A COMPARISON OF FUZZY AND NON-FUZZY CLUSTERING TECHNIQUES IN CANCER DIAGNOSIS

X. Y. Wang **J. M. Garibaldi**
The University of Nottingham

ABSTRACT

In this paper, we apply K-means and Fuzzy C-Means, two widely used clustering algorithms, to cluster a lymph node tissue section which had been diagnosed with metastatic infiltration (cancer spread from its original location). Each cluster algorithm is run 10 times as different initialisation states may lead to different clustering results. We compare the performance of the two algorithms by subjectively altering the number of clusters from 2 to 9 and we analyse the results using false-colour images (which are produced as a function of the spatial coordinates on the tissue section).

In the initial stages of this experiment, we found that the ranges of the first three principal components were too small and may lead to small objective function values in Fuzzy C-Means. Therefore, the minimal amount of improvement must be set to a small enough value to allow the cluster centre positions to improve, otherwise the iteration will stop prematurely. After adjusting this setting, the performance of Fuzzy C-Means was significantly better. The results show that Fuzzy C-Means can separate the major different tissue types using just a small number of clusters, whereas K-means is only able to separate them if a larger cluster number is used.

Keywords: K-means, Fuzzy C-Means, principal components.

1. INTRODUCTION

Breast cancer is the most common form of cancer found among women. In the United Kingdom alone, around 13,000 deaths are reported each year from the disease, and approximately 40,000 per year are recorded in the United States of America [1]. If the cancer can be detected early, the options for treatment and thus chance of total recovery are increased. Intra-operative diagnosis of the disease has steadily become more important with respect to the recent introduction of sentinel lymph node biopsy. A sentinel lymph node is classed as any node that has a direct lymphatic connection to the cancer, and would therefore be the most likely location of cancer spreading from the breast. In surgical studies it has also been indicated that the chance of finding the disease further down the chain of lymph nodes that drain the breast, is significantly smaller if the cancer cannot be found in sentinel lymph node [2].

Currently, histological techniques used for intra-operative sentinel lymph node diagnosis are not consistently available and reliable. Therefore, various spectroscopic methods have been proposed and are being investigated to assist in providing a reliable diagnostic tool [3-6]. Although the techniques applied have showed promise, all display limitations in their diagnostic ability.

Fourier Transform Infrared (FTIR) Microspectroscopy has increasingly been applied to problems in the biomedical field. The technique is based upon detecting changes in the cellular composition of tissue molecules, by analysing the vibrational and rotational modes of biochemical bonds and atomic nuclei. Such changes can reflect the onset of a disease, and thus a diagnostic model could be produced. More technical details regarding FTIR Microspectroscopy can be found in [7]. The spectroscopic technique has been used to examine several different types of disease, such as lung [8], skin [9], colon [10], breast [11] and prostate cancers [12]. However, FTIR Microspectroscopy has yet to be utilised in the assessment of cancers that specifically spread from the breast to lymph nodes.

Clustering is a multivariate analysis technique widely adopted in medical diagnosis studies and pattern recognition areas. By examining the underlying structure of a dataset, cluster analysis aims to class data (IR spectra in this study) into separate groups according to their characteristics. The clustering is performed such that spectra held within a cluster are as similar as possible, and those found in opposing clusters as dissimilar as possible. Thus different cells types found within biological tissue can be separated and characterised. In previous work, three different clustering algorithms, Fuzzy C-Means (FCM) Clustering, Hierarchical Clustering Analysis (HCA), and Simulated Annealing Fuzzy Clustering (SAFC), were investigated using data sets that comprised of FTIR spectra collected from oral cancers [7,13,14]. The experimental results indicated that FCM clustering performed significantly better than HCA when classifying spectra into their specific diagnosis [14]. It was also shown that when FCM clustering was combined with simulated annealing, the algorithm was able to automatically obtain the optimal number of clusters with respect to the Xie-Beni cluster validity measure [7,13]. However, although results were exceptionally promising, the seven data sets examined were relatively small in size, varying between 11 and 42 IR spectra in each.

In this paper, FCM (fuzzy) and K-means (non-fuzzy) clustering techniques are used to cluster IR spectra that were collected from a large area of an axillary lymph node tissue section. The map or IR image created was

composed of 7497 spectra, a significantly higher number than in previous studies, and proves a more stern test in assessing clustering techniques diagnostic ability. The setting for the numbers of clusters used in each algorithm was subjectively set from 2 to 9.

2. MATERIAL AND METHODS

2.1 FTIR Data sets

With approval from the Gloucestershire Research Ethics Committee (UK) and fully informed consenting patients, lymph node tissue specimens were collected during routine breast cancer surgery for use in this project. Tissue specimens were cut using a microtome to providing a 7μm thick tissue section that were fixed onto barium fluoride disc ideally suited for transmission FTIR Microspectroscopy. To allow histological diagnosis of the sample, the remaining specimens were formalin-fixed and wax embedded as done conventionally. An adjacent tissue section was then cut and further Hematoxylin and Eosin (H&E) stained for comparative analysis by a consultant breast histopathologist. For the purpose of this paper, we have focused on one such tissue section analysed by FTIR, diagnosed with metastatic infiltration (cancer spread from its original location), which we have named LNII5. By use of the commercially available Perkin Elmer Spotlight Imager®, an IR spectroscopic map or image was collected from lymph node LNII5, encompassing an area with several different tissue types, including both normal and cancerous cells. The image collected is made up of 7497 transmission spectra in total, expressing absorbance values from 821 wave numbers, in the range of 4000cm^{-1} to 720cm^{-1}. All transmission spectra were recorded using a spatial resolution of $6.25\mu m \times 6.25\mu m$, an instrumental resolution of $8cm^{-1}$, and the accumulation of 16 interferograms per spectrum.

2.2 K-means algorithm

K-means algorithm was originally introduced by McQueen in 1967 [15]. It is a non-fuzzy clustering method whereby each pattern can only belong to one centre at any one time. Let $X = \{x_1, x_2, ...x_n\}$ represent a set of data, where n is the number of data points. $V = \{v_1, v_2, ...v_c\}$ is the corresponding set of centres, where c is the number of clusters. The aim of K-means algorithm is to minimize the objective function *J (V),* in this case a squared error function:

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} \| x_{ij} - v_j \|^2 \qquad \textbf{(1)}$$

where $\| x_{ij} - v_j \|$ is the Euclidean distance between x_{ij} and v_j. c_i is the number of data points in the cluster *i* . The *ith* centre v_i can be calculated as:

$$v_i = \frac{1}{c_i} \sum_{j=1}^{c_i} x_{ij}, \qquad i=1...c \qquad \textbf{(2)}$$

The procedure of this algorithm can be described as followed:

1) Randomly select *c* cluster centres.
2) Calculate the distance between all of the data points and each centre.
3) Data is assigned to a cluster based on the minimum distance.
4) Recalculate the centre positions using (2).
5) Recalculate the distance between each data point and each centre.
6) If no data was reassigned, then stop, otherwise repeat 3).

2.3 Fuzzy C-Means algorithm

The FCM algorithm, also known as Fuzzy ISODATA, is one of the most frequently used methods in pattern recognition. It is based on the minimization of the objective function (3) to achieve a good classification. *J(U,V)* is a squared error clustering criterion, and solutions of minimization of (3) are least-squared error stationary points of *J(U,V)*.

$$J(U,V) = \sum_{i=1}^{n} \sum_{j=1}^{c} (\mu_{ij})^m \| x_i - v_j \|^2 \qquad \textbf{(3)}$$

Once again, the expression $X = \{x_1, x_2, ...x_n\}$ is a collection of data, where *n* is the number of data points and $V = \{v_1, v_2, ...v_c\}$ is the set of corresponding cluster centres in the data set ***X***, where *c* is the number of clusters. μ_{ij} is the membership degree of data x_i to the cluster centre v_j. Meanwhile, μ_{ij} has to satisfy the following conditions:

$$\mu_{ij} \in [0,1], \qquad \forall i = 1,...n, \forall j = 1,...c \qquad \textbf{(4)}$$

$$\sum_{j=1}^{c} \mu_{ij} = 1, \qquad \forall i = 1,...n \qquad \textbf{(5)}$$

$U = (\mu_{ij})_{n*c}$ is a fuzzy partition matrix. $\| x_i - v_j \|$ means the Euclidean distance between x_i and v_j. Parameter *m* is called the "fuzziness index", it is used to control the fuzziness of membership of each datum. The value of *m* should be within the range $m \in [1, \infty]$. There is no theoretical basis for the optimal selection of

m, but value of $m = 2.0$ is usually chosen. The FCM algorithm can be performed by following steps (see [16]):

1) Initialize the cluster centres $V = \{v_1, v_2, ...v_c\}$, or initialize the membership matrix μ_{ij} with random value, and make sure it satisfies conditions **(4)** and **(5)**, then calculate the centres .

2) Calculate the fuzzy membership μ_{ij} using

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{2}{m-1}}} \qquad (6)$$

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

3) Compute the fuzzy centers v_j using

$$v_j = \frac{\sum_{i=1}^{n} (\mu_{ij})^m x_i}{\sum_{i=1}^{n} (\mu_{ij})^m}, \forall j = 1,...,c \qquad (7)$$

4) Repeat step 2) to 3) until the minimum J value is achieved.

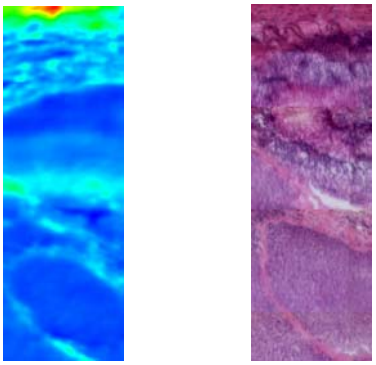
In this paper, both K-means and FCM clustering algorithms were implemented using MATLAB (version 6.5.0, release 13.0.1).

3. RESULTS AND DISCUSSION

The data sets that have been used in our previous work (seven sets of oral cancer IR spectra) are comparatively small, the number of spectra is: 15, 18, 11, 31, 30, 15 and 42 in data set 1 to data set 7 respectively. 901 wave number absorbencies were recorded from each spectrum. In each data set, some scattered individual points from different types of tissue were chosen for the FTIR scan. In contrast, the experiments reported in this paper are using 7497 spectra which have been taken from a whole sub area of one axillary lymph node (LNII5). Each spectrum contains 821 wave number absorbencies. It is apparent that if we can reduce the dimensionality from the original data without losing too much useful information, the performance of K-means and FCM clustering algorithm will be computationally efficient. In respect of this, principal component analysis was used to adjust the coordinates of the original data [17,18]. In this experiment, the data dimensions of the first 10 principal components were selected on which to do the data analysis. These principal components were found to contain 99.08% of the variances from the original data and, thus, the principal components are still highly reflective of the original data.

The number of clusters is subjectively set from 2 to 9. This is because different initialisation states often lead to different clustering results, both the K-means and FCM experiments were each run 10 times. For K-means, the squared Euclidean distance was used as the distance measure; the initial cluster centres positions were randomly selected. The maximum number of iterations was set to 100. In FCM, the fuzziness index m was set to a value of 2; the maximum number of iterations was also set to 100. The minimal amount of improvement was firstly set 10^{-5} (the stopping criterion of the iteration).

The K-means and FCM clustering algorithms were applied on tissue section LNII5. This is a mixed type of tissue section, which mainly contains cancer, normal, capsule (tissue with surrounds and protects lymph node) and reticulum (fibrous tissue which is major constituent of support structure). As mentioned in section 2.1, two adjacent layers of tissue were cut for IR scan and H&E stained individually, so it is possible that slight differences exist between the two sections. The spectra image of LNII5 is shown in Fig. 1a and the stained pictures in Fig. 1b.



(a) **(b)**

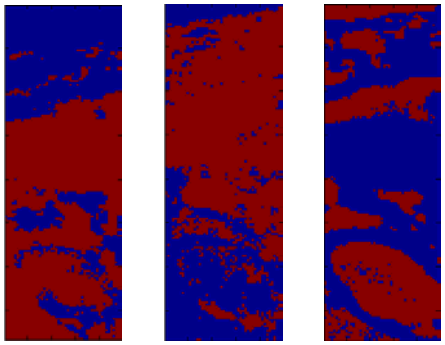
Fig.1 LNII5 in (a) spectra image (b) stained pictures.

In top left corner of Fig. 1a, there are some fat tissues, however, in Fig. 1b the stained piece was cut slightly below the fat and, therefore, we can see the scale in the two pictures are not exactly the same.

Since each spectrum on the IR image has a unique spatial x, y position, false-color images can be generated by plotting specially colored pixels as a function of the spatial coordinate. In the results, in order to distinguish between different clusters, each cluster was assigned a unique colour.

During the initial stages of the experiments, the FCM produced considerably varied results on each run. Fig. 2. shows examples from 3 separate runs (the number of clusters was 2). In these examples, the FCM clustering did not perform well. The main types of tissue were mixed throughout the section. In each color region, it could not clearly separate individual or groups of several main types of tissue. For example, in Fig. 2a,

the red region included cancerous and normal node, reticulum, and capsule; blue area included fatty capsule and all types of tissue covered in red region. The situation is similar happened in Fig. 2b and Fig. 2c.



(a) **(b)** **(c)**

Fig. 2 Results from three separate runs with FCM.

Based on this observation, when the spectra data were plotted in first three principal components in which contained 93.08% variances from the original data, we found that the ranges of the components are:

[-0.0075, 0.0751],
[-0.0117, 0.0069]
and [-0.0096, 0.0047] respectively.

The order of -.0.0075 is -10^{-2} and 0.0751's is 10^{-1} and so on. Thus their corresponding sizes of the range are 10^{-1}, 10^{-2} and 10^{-2}. As we know, the smaller the size of the component range, the more compact the data, thus the distances between the data and their ideal centres are smaller. In FCM, the objective function *J(U,V)* (see equation (3)) is proportional to the squared Euclidean distance between the data and the centres. In this case, the range sizes are 10^{-1} and 10^{-2}, their squared Euclidean distances are then 10^{-2} and 10^{-4} (even smaller). Hence a small range size may lead to a very small objective function value. One of the iteration stopping criterions occurs when the difference between two objectives functions values is less than the minimal amount of improvements. In such instances the iteration stops. Therefore, if the minimal amount of improvement was not small enough (i.e. 10^{-5} as initial setting) to allow improvements in the centre positions, the performance of FCM is bad. Due to this finding, we used 10^{-7} as the minimal amount of improvement for the remainder of the experiments. It was found that the performance of FCM improved and always achieved stable clustering results.

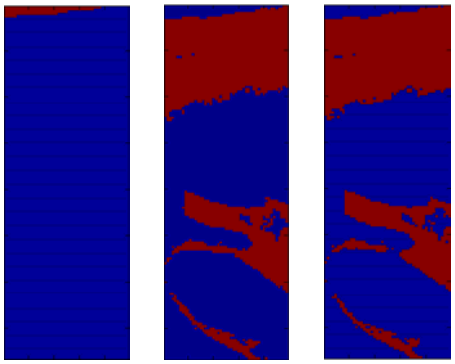
This finding also can be demonstrated in the seven oral cancer FTIR spectra data sets which were used in our previous experiments (see [7,13,14]). The minimal amount of improvement was set 10^{-5}. In these data sets, FCM performed well because the range sizes were compatible. The range sizes in the first three principal components (pc) are shown in Table 1.

Table 1. First three principal components size of ranges in seven oral cancer FTIR data sets.

Data sets	Variances ranges of		
	First pc	Second pc	Third pc
Data set 1	1.8458	0.6141	0.2575
Data set 2	2.8569	1.2562	0.4795
Data set 3	1.5960	0.5859	0.6600
Data set 4	2.1702	0.8569	0.4112
Data set 5	1.6224	0.8317	0.5496
Data set 6	1.7023	0.8902	0.3342
Data set 7	1.8750	1.5900	0.8741

We can see from Table 1 that the order of the first principal component range is 10^0 for all of the datasets (the squared Euclidean distance is therefore also of order 10^0). In the second and third principal components, the range sizes are either 10^0 or 10^{-1}, so the corresponding squared Euclidean distance are 10^0 and 10^{-2} respectively. Compared to these, a value of 10^{-5} for the minimal amount of improvement is sufficiently small to allow the centre positions to improve. This is also the reason why FCM obtained good results for these data sets.

For the remainder of the experiments, FCM produced consistent clustering results. K-means also produced fairly stable results except for the case where there are 2 clusters. Fig. 3 shows when number of cluster is 2 the clustering plot results which obtained from K-means and FCM.

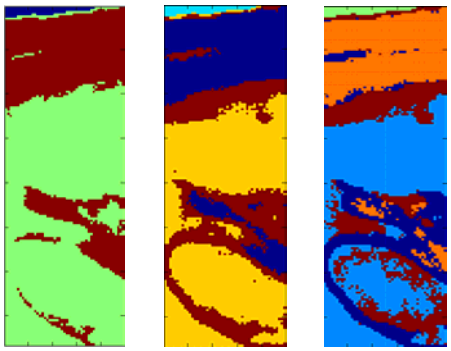


(a) **(b)** **(c)**

Fig. 3. Clustering results from K-means (a & b) and FCM (c) in 2 clusters

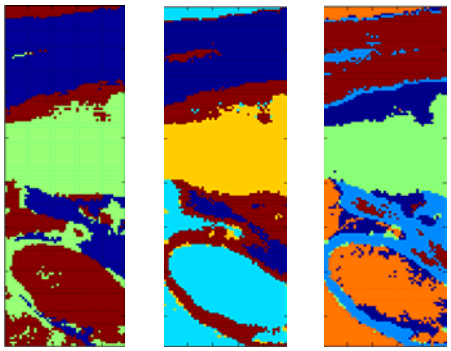
Fig 3a and Fig.3b are two types of results from K-means; Fig.3c is from FCM. In Fig. 3a, K-means separated the fat with other types of tissue, Fig. 3b, and Fig. 3c are almost the same, the red region covered capsule and reticulum nodes. The blue area includes the fatty capsule and nodal tissue (including both cancerous and normal). The variation in results obtained using the K-means algorithm can clearly be seen from Fig. 3a and Fig. 3b. This is because the K-means algorithm is very sensitive to the outliers, if one data point is assigned to one cluster rather than another, the results may substantially distort the distribution of the data.

Fig 4 shows the results when number of clusters is set to 3, 4 and 5 using K-means clustering. Fig 5 shows the result of applying the FCM algorithm using the same number of clusters.



(a) 3 clusters **(b)** 4 clusters **(c)** 5 clusters

Fig. 4. K-means clustering results in 3, 4 and 5 clusters.



(a) 3 clusters **(b)** 4 clusters **(c)** 5 clusters

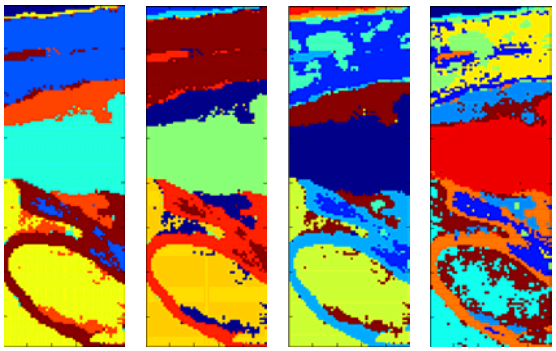
Fig. 5. FCM clustering results in 3, 4 and 5 clusters.

When the number of clusters is 3 in K-means (see Fig. 4a) the three clusters are distributed in: i) the blue region where the fatty capsule is located, ii) the red region containing capsule (top) and reticulum (down middle) and iii) the green region which is nodal tissue. In FCM (see Fig. 5a), the three clusters are: i) the red region which covers cancerous node (down middle), fatty capsule (top left corner) and reticulum (possibly subcapsular sinus) (middle up area), ii) the blue region includes capsule (top) and reticulum (possibly cortical sinus) and iii) the green region contains normal node (possible germinal centre) (middle) and reticulum (possible cortical sinus) (middle down).

In general, FCM clusters the spectra into three main tissue types, cancerous node, normal node and reticulum lined tissues (subcapsular and cortical sinus). However these clusters are partially intermingled with different tissue subtypes. E.g. cancer mixed with reticulum (subcapsular sinus) and fatty capsule; normal node mixed with reticulum (cortical sinus). K-means is able to split the fatty tissue, reticulum and nodal tissue. However, it was unable to differentiate between the cancerous and normal tissue.

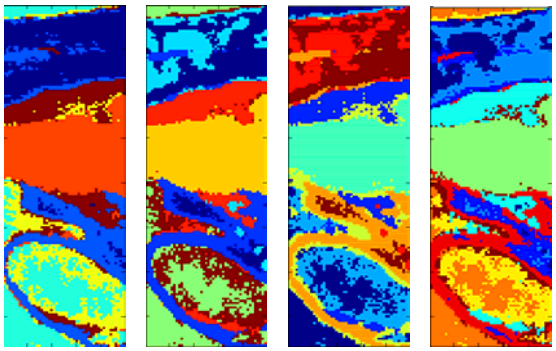
When the number of cluster increased to 4 and 5, the clusters are almost same as with 3 clusters in both K-means and FCM. The difference is that, using 4 and 5 clusters, both clustering algorithms were able to discover all subtypes of reticulum (up and down side of normal node – possible germinal centre). In 5 clusters, both clusters started to identify some reticulum in the cancerous node area. It can be seen that there are always major differences between the classifications obtained using FCM to those using K-means and, whilst the cancerous node and normal nodes can consistently be separated using FCM, K-means is inconsistent in distinguishing between these regions. This is a very useful observation for diagnosis purposes.

In the following results, the number of clusters is increased up to 9 for both algorithms. The cluster plots are presented in Fig. 5 and Fig. 6.



6 clusters 7 clusters 8 clusters 9 clusters

Fig. 6. K-means clustering results from 6 to 9 clusters.



6 clusters 7 clusters 8 clusters 9 clusters

Fig. 7. FCM clustering results from 6 to 9 clusters.

Starting from 6 clusters, the K-means algorithm begins to separate cancerous node and normal node although there is a small amount of reticulum that has also been classified with the cancerous node region. However, the FCM algorithm also started to mix reticulum in with the cancerous node area. When number of clusters increase to 7 and 8, both clustering algorithms classified more subtypes of tissues in the capsule area. FCM also classified more mixed types of tissue in the cancerous node region. Finally, increasing the number of clusters

to 9 yields more and more types of tissue being mixed together. The additional clusters within these tissue sections may identify potential subtypes of tissue which cannot currently be identified by pathological analysis but may be useful for diagnosis (of course, they may also be clustering noise).

Overall, FCM can split the two main different types of clusters in the early stage of clustering. When increasing the number of clusters, the main different types of tissue can be separated by K-means as well. However, the fatty capsule region can not be separated from the cancerous node using FCM with any of the cluster numbers used within this paper (2 to 9 clusters). As the number of clusters increases, more and more information is obtained about the tissue which cannot be identified by pathologists.

4. CONCLUSIONS

Fourier Transform Infrared Microspectroscopy has been widely used to study several different diseases. However, it has not been used in the assessment of cancers that specifically spread from the breast to lymph nodes. In this paper, two well-known clustering algorithms, K-means and Fuzzy C-Means, were applied to cluster a tissue section analysed by FTIR that had been diagnosed with metastatic infiltration (cancer spread from its original location). Initially, it was shown that the performance of FCM was poor. However, by investigating the size of first three principal components ranges, we found that when the minimal amount of improvement is not small enough, the iterations will stop without doing any further improvement. Based on this finding, we adjusted the minimal amount of improvement setting; the results showed that the performance of the FCM algorithm was improved significantly.

The number of clusters on this tissue section was subjectively set from 2 to 9 within the two clustering algorithms. The results showed that FCM can separate the main different tissue types using a smaller number of clusters. K-means was only able to satisfactorily classify them when the number of clusters was increased. After increasing the number of clusters, more and more information was obtained within the classification (in particular, the possible identification of tissue subtypes) which cannot be recognized by the pathologist.

In the future, this research programme will continue to investigate both K-means and FCM clustering algorithms. In particular, we are investigating methods to enable the optimal number of clusters to be automatically and consistently identified. Further tissue sections will be collected and used to evaluate our findings in this paper and future research.

ACKNOWLEDGEMENTS

The authors are grateful to Benjamin Bird for providing the FTIR spectra and clinical analysis used within this study and for his valuable comments and discussion.

REFERENCES

1. Landis, S. H., Murray, T., Bolden, S. and Wingo, P. A., 1999, "Cancer Statistics", CA: A Cancer Journal for Clinicians, 49, 1, 8-31

2. Turner, R. R., Ollila, D. W., Krasne, D. L. and Giuliano, A. E., 1997, "Histopathologic validation of the sentinel lymph node hypothesis for breast cancer", Annals of Surgery, 226, 271-278

3. Godavarty, A., Thompson, A. B., Roy, R., Gurfinkel, M., Eppstein, M. J., Zhang, C. and Sevick-Muraca, E. M., 2004, "Diagnostic imaging of breast cancer using fluorescence-enhanced optical tomography: phantom studies", Journal of Biomedical Optics, 9, 488-496

4. Johnson, K. S., Chicken, D. W., Pickard, D. C. O., Lee, A. C., Briggs, G., Falzon, M., Bigio, I. J., Keshtgar, M. R. and Bown, S. G., 2004, "Elastic scattering spectroscopy for intraoperative determination of sentinel lymph node status in the breast", Journal of Biomedical Optics, 9, 6, 1122-1128

5. Stone, N., Kendall, C., Shepherd, N., Crow, P. and Barr, H., 2002, "Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers.", Journal of Raman Spectroscopy, 33, 7, 564-573

6. Yeung, D. K. W., Yang, W. and Tse, G. M. K., 2002, "Breast Cancer: In Vivo Proton MR Spectroscopy in the Characterization of Histopathologic Subtypes and Preliminary Observations in Axillary Node Metastases", Radiology, 225, 1, 190-197

7. Wang, X. Y. and Garibaldi, J., 2005, "Simulated Annealing Fuzzy Clustering in Cancer Diagnosis", European Journal of Informatica

8. Benedetti, E., Teodori, L., Trinca, M. L., Vergamini, P., Slavati, F., Mauro, F. and Spremolla, G., 1990, "A new approach to the study of human solid tumor cells by means of FT-IR microspectroscopy", Applied Spectroscopy, 44, 1276-1280

9. Crupi, V., De Domenico, D., Interdonato, S., Majolino, D., Maisano, G., Migliardo, P. and Venuti, V., 2001, "FT-IR spectroscopy study on cutaneous neoplasie ", Journal of Molecular Structure, 563-564, 115-118

10. Mantsch, M. and Jackson, M., 1995, "Spectroscopy in biodiagnostics (From Hippocrates to Herschel and beyond)", Journal of Molecular Structure, 347, 187-206

11. Eckel, R., Huo, H., Guan, H. W., Hu, X., Che, X. and Huang, W. D., 2001, "Characteristic infrared spectroscopic patterns in the protein bands of human breast cancer tissue", Vibrational Spectroscopy, 27, 2, 165-173

12. Paluszkiewicz, C. and Kwiatek, W., 2001, "Analysis of human cancer prostate tissue using FTIR microspectroscopy and SRIXE techniques", Journal of Molecular Structure, 565-566, 329-334

13. Wang, X. Y., Whitwell, G. and Garibaldi, J., 2004, "The Application of a Simulated Annealing Fuzzy Clustering Algorithm for Cancer Diagnosis″, Proceedings of the IEEE 4th International Conference on Intelligent System Design and Application, Hungary, 467-472

14. Wang, X. Y., Garibaldi, J. and Ozen, T., 2003, "Application of The Fuzzy C-Means Clustering Method on the Ananlysis of non Pre-processed FTIR Data for Cancer Diagnosis″, Proceedings of the 8th Australian and New Zealand Conference on Intelligent Information Systems, December 10-12, Australia, 233-238

15. McQueen, J. B., 1967, "Some methods of classification and analysis of multivariate observations″, Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, 281-297

16. Bezdek, J., 1981, "Pattern Recognition With Fuzzy Objective Function Algorithms″, Plenum, New York

17. Causton, D. R., 1987, "A Biologist's Advanced mathematics″, Allen & Unwin, London

18. Jolliffe, I. T., 1986, "Principal Component Analysis″, Springer-Verlag, New York