The Application of a Simulated Annealing Fuzzy Clustering Algorithm for Cancer Diagnosis

Xiao Ying Wang, Glenn Whitwell and Jonathan M Garibaldi
Automated Scheduling, Optimisation and Planning (ASAP) Research Group
Department of Computer Science & Information Technology
The University of Nottingham, Jubilee Campus, Wollaton Road, UK
{xyw,jmg}@cs.nott.ac.uk

Abstract – **Fourier Transform Infrared Spectroscopy (FTIR) is becoming a powerful tool for use in the study of biomedical conditions, including cancer diagnosis. As part of an ongoing programme of research into the potential early diagnosis of cervical cancer, Hierarchical Cluster Analysis (HCA) and Fuzzy C-Means (FCM) have been applied to distinguish FTIR spectra obtained from cancerous and non-cancerous cells. In recent experimentation on non pre-processed FTIR spectra data, the FCM method has been shown to achieve significantly better results. Nevertheless, two limitations apply to this technique. Firstly, *a priori* assumption of the number of clusters is needed. This is a problem because, in general, this is not known in medical diagnosis. The other limitation is that it involves a greedy local search methodology such that sub-optimal solutions may be returned and, thus, misdiagnosis could occur. Bandyopadhyay [8] has recently proposed a Simulated Annealing Fuzzy Clustering algorithm (SAFC) which can avoid these two limitations. However, when we implemented the proposed algorithm, it was found that sub-optimal solutions could be obtained in certain circumstances. In this paper, we extend the SAFC algorithm to overcome this difficulty and apply this modified version to the classification of seven sets of FTIR spectra data which have been taken from three oral cancer patients. With no prior specification of cluster number, our modified SAFC algorithm is shown to obtain the correct (clinical) classification of clusters in 4 out of 7 data sets. In the remaining 3 data sets it produces a number of clusters which, while differing from the clinical classification, appear to better match the underlying data when subjectively visualised using Principal Component Analysis (PCA).**

I. INTRODUCTION

Cancers have become one of the major health problems that humankind endures and thus research into diagnosis methods and treatment has become an important research area for the scientific community. In Britain, more than one in three people will be diagnosed with cancer during their lifetime and one in four will die from cancer. If cancers can be diagnosed earlier then patients can begin treatment programmes quicker and, ultimately, a lower mortality rate can be achieved. FTIR is becoming a valuable tool for use in the study of a wide variety of biomedical conditions, including the diagnosis of cancer and other diseases [1-5]. It is based on the differential absorption of a range of wavelengths of Infrared Radiation (IR) dependent on the molecular structures and bonds of the substance(s) the IR passes through. Thus, this technology can detect changes in the chemical composition of cell nuclei that can indicate the early onset of cancer. Unfortunately, it has also been established that 'simple' measures, such as peak height(s) or peak location(s) are not sufficient to permit accurate diagnosis. As a precursor to a diagnostic system, an investigation was carried out in order to establish whether it is indeed possible to distinguish normal and abnormal cells from analysis of the entire spectra. If the characteristic spectra of abnormal and normal tissue components are known, it may be possible to compare the spectra in each cluster to these reference spectra and hence achieve accurate diagnosis.

In previous clinical work [6], FTIR spectral data was first empirically pre-processed, using such techniques as mean-centring and variance scaling, and then various data clustering techniques were compared to the manually obtained classifications. This showed that correct clustering could only be achieved by applying pre-processing techniques which were hand-tuned to the particular sample characteristics. Moreover, these pre-processing procedures need extra tools, time and expertise which are undesirable and, indeed, not practical in a clinical environment. Following this, a comparison was made of FCM and HCA in classifying non pre-processed FTIR oral cancer data [7], in which the FCM method was found to perform significantly better. However, in this study it was necessary to use the (previously established) actual number of clusters as an input parameter to FCM clustering, Obviously, in a fully automated diagnostic tool, such cluster numbers would not be know *a priori*. Efficient clustering techniques are required to automatically detect the number of disease categories or different levels of disease.

In this paper an algorithm that extends upon the ideas presented by Bandyopadhyay for fuzzy clustering [8] is developed that is able to automatically detect the number of clusters. A detailed discussion of the standard FCM algorithm and the proposed extensions is presented in section II and section III. Simulated annealing is combined with the conventional FCM method to guide a search through all cluster configurations. The Xie-Beni validity measure (XB) [9] is used to optimise the number of clusters and their locales. The results show that the proposed simulated annealing fuzzy clustering (SAFC) algorithm can produce better clustering than the standard FCM method as used in our previous work [7].

II. BACKGROUND

Fuzzy clustering has been widely used in medical diagnosis, pattern recognition, exploratory data analysis and image segmentation. The most prominent fuzzy clustering algorithm used is the Fuzzy C-Means (FCM), a fuzzification of K-Means or ISODATA. It was originally introduced by Bezdek in 1981 [10] as an extension to Dunn's algorithm [11]. Nevertheless there are limitations to the FCM technique. Firstly, it needs a priori assumption on the number of clusters which is not practical in real

medical diagnosis applications; secondly, it is a nonconvex method [12], so may often lead to local minimum solutions and hence misdiagnosis could occur. In order to overcome these drawbacks, the SAFC algorithm was proposed by Bandyopadhyay [8]. It avoids these limitations in the FCM technique and provides good quality solutions with accurate clusters.

A. Fuzzy C-Means

The aim of this clustering technique is to group a given set of data $X = \{x_1, x_2, ...x_n\}$ into a number of clusters c so that the data in the same group are as similar as possible and the data in different groups are as dissimilar as possible. FCM achieves this classification by minimising the objective function:

$$J(U,V) = \sum_{i=1}^{n}\sum_{j=1}^{c}(\mu_{ij})^m \| x_i - v_j \|^2 \qquad (1)$$

Where μ_{ij} represents the membership degree of data point x_i to the cluster centre v_j and μ_{ij} also satisfies the following conditions:

$$\mu_{ij} \in [0,1], \quad \forall i = 1,...n, \forall j = 1,...c \qquad (2)$$

$$\sum_{j=1}^{c}\mu_{ij} = 1, \quad \forall i = 1,...n \qquad (3)$$

The Euclidean distance between x_i and v_j is represented by $\| x_i - v_j \|$. The parameter m is used to control the fuzziness of membership of each datum, $m > 1$. There is no theoretical basis for the optimal selection of m, but a value of m = 2.0 is usually chosen. $U = (\mu_{ij})_{n*c}$ is a fuzzy partition matrix and $V = \{v_1, v_2, ...v_c\}$ is a set of cluster centres. FCM is described by the following steps [10]:

1) Initialize membership matrix μ_{ij} with random value, make sure it satisfies conditions **(2)** and **(3)**.

2) Compute the fuzzy centres v_j using

$$v_j = \frac{\sum_{i=1}^{n}(\mu_{ij})^m x_i}{\sum_{i=1}^{n}(\mu_{ij})^m}, \forall j = 1,...,c \qquad (4)$$

3) Calculate the new distance

$$d_{ij} = \| x_i - v_j \|, \forall i = 1,...,n, \forall j = 1,...c$$

4) Update the fuzzy membership μ_{ij}

$$\text{If } d_{ij} \neq 0 \qquad \mu_{ij} = \frac{1}{\sum_{k=1}^{c}(\frac{d_{ij}}{d_{ik}})^{\frac{2}{m-1}}} \qquad (5)$$

$$\text{Else} \qquad \mu_{ij} = 1$$

5) Repeat step 2) to 4) until the minimum J value is achieved.

B. Simulated Annealing

Simulated annealing is a Monte Carlo method to simulate the physical annealing of solids. It was originally proposed by Metropolis et al. in 1953 [13]. The process is as follows; an object is firstly heated from a high temperature then slowly cooled so that the system at any time is approximately in thermodynamic equilibrium. At equilibrium, the object would have many configurations which correspond to a specific energy level. A random perturbation from current configurations to next configuration is applied and the corresponding new energy level is obtained from current energy level. Let E_c and E_n represent the current energy level and new energy level. If $E_c > E_n$, then the new energy level is accepted, otherwise, the new energy level is accepted with probability $\exp(-(E_n - E_c)/T)$, where T is the temperature at that equilibrium. Therefore, the probability of accepting a worse energy level is greater when T is high and smaller when T is low. By gradually decreasing the temperature and repeating the Metropolis simulation, new energy levels will be achieved until no more improvements are possible.

The first attempt to bring simulated annealing into optimisation problems was by Kirkpatrick et al [14] who used simulated annealing as a new optimisation search paradigm to escape local optima and hopefully converge to the global optimum. Since this time, simulated annealing has been used on a wide range of combinatorial optimisation problems and achieved good results. Al-Sultan and Selim [15, 16] and Klein and Dubes [17] have developed algorithms based on simulated annealing to find the global minimum solution using FCM and other crisp (non-fuzzy) clustering methods. However the (fixed) number of clusters is a prerequisite parameter in both these techniques. Maulik and Bandyopadhyay [18] developed a fuzzy clustering technique which combines a genetic algorithm and FCM clustering to automatically segment satellite images obtained by remote sensing.

III. THE PROPOSED ALGORITHM

A. The Modified SAFC Algorithm

A good clustering has the properties that data points within the same cluster are compact and data points in different clusters are separated. Xie and Beni [9] created a fuzzy clustering criterion based on this idea, referred to as the XB validity measure. Let S represent the compactness and separation validity function. It can be defined as shown in the following equation:

$$S = \frac{\sum_{j=1}^{c}\sum_{i=1}^{n} \mu_{ij}^{\ 2} \| x_i - v_j \|^2}{n \min_{ij} \| v_i - v_j \|^2} \quad (6)$$

The parameters are the same as in section II A. In a simpler form, it can be expressed as: $S = \frac{\pi}{s}$, where π is the compactness of the fuzzy c-partition of the data set.

The $\pi = \frac{\sum_{j=1}^{c}\sum_{i=1}^{n} \mu_{ij}^{\ 2} \| x_i - v_j \|^2}{n}$ component of equation (6) measures the compactness of each class. The smaller π indicates more compact the clusters are. s is called the separation of the fuzzy c-partition of the data set. $s = (d_{\min})^2$, where $d_{\min} = \min_{ij} \| v_i - v_j \|$ is the minimum distance between cluster centres. If s is bigger, then the clusters are separated. Therefore, a smaller S indicates all of the clusters are more compact and with greater separation from each other.

The corresponding process of simulated annealing and SAFC is shown in Table I. This table shows the XB validity function value as the energy associated with a configuration. In other words, it is used as an evaluation function to measure how the FTIR oral cancer data sets are classified. There are two reasons for choosing the XB validity measure as the evaluation function. Firstly, the FCM algorithm needs *a priori* assumption of number of clusters. However, in the SAFC algorithm the number of clusters is unknown, thus the original evaluation function of FCM is not suitable. Secondly, of the existing validity measures available for fuzzy clustering, the XB validity measure has been shown to be able to detect the correct number of clusters in several experiments [19].

The most interesting part of the proposed SAFC algorithm is how to perturb the current centres (current configuration). Three perturbation functions are randomly chosen on each perturbation. They are: perturbing an existing centre (*Perturb Centre*), splitting an existing centre (*Split Centre*) and deleting an existing centre (*Delete Centre*). At the beginning of each perturbation function, an existing centre is chosen based on the size of the cluster. Let $|C_j|$ represent the size of cluster j; it is expressed by:

$$|C_j| = \sum_{i=1}^{n} \mu_{ij}, \ \forall j = 1,...c \quad (7)$$

where c is the number of clusters.

The three functions are described below and are followed by a summary of the whole SAFC process.

1) Perturb Centre

The smallest cluster will be chosen computed by computing the cluster sizes using equation (7) in the set of centres. This centre position is then modified through addition of the change rate $cr = perturb\_rate * rand$, where $perturb\_rate \in$ [-0.007, 0.007] and $rand \in$ [0, 1]. Let $v_{current}[d]$ and $v_{new}[d]$ represent current centre and new centre respectively, and $d = 1,...N$, where N is the number of dimensions. *Perturb Centre* can then be expressed as: $v_{new}[d] = v_{current}[d] + cr$. The parameter *perturb_rate* was set through initial experimentation.

2) Split Centre

As opposed to *Perturb Centre*, *Split Centre* will pick up the centre of the biggest cluster and split it into two centres. This is performed by choosing a data point from the data set as a reference point, expressed as $w_{chosen}[d]$. There are two kinds of situation that must be taken into account when choosing the reference data point. The first exists when the boundary between the biggest cluster and the other clusters are not obvious, that is to say there are some data points whose membership degree to the chosen centre is close to 0.5. In this case, the data point which has membership degree that is smaller than but closest to 0.5 is chosen as the reference point. The second situation occurs when the biggest cluster is clearly separate from the other clusters. This means that there exist some data points which have a proportionally large membership degree to the biggest cluster centre (at least greater than 0.5) and all of the other data points have membership degrees that are 'small'. In our experimentation, we define 'small' as a membership degree which is less than 0.4. In this case, the mean value of all the membership degrees above 0.5 is calculated and the data point that has a membership degree closest to the mean is chosen as the reference point. The distance between this chosen data point and current chosen centre is $dist[d] = | v_{current}[d] - w_{chosen}[d] |$. The two new centres are then generated by using $v_{new}[d] = v_{current}[d] \pm dist[d]$. Finally, $v_{current}[d]$ is deleted.

3) Delete Centre

As with *Perturb Centre*, the smallest cluster will be chosen but this time is deleted from the current cluster set.

4) SAFC process summary:

i. Randomly initialise the number of clusters to num, where $2 \le num \le \sqrt{n}$ and n is number of data points.
ii. Initialise fuzzy membership μ_{ij}, subject to conditions (2) and (3) being satisfied.
iii. Calculate data centres using (4).
iv. Calculate current XB value using (6).
v. $T = T_{\max}$, setting maximum temperature as starting temperature.
vi. If $T \ge T_{\min}$, i = 1 to k
vii. Alter centres by randomly calling one of *Peturb Centre*, *Split Centre* or *Delete Centre*.
viii. Calculate new XB value using (6) based on the perturbation.
ix. If new XB < current XB, accept new centre. Set new XB to current XB, save best XB and best centres position.

Table I
The corresponding process of simulated annealing and SAFC

Simulated Annealing	SAFC
1. The solid undergoing annealing	1. Optimize the FTIR oral cancer data clustering
2. Current configuration	2. Current data set centres
3. Perturbation of current configuration	3. Perturbation of a centre in the current solution
4. The next configuration	4. New solution after perturbation
5. Calculate energy associated with the new configuration	5. Calculate XB validity function value of new solution
6. Accepting the next configuration if its energy level is less than the current energy level	6. Accepting the new solution if the new XB validity value is less than the current XB value
7. Accepting the next configuration with some probability	7. Accepting the new solution with high XB value, if acceptance probability is satisfied (moving to worse solution may escape from local optima).

x. If new XB > current XB, accept new centre with a probability given by the expression: $\exp(-(XB_{new} - XB_{current})/T)$.
xi. $T = rT$, $0 < r < 1$.
xii. $i = i + 1$. If $i > k$, stop, Else go to vi.
xiii. Return the best XB and best centre positions.

In above steps, T represent current temperature; T_{max} and T_{min} are maximum and minimum temperature respectively, i is iteration counter, k is the number of iterations at each temperature and r is the cooling rate that is used to decrease the temperature.

B. Extensions from the Original SAFC Algorithm

When the original SAFC algorithm was implemented by ourselves on a wider set of test cases than Bandyopadhyay originally used, it was found to suffer from several difficulties. In order to overcome these difficulties, three extensions to the algorithm originally presented in [8] were introduced (and are included above). These difficulties and the extensions introduced are now summarized.

The first extension is in part *A 1) Perturb Centre*. The original algorithm picked a random centre to perturb. If a random centre is chosen, it is possible that even a good centre is altered. In contrast, by choosing the weakest (smallest) centre, we avoid destabilising an already good (large) centre. Ultimately, this can lead to a quicker and more productive search.

The second extension is in part *A 2) Split Centre*. In the original algorithm, all cases were handled as described in situation one (choosing the data point that has the membership degree that is closest to but less than 0.5). There is a drawback to this approach when splitting a centre where clusters are separate and distinct. For example, let there be two clusters in a set of data points which are separated, with a clear boundary between them. v_1 and v_2 are the corresponding cluster centres at a specific time in the search as shown in Fig. 1 (shown in two-dimensions). The biggest cluster is chosen, say v_1. Then a data point whose membership degree is closest to but less than 0.5 can only be chosen from the data points that belong to v_2 (where the data points have membership degrees less than 0.5 to v_1). So, for example, the data point w_1 (which is closest to v_1) is chosen as the reference data point. The new centres will then move to v_{new1} and v_{new2}.

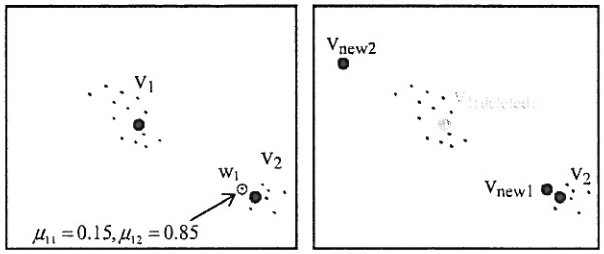

Fig. 1. An illustration of *Split Centre* from original algorithm with distinct clusters (where μ_{11} and μ_{12} represent the membership degree of w_1 to the centres v_1 and v_2 respectively)

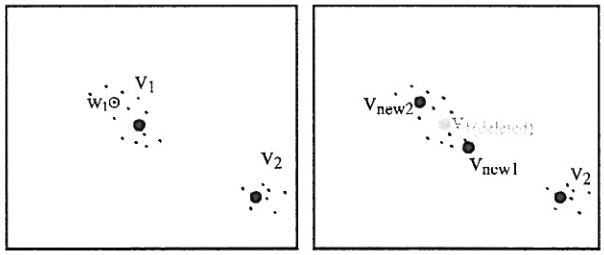

Fig. 2. The new *Split Centre* applied to the same data set as Fig. 1 (where w_1 is now the data point that is closest to the mean value of the membership degree above 0.5)

Obviously these centres are far from the good solution. Although the new centres would be changed by the *Perturb Centre* function afterwards, it will inevitably take a longer time to 'repair' the solutions. In our approach, two new centres are created within the biggest cluster; the same dataset as in Fig. 1 is used to illustrate this process. A data point is chosen, w_1, that is closest the mean value of the membership degree above 0.5. Then two new centres v_{new1} and v_{new2} are created according the distance between v_1 and w_1. This is shown in Fig. 2 above. Obviously the new centres are better than the ones in Fig. 1 and therefore less time will be required to find the better centre positions.

The third extension is in part *A 4)*ix. In the original algorithm, the current solution was returned as the final solution. This causes a problem if, for example, the last accepted solution was a worse solution that had been previously found. In the modified algorithm, the best centre positions (with the best XB) that have been encountered are now saved. At the end of the search, rather than returning the current solution, the best solution seen throughout the *whole* duration of the search is returned.

IV. RESULTS

Seven sets of FTIR spectra data were used during experimentation. All of the FTIR spectra data have been produced by FTIR spectroscopy and contain a mixture of tumour, stroma, early keratinisation and necrotic specimens from three oral cancer patients. These data have been provided by Derby City General Hospital, UK. Infometrix Pirouette multivariate analysis software was used to perform the FTIR data analysis. The frequency of the spectra data is limited within 900-1800 cm^{-1}. Each set of FTIR data was obtained by scanned pieces of oral tissue from patients. Data set 1 to data set 4 are tissue sections from first patient, data set 5 and data set 6 get from second patient, data set 7 is from third patient. Fig. 3 shows the tissue section that was scanned for data set 1 and Fig. 4 shows its FTIR spectra. Fig. 3a shows the stained oral tissue section which would be used for diagnosis under a microscope by a medical technician. The tumour and stroma regions are clearly discriminated by their light and dark coloured stains respectively. Fig. 3b is from a portion of a parallel, unstained section. The superimposed dashed white lines separate the different morphologies. Fifteen single points were chosen to scan on FTIR spectroscopy in this tissue section and these are also shown in Fig. 3b.

At this stage of the study, it is the spectral characteristics of essentially distinct classes of tissue cells that are of interest, rather than gradation processes or mixed types. Therefore some data points on the boundary region are excluded from data set 4 and data set 7. The number of data points from data set 2 to data set 7 is: 18, 11, 31, 30, 15 and 42 respectively. The parameters of the SAFC algorithm were determined through initial experimentation and are as follows:

$$T_{max} = 100,\ T_{min} = 10^{-5},\ k = 40,\ r = 0.9$$

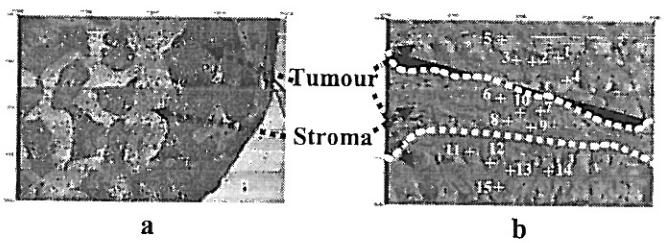

Fig. 3. Tissue samples from data set 1

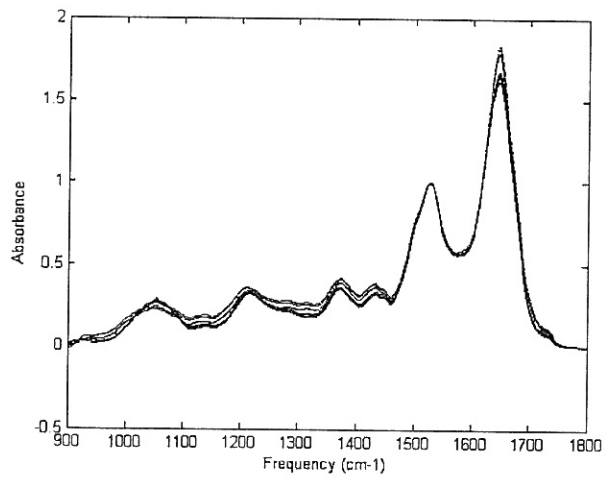

Fig. 4. FTIR spectrum of data set 1

TABLE II

COMPARISON BETWEEN CLINICAL CLUSTER NUMBER AND SAFC CLUSTER NUMBER

Data set names	Number of clinical clusters	Number of SAFC clusters
Data set 1	2	2
Data set 2	2	2
Data set 3	2	3
Data set 4	3	2
Data set 5	2	3
Data set 6	2	2
Data set 7	3	3

TABLE III

EVALUATION VALUE COMPARISON BETWEEN FCM CLUSTERING ALGORITHM AND SAFC CLUSTERING ALGORITHM

Data set names	*FCM*	*SAFC*
Data set 1	0.048038	0.047835
Data set 2	0.078894	0.078698
Data set 3	0.291327	0.077379
Data set 4	0.449369	0.046062
Data set 5	0.296051	0.211870
Data set 6	0.071461	0.070518
Data set 7	0.140332	0.134

TABLE II shows the comparison between clinical cluster number and SAFC cluster number. TABLE III shows a comparison of the evaluation values between the FCM and the modified SAFC method. The FCM technique was specified to find a solution with the same number of clusters as determined by clinical analysis (provided by Derby General Hospital, UK).

From the results, it can be observed that SAFC determines the same number of clusters as the clinical identification in 4 out of the 7 data sets. It can also be noted that the cluster quality obtained is better than that achieved by standard FCM clustering on all data sets. This indicates that the SAFC is able to escape the local optima which occur in FCM clustering and is, therefore, able to find improved solutions. The ability of the SAFC algorithm to accurately detect the classes present in previously unseen data is a valuable first step towards a fully automated diagnostic system.

However, the SAFC algorithm gets different numbers of clusters in data sets 3, 4 and 5 as compared to the number of clusters from clinical analysis. The cluster validity measure obtained by SAFC in these three data sets is much smaller than those obtained using standard FCM (with the clinical number of clusters). As mentioned in Section III A, the smaller the evaluation functions, the more compact are the clusters. These findings were examined further by plotting the data by principal components obtained through PCA. Space limitations preclude the inclusion of such plots here. However, inspection of such plots indicates that the number of clusters obtained by SAFC look more compact than those obtained by FCM. Although by no means definitive this may indicate that the clinical numbers of clusters are incorrect.

V. DISCUSSION AND CONCLUSION

In this paper, three modifications have been proposed to the simulated annealing method of fuzzy clustering of Bandyopadhyay [8]. This modified SAFC algorithm has been applied to the analysis of seven real world Fourier Transform Infrared Spectroscopy (FTIR) spectra gathered from three oral cancer patients using. On all seven datasets the modified SAFC algorithm obtains better solutions (as assessed by the Xie-Beni validity measure [9]) than the standard Fuzzy C-Means algorithm. Although this new approach achieved better cluster validity measures, there are sometimes discrepancies with respect to the number of clusters that were specified in the clinical classification. It is possible that the clinically provided number of clusters is not correct. There may be tissue samples of different types present which were not noticed by the clinical observers. More results are needed before any definitive conclusion can be made on this matter. However, this method can avoid the problem of local optima and pre-defined number of clusters that occur in the standard Fuzzy C-Means algorithm.

Further development of the SAFC algorithm is an ongoing research area. At the same time, we are trying to obtain a wider source of sample data with which the classification is known from a number of clinical domains such as oral cancer and cervical smear test screening. Establishing the techniques necessary to develop clinically useful tools in number of domains is the ultimate goal of this research.

VI. ACKNOWLEDGMENT

The authors are grateful to John Chalmers et al for providing the FTIR spectral data used in this study, and for making available their internal report on the clinical study carried out at Derby General City Hospital. The authors would like to express their kind thanks to Sanghamitra Bandyopadhyay for her permission to use and extend the algorithm. Finally, we also thank the anonymous reviewers for their valuable comments and suggestions which have resulted in an improved paper.

VI. REFERENCES

[1] P.T.T. Wong, B. Rigas, Infrared spectra of microtome sections of human colon tissues. *Applied Spectroscopy*, Vol. 44, pp. 1715–1718, 1990.

[2] B. Rigas, S. Morgello, I.S. Goldan, P.T.T. Wong, Human colorectal cancers display abnormal Fourier-transform infrared spectra. *Proceedings of the National Academia of Science USA*, Vol. 87, No. 20, pp. 84–88, 1990.

[3] P.T.T Wong, S.M. Goldstein, R.C. Grekin, T.A. Godwin, C. Pivik, B. Rigas, Distinct infrared spectroscopic patterns of human basal cell carcinoma of the skin. *Cancer Research*, Vol. 53, No. 4, pp. 762–765, 1993.

[4] B.J. Morris, C. Lee, B.N. Nightingale, E. Molodysky, L.J. Morris, R. Appio, Fourier transform infrared spectroscopy of dysplastic, papillomavirus-positive cervicovaginal lavage speciens. *Gynecological Oncology*, Vol. 56, No. 2, pp. 245–249, 1995.

[5] D.C. Malins, N.L. Polissar, K. Nishikida, E.H. Holmes, H.S. Gardner, S.J. Gunselman, The etiology and prediction of breast cancer. Fourier transform-infrared spectroscopy reveals progressive alterations in breast DNA leading to a cancer-like phenotype in a high proportion of normal women. *Cancer*, Vol. 75, No. 2, pp. 503–517, 1995.

[6] R. Allibone, J. M. Chalmers, M. A. Chesters, S. Fisher, A. Hitchcock, M. Pearson, F. J. M. Rutten, I. Symonds, M. Tobin, "FT-IR microscopy of oral and cervical tissue samples", *Internal Report,* Derby City Gerneral Hospital, England, 2002.

[7] X. Y. Wang, J. Garibaldi, T. Ozen, "Application of the Fuzzy C-Means Clustering Method on the Analysis of non Pre-processed FTIR Data for Cancer Diagnosis", Macquarie University, Sydney, Australia, pp. 233-238, 2003.

[8] S. Bandyopadhyay, "Simulated Annealing for Fuzzy Clustering: Variable Representation, Evolution of the Number of Clusters and Remote Sensing Application", Machine Intelligence Unit, Indian Statistical Institute. 2003. (unpublished personal communication).

[9] X. L. Xie and G. Beni, " A validity measure for fuzzy clustering", *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 13, pp. 841-847, 1991.

[10] J. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. New York: Plenum, 1981.

[11] J. Dunn, "A fuzzy relative of the isodata process and its use in detecting compact, well-separated clusters", *J.Cybernet.,* vol. 3, no. 3, pp. 32-57, 1974.

[12] S. Z. Selim, "Using nonconvex programming techniques in cluster analysis", *Jt Meet. Ops Res. Soc. Am. And Inst. Mgmt Sci.,* Houston, Texas, 1981.

[13] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller, "Equation of State Calculations by Fast Computing Machines", *J. Chem. Phys.*, vol. 21, no. 6, pp. 1087-1092, 1953.

[14] S. Kirkpatrick, C. D. Gelatt Jr, M. P. Vecchi, "Optimization by Simulated Annealing", *Science*, 220, 4598, pp. 671-680, 1983.

[15] K. S. Al-Sultan, S. Z. Selim, "A Global Algorithm for the Fuzzy Clustering Problem", *Pattern Recognition*, Vol. 26, No. 9, pp. 1357-1361, 1993.

[16] S. Z. Selim, K. Alsultan, "A Simulated Annealing Algorithm for the Clustering Problem", *Pattern Recognition*, Vol 24. No. 10, pp. 1003-1008, 1991.

[17] R.W. Klein, R. C. Dubes, "Experiments in Projection and Clustering by Simulated Annealing", *Pattern Recognition*, Vol 22. No. 2. pp. 213-220, 1989.

[18] U.Maulik, S. Bandyopadhyay, "Fuzzy Partitioning Using a Real-coded Variable Length Genetic Algorithm for Pixel Classification", *IEEE Trans. On Geosciences and Remote Sensing*, Vol. 41, No. 5, 5 May 2003.

[19] N. R. Pal, J. C. Bezdek, "On cluster validity for the fuzzy c-means model", *IEEE Trans. Fuzzy Systs.*, Vol. 3, pp. 370-379, 1995.