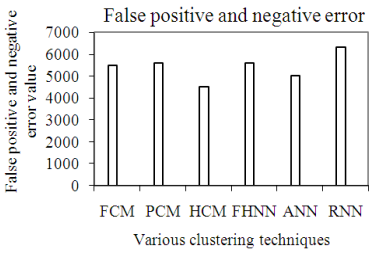

Fig. 11: False positive and negative error for different clustering methods

The adaline network is constructed and is trained with about 200 training samples and the corresponding training targets. The training samples are the pixels taken from the Melanoma digital image. The GRNN (Fig.7) is trained by the input samples and the corresponding target vectors that are taken from the malignant melanoma image.

In these cases, all the methods produce segmentation results which are close to the ground truth segmentation as shown in Fig.8. This happens when there is a good contrast between the lesion and the skin, thus the lesion boundaries are well defined. For obtaining the more accurate result in performance analysis, various parameters are to be considered. Fig. 9-11 shows the performance analysis using various parameters. The segmentation results were compared with the reference image (ground truth) and various parameters (Coefficient of Similarity, Spatial Overlap and False Positive and Negative Error) were evaluated and shown in Fig. 9-11.

The Coefficient of Similarity clearly shows the closeness or similarity of the segmented result with the Ground truth image. While comparing the coefficient of similarity of the results obtained from the 6 segmentation methods, the Hierarchical C Means algorithm performs better with a value of 0.9612 which is close to 1.

The best method according to the Spatial Overlap is the Hierarchical C Means algorithm with the value of 0.9421. It is more sensitive to small unmatched errors.

The best false positive and negative error (4459) is obtained by Hierarchical C Means algorithm. Among the two types of error, the false negative error is the worst type. Thus the Hierarchical C Means method was considered the most relevant metric from clinical point of view.

CONCLUSION

We proposed and evaluated six methods (intelligent fuzzy and neural networks based clustering techniques) for the segmentation of skin lesions in dermoscopic images. The various segmentation methods employed are Fuzzy C Means Algorithm (FCM), Possibilistic C Means Algorithm (PCM), Hierarchical C Means Algorithm (HCM), Fuzzy Hopfield Neural Network method (FHNN method), Regression Neural Network and Adaline Neural Network. The Segmentation methods was compared with the manually segmented image (the ground truth image) using various parameters such as Coefficient of Similarity, Spatial Overlap and False Positive and Negative Error to evaluate the performance of the proposed intelligent system based clustering techniques. The experimental results show that the Hierarchical C Means algorithm(Fuzzy) provides better segmentation than other (Fuzzy C Means ,Possibilistic C Means, Adaline Neural Network , FHNN and GRNN) clustering algorithms. Thus Hierarchical C Means approach can handle uncertainties that exist in the data efficiently and useful for the lesion segmentation in a computer aided diagnosis system to assist the clinical diagnosis of dermatologists.

REFERENCES

Bezdek, J.C. and S.K. Pal, 1992. Fuzzy Models for Pattern Recognition: Methods That Search for Structures in Data. 1st Edn., Institute of Electrical and Electronics Engineers, New York, ISBN: 0780304225, pp: 539.

Ganster, H., P. Pinz, R. Rohhrer, E. Wildling and M. Binder *et al*., 2001. Automated melanoma recognition. IEEE Trans. Med. Imag., 20: 233-239. DOI: 10.1109/42.918473

Jaya, J. and K. Thanushkodi, 2011. Implementation of Computer Aided Diagnosis System Based on Parallel Approach of Ant Based Medical Image Segmentation. J. Comput. Sci., 7: 291-297. DOI: 10.3844/jcssp.2011.291.297

Kapur, T., W.E.L. Grimson, W.M. Wells and R. Kinkinis, 1996. Segmentation of brain tissue from magnetic resonance images. Med. Image Anal., 1: 109-127. DOI: 10.1016/S1361-8415(96)80008-9

Lin, J.S., K.S. Cheng and C.W. Mao, 1996. A fuzzy Hopfield neural network for medical image segmentation. IEEE Trans. Nucl. Sci., 43: 2389-2398. DOI: 10.1109/23.531787

Roozbahani, R.G., M.H. Ghassemian, A.R. Sharafat, 2001. Robust segmentation of medical images using competitive Hopfield neural network as a clustering tool. Iranian J. Sci. Tech. Technol., 25: 427-439. Scavino, E., D.A. Wahab, H. Basri, M.M. Mustafa and A. Hussain, 2009. A Genetic algorithm for the segmentation of known touching objects. J. Comput. Sci., 5: 711-716. DOI: 10.3844/jcssp.2009.711.716

Sebri, A., J. Malek and R. Tourki, 2007. Automated Breast Cancer Diagnosis based on GVF-Snake Segmentation, Wavelet Features Extraction and Neural Network Classification . J. Comput. Sci., 3: 600-607. DOI: 10.3844/jcssp.2007.600.607

Silveira, M., J.C. Nascimento, J.C. Marques, A.R.S. Marcal and T. Mendonca *et al*., 2009. Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. IEEE J. Selected Topic Signal Proc., 3: 35-45. DOI: 10.1109/JSTSP.2008.2011119

Steck, J.E. and S.N. Balakrishnan, 1994. Use of hopfield neural networks in optimal guidance. IEEE Trans. Aerosp. Electron. Syst., 30: 287-293. DOI: 10.1109/7.250431

Sujaritha, M. and S. Annadurai, 2011. A new modified gaussian mixture model for color-texture segmentation. J. Comput. Sci., 7: 279-283. DOI: 10.3844/jcssp.2011.279.283

Zhang, W., 2006. System identification based on generalized adaline neural network. Int. J. Intell. Control Syst., 11: 17-25.

Zhang, X.B. and L. Jiang, 2009. An image segmentation algorithm based on fuzzy c-means clustering. Proceedings of the International Conference on Digital Image Processing, Mar. 7-9, IEEE Xplore Press, Bangkok, pp: 22-26. DOI: 10.1109/ICDIP.2009.15