

# SISTEMA DE CLASSIFICAÇÃO DE ESPECTROS RAMAN PARA DETECÇÃO DE LESÕES ATEROSCLERÓTICAS

RICARDO S. SANTOS, HASSAN SIDAOU, LANDULFO SILVEIRA JR.

*Instituto de Pesquisa e Desenvolvimento, Universidade do Vale de Paraíba  
Av. Shishima Hifumi, 2911 CEP 12244-000 São José dos Campos, SP  
E-mail: rsantos@mgr.com.br*

**Resumo** — As técnicas da Espectroscopia Raman, são amplamente utilizadas, sobretudo, como métodos de caracterização de substâncias. O processamento e análise dos espectros obtidos a partir de uma determinada substância apresentam variações, que podem ser decorrentes de alterações no ambiente de aquisição do sinal, ou até mesmo provenientes de pequenas diferenças entre as amostras do conjunto de aquisições. É necessário um exaustivo trabalho para identificação e classificação dos espectros que apresentam diferenças. Neste trabalho apresentamos uma ferramenta computacional baseada em métodos de reconhecimento de padrões, para classificar os espectros de um conjunto de amostras de tecido arterial. O objetivo é a identificação de subgrupos de espectros, dentro de um grupo inicial, de acordo com as características dos picos destes espectros. Os subgrupos identificados representam diferentes estados da parede íntima arterial.

**Abstract** – The Raman Spectroscopy, are largely used, especially, as substance characterization method. The processing and analysis of the a specific substance show variations. The origin of the variations can be alteration in signal acquisition environment, or the small differences among the individual elements of the set sample. The exhaustive work is necessary to identify and classify the different spectra in a sample set. This work show a computational tool that use pattern recognition method to classification of human artery spectra. The objective is the identification of subsets in a sample spectra set. The subgroups identified, represent the differents states of artery wall.

**Key Words**— Pattern Recognition, Clustering, Signal Processing, Intelligent Instrumentation, Raman Spectroscopy.

## 1 Introdução

A Espectroscopia Raman para caracterização de substâncias consiste na aquisição e interpretação de sinais espectrais, obtidos de uma amostra que é excitada por um laser. A figura 1 mostra um sistema de aquisição. (Sidaoui et al. 1998).

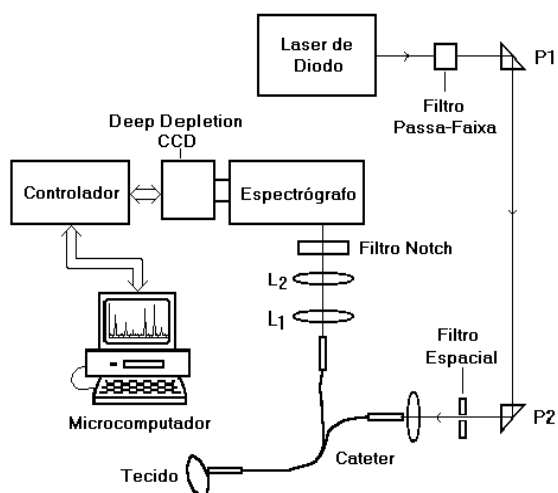


Figura 1 - Sistema de Aquisição Espectral

O sinal espectral coletado deve ser interpretado para a caracterização da amostra. A interpretação do espectro é baseada na análise dos picos de frequências presentes no sinal e requer um conhecimento especialista.

A automatização da interpretação para caracterizar a presença de uma determinada substância requer a determinação de padrões a partir de um conjunto numeroso de espectros, visando capturar o conhecimento especialista envolvido.

A informação contida no espectro é representada por picos existentes, sendo que os dados representativos de um pico são: posição, intensidade e largura a meia altura (FWHM), conforme mostra a figura 2. (Sidaoui et al. 1998).

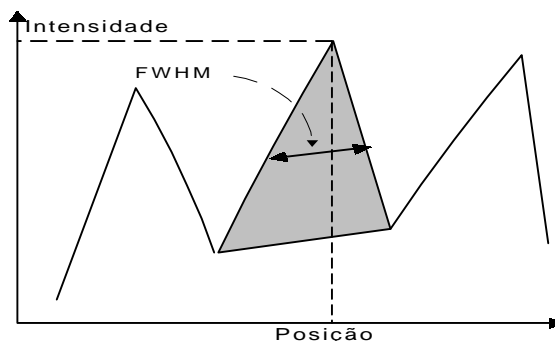


Figura 2 - As features de um pico

A determinação de padrões espectrais é dificultada pela existência de variações nos picos presentes em espectros da mesma substância. Essas variações podem ser provocadas por ruídos no processo de aquisição ou decorrentes de pequenas diferenças entre as amostras do conjunto de aquisições. A figura 3 mostra dois espectros obtidos de artérias coronárias normais, porém apresentando variações nos picos.

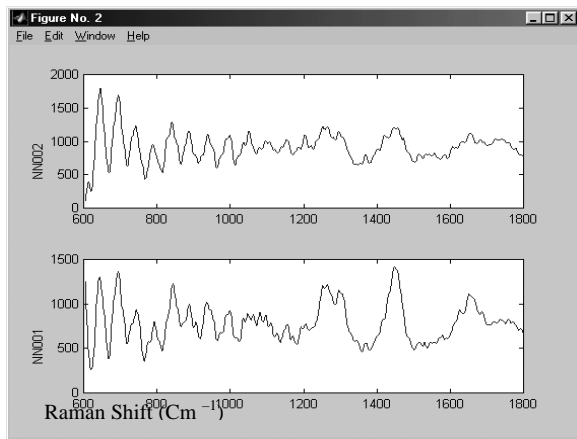


Figura 3– Espectros da mesma substância com formas diferentes

As variações assinaladas tornam as tarefas de agrupamento e classificação difíceis e sujeitas a erros. A automatização do processo de agrupamento baseando-se na análise das diferentes características dos picos deve amenizar esse problema e tornar o processo de classificação mais confiável.

Neste trabalho apresentamos uma ferramenta computacional que usa métodos de reconhecimento de padrões para classificação espectral. O método utilizado é uma técnica de ‘clustering’ hierárquico, baseado no algoritmo “Minimal Spanning Tree” (McAndrews, 1995), que está fundamentado na fórmula da distância euclidiana.

## 2 Descrição do Sistema de Agrupamento e Classificação

A ferramenta computacional foi desenvolvida usando a biblioteca MATLAB (MathWorks, MA, USA, versão 6.0). A ferramenta permite a criação de subgrupos de sinais espectrais, dado um determinado conjunto de espectros de entrada.

Para flexibilizar o uso da ferramenta à diferentes situações e diferentes tipos de sinais espectrais, foi estabelecido um conjunto de parâmetros iniciais que devem ser ajustados para cada situação, levando em consideração principalmente a natureza do espectro.

### 2.1 Dados de entrada para Determinação dos Padrões

Para obter um resultado satisfatório é necessário um conhecimento prévio da natureza dos espectros do conjunto de entrada, pois a informação relevante está contida em picos localizados em regiões específicas. Entre os parâmetros iniciais estão as regiões de análise dos picos, que definem as regiões dos picos relevantes para análise e a importância relativa de cada faixa.

Os parâmetros “Prioridade de Intensidade” e “Prioridade de Largura” permitem a ponderação da intensidade dos picos e da sua largura no processo de comparação dos sinais.

O “número de clusters” é o parâmetro que determina a quantidade de subgrupos de saída a serem formados.

### 2.2 Metodologia Empregada

Com os parâmetros iniciais fornecidos, assim como o conjunto de espectros a serem agrupados, calcula-se para cada espectro as características dos picos de interesse (posição, intensidade e largura) usando um algoritmo apresentado em Sidaoui et al. 1998.

O método empregado utiliza o princípio de distância entre espectros.

A distância entre dois espectros é calculada através da fórmula de distância euclidiana entre as características dos picos dos espectros considerados:

$$D = \sqrt{(I(n) - I(n-1))^2 + (L(n) - L(n-1))^2}$$

onde

$$I = \text{Intensidade do Pico} * \text{Prioridade da Intensidade} * \text{Prioridade Região}$$

$$L = \text{Largura do Pico} * \text{Prioridade da largura} * \text{Prioridade Região}$$

Após cálculo das características dos picos para cada espectro, é estabelecido um número de subgrupos de saída (clusters) igual ao número de espectros de entrada (agrupamento 0). Em seguida, são calculadas as distâncias entre os subgrupos, baseadas nas características (médias) dos picos dos subgrupos considerados. A cada passo o número de subgrupos reduz-se, agrupando os subgrupos que apresentam a menor distância entre si. O processo é encerrado quando a quantidade de subgrupos obtidos torna-se igual à quantidade desejada (“Minimal Spanning Tree”, McAndrews, 1995).

Para poder comparar sinais obtidos com tempos de aquisição diferentes, é necessário ‘normalizar’ estes sinais, isto é dividi-los pela intensidade máxima obtendo dessa forma sinais com faixas de intensidade entre 0 e 1.

A figura 4 ilustra a arquitetura básica do sistema.

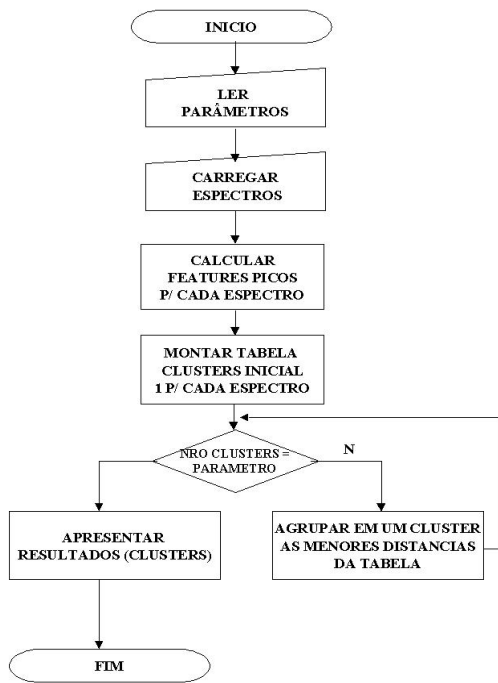


Figura 4 - Arquitetura básica do sistema

O método utiliza o seguinte algoritmo para classificação dos espectros:

1. Para cada espectro do conjunto de espectros iniciais, são determinadas as características dos picos das regiões de interesse :

$$i_1, f_1, i_2, f_2$$

onde  $i$  e  $f$  representam respectivamente a intensidade e largura (FWHM) do pico. Para reconhecimento dos picos do espectro utilizamos a técnica descrita em Horowitz, 1977.

2. É criada uma tabela (tabela de clusters) atribuindo cada espectro para um único cluster. O número de linhas da tabela coincide com o número de espectros. A tabela 1 mostra o modelo da tabela de clusters.

		Região 1		Região 2		Região n	
		Int.	Lar.	Int.	Lar.	Int.	Lar.
C1	E1	$I_{11}$	$L_{11}$	$I_{12}$	$I_{12}$	$I_{1n}$	$I_{1n}$
C2	E2	$I_{21}$	$L_{21}$	$I_{22}$	$I_{22}$	$I_{2n}$	$I_{2n}$
...	...	...	...	...	...	...	...
Cn	E <sub>m</sub>	$I_{n1}$	$L_{n1}$	$I_{n2}$	$I_{n2}$	$I_{nn}$	$L_{nn}$

Tabela 1 – Tabela inicial de Clusters

Onde

C1 = Cluster 1, C2 = Cluster 2, ..., Cn = Cluster n  
 E1=Espectro 1, E2=Espectro 2, ..., En=Espectro n.

$I_{11}$  = Intensidade do espectro 1, na região 1

$I_{12}$  = Intensidade do espectro 1, na região 2

$I_{1n}$  = Intensidade do espectro 1, na região n

$L_{11}$  = Largura do espectro 1, na região 1

$L_{12}$  = Largura do espectro 1, na região 2

$L_{1n}$  = Largura do espectro 1, na região n

3. Calcula-se a distância entre os clusters, considerando os valores dos picos dos espectros conforme a tabela de clusters e estabelece uma tabela de distâncias relacionando cada cluster. Nesta tabela de distâncias, as células que representam distâncias já estabelecidas anteriormente, recebem o valor '-2', e as células que representam a distância de um cluster para si mesmo recebem o valor '-1'. A tabela 2 mostra a tabela das distâncias entre os clusters.

Cluster	C1	C2	C3	C4
C1	-1	$d(E1-E2)$	$d(E1-E3)$	$d(E1-E4)$
C2	-2	-1	$d(E2-E3)$	$d(E2-E4)$
C3	-2	-2	-1	$d(E3-E4)$
C4	-2	-2	-2	-1

Tabela 2 – Tabela de distâncias entre os clusters

4. Os clusters que apresentam a menor distância são agrupados em um único cluster, e a tabela inicial que continha um espectro em cada cluster, agrupará os espectros de dois clusters em uma única linha. Os valores de Intensidade e Largura da nova linha, será a média dos valores dos dois clusters que foram agrupados.
5. Os passos 3 e 4 são repetidos enquanto a tabela de clusters conter um número de clusters maior que o desejado.

### 3 Resultados

Uma das áreas de aplicação da Espectroscopia Raman é o diagnóstico de lesões ateroscleróticas, que se caracteriza por formações de placas na camada interna da parede arterial (Richards-Kortum et al., 1989. Hoyt et al., 1988). Estudos anteriores, já caracterizaram três classes espectrais obtidas de artérias: Tecido Normal, Tecido com placa de Ateroma e Tecido com placa de ateroma Calcificada.

A comparação dos resultados da análise histológica (fornecida) com os resultados da análise

do sistema proposto permite uma avaliação do mesmo.

Primeiramente um conjunto de 18 espectros foi utilizado para testar o método e ajustar os parâmetros iniciais da ferramenta computacional. Este primeiro conjunto de espectros era constituído por 6 espectros de tecido arterial normal, 6 espectros de artérias com placas de ateroma e 6 espectros de artérias com placas calcificadas. Este conjunto de espectros foi submetido à ferramenta e os parâmetros foram ajustados até conseguirmos os 3 grupos de seis espectros, representantes de cada estado da parede arterial.

A tabela 3 mostra os parâmetros de entrada ideais para o reconhecimento dos três grupos de espectros arteriais desejados.

<b>Regiões / Prioridades</b>
600-700 / 1
900-1000 / 1
1400-1500 / 1
1600-1700 / 1
<b>Prioridade Intensidade</b>
1
<b>Prioridade Largura</b>
1
<b>Faixa Espectral</b>
600 – 1800

Tabela 3 – Configurações do sistema

Definidos os parâmetros de classificação, utilizamos um segundo grupo de espectros obtidos de outras amostras de tecido arterial, para validar a ferramenta de classificação. Com este grupo realizamos diferentes testes visando o agrupamento e a classificação dos espectros.

Os teste realizados foram os seguintes:

1. Três clusters de saída. Dez espectros de entrada sendo quatro de artéria normal, três de artérias com placas de ateroma e três de placas calcificadas.
2. Três clusters de saída. Aos espectros do primeiro teste acrescentamos quatro espectros sendo dois de artérias com placas de ateromas e dois de artérias com placas calcificadas.
3. Dois clusters de saída com os espectros do primeiro teste.
4. Quatro clusters de saída com os espectros do primeiro teste.

5. Cinco clusters de saída com os espectros do primeiro teste.

Os resultados obtidos são descritos a seguir:

- A) Os três grupos correspondentes aos três tipos de artérias (normal, ateroma e calcificada) foram criados (figuras 5A, 5B e 5C) Todos os espectros considerados foram agrupados de forma correta de acordo com os resultados histológicos.
- B) Os novos espectros introduzidos no segundo teste foram classificados de forma correta de acordo com os resultados histológicos (figuras 6A e 6B).
- C) A redução para 2 subgrupos de saída provoca o agrupamento dos espectros de artérias normais e os espectros de artérias com placas de ateroma em um único grupo. O segundo grupo, que corresponde aos espectros de artérias calcificadas, permanece inalterado. Os espectros das artérias calcificadas apresentam um formato bem específico. A figura 7 mostra o grupo ‘Normal + Ateroma’.
- D) Ao aumentar o número de subgrupos de saída para 4, foram gerados os três grupos (normal, cálcio e ateroma) dos testes anteriores e um novo grupo que contém um espectro extraído do grupo Ateroma com algumas diferenças em relação aos demais espectros deste grupo. Apresenta um pico mais intenso na região de 600-700, conforme mostra a figura 8.
- E) Com 5 subgrupos de saída, os espectros normais são separados em 2 subgrupos. A análise histológica mostra que os espectros extraídos do grupo de artérias normais, correspondem a artérias com camada íntima mais densa (espessamento da íntima).

Além dos testes apresentados neste trabalho, um segundo teste composto por um conjunto de 81 espectros, foi testado para avaliar o desempenho dos agrupamentos obtidos, e a ferramenta apresentou um resultado satisfatório, conseguindo uma classificação correta em 90 % dos casos.

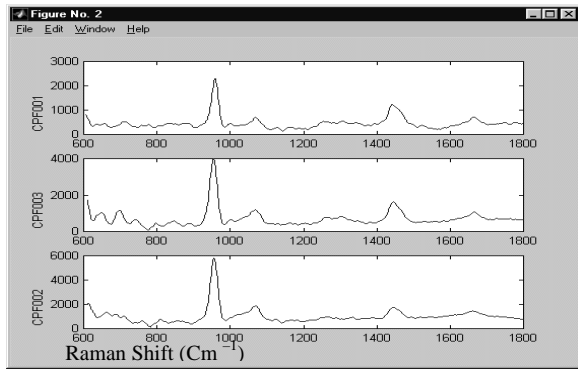


Figura 5(A) – Espectros do grupo “Calcificadas”

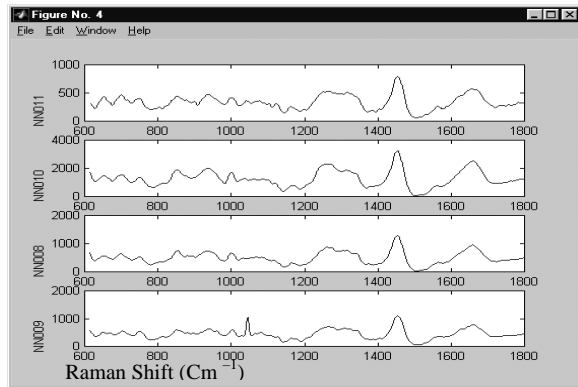


Figura 5(B) – Espectros do grupo “Normal”

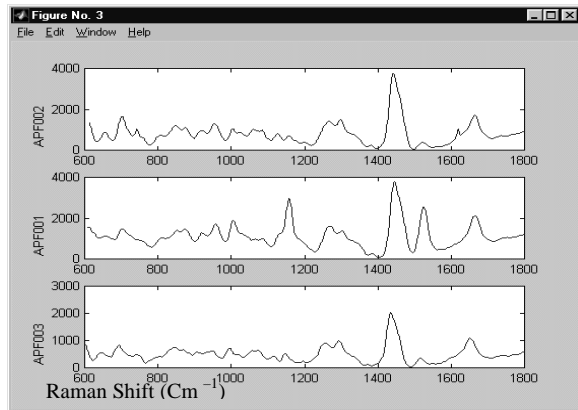


Figura 5(C) – Espectros do grupo “Ateroma”

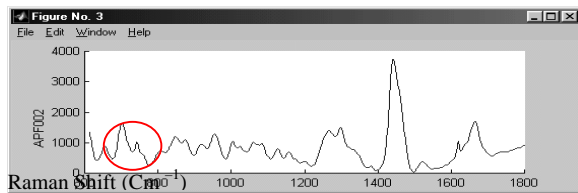


Figura 8 – Elemento extraído do grupo “Ateroma”.

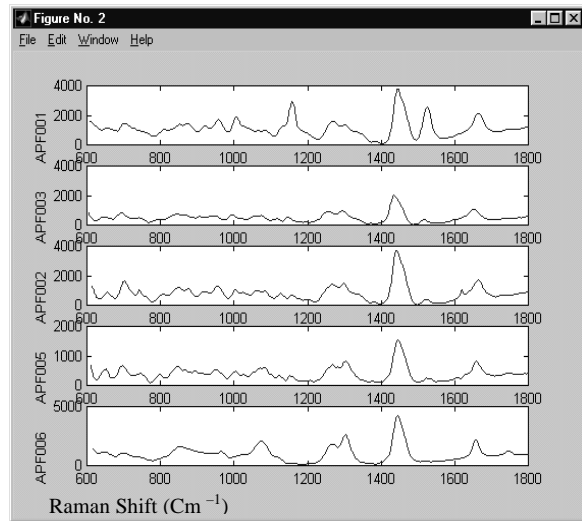


Figura 6(A) – Espectros do grupo “Ateroma”, acrescido dos dois novos espectros

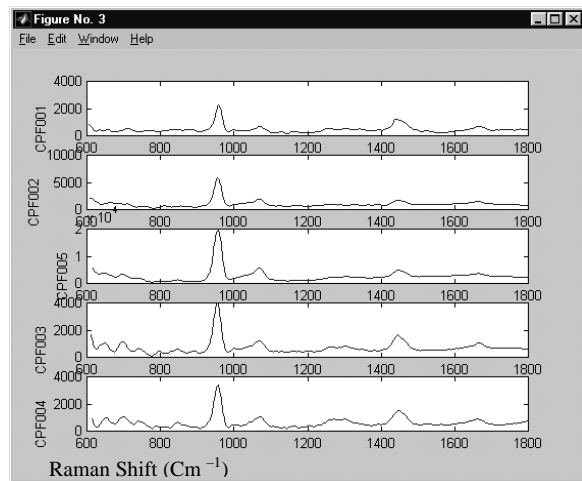


Figura 6(B) – Espectros do grupo “Calcificadas”, acrescido dos dois novos espectros

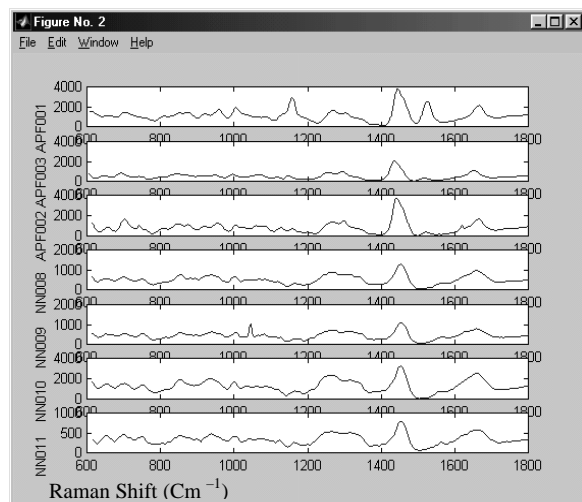


Figura 7 – Espectros do grupo “Ateroma” + “Normal”

#### 4 Conclusão

Os resultados obtidos são satisfatórios e permitem comprovar a eficiência da ferramenta desenvolvida para classificação de sinais espectrais.

Essa ferramenta possibilita a identificação automática de padrões espectrais com um alto grau de acerto, pois, em 90% dos casos, os resultados obtidos coincidiram com os resultados histológicos, conseguindo até agrupar espectros que dificilmente seriam considerados semelhantes através de uma análise visual.

Recentemente existem outras técnicas utilizadas para a classificação de espectros Raman (Deinum, 1999). A próxima fase do trabalho é demonstrar a validade do método adotado, comparando sua eficiência com outros métodos de classificação.

#### 5 Referências bibliográficas

Hassan Sidaoui, Landulfo Silveira JR e Marcos T. T. Pacheco (1998). “Sistema Automático de Tratamento de Sinais Espectrais para Detecção em Tempo Real de Lesões Ateroscleróticas”. *Anais do XII Congresso Brasileiro de Automática, Volume II*, pp. 411-415.

Priscilla McAndrews, et Al. (1995). “Tutorial of Probability and Statistic”. *Center for the New Engineer, George Mason University, Fairfax, Virginia*.  
[http://cne.gmu.edu/modules/dau/stat\\_idx\\_frm.html](http://cne.gmu.edu/modules/dau/stat_idx_frm.html)

Hassan Sidaoui e Marcos T. T. Pacheco (1997). “Técnicas de Reconhecimento de Padrões Aplicada ao Diagnóstico de Lesões Ateroscleróticas”. *Anais do XX Encontro Nacional de Física da Matéria Condensada*, pp. 3 e 28.

Horowitz S.L. (1977). “Peak Recognition in Waveforms”, in: K.S. Fu et al.: *Syntactic Pattern Recognition, Applications*, Springer-Verlag, pp. 30-49.

Hoyt C.C., R.R. Richards-Kortum, B. Costello, B.A. Sacks, C. Kittrel, N.B. Ratliff, J.R. Krammer and M.S. Feld (1988). Remote Biomedical Spectroscopic Imaging of Human Artery Wall. *Lasers in Surgery and Medicine*. Vol. 8, pp.1-9.

Richards-Kortum, A. Mehta, G. Hayes and R. Cothren (1989). Spectral Analysis of Atherosclerosis Using an Optical Fiber Laser

Catheter. *American Heart Journal*, Vol. 118, No 2, St.Louis., pp. 381-391.

Geurt Deinum, et al. (1999). “Histological Classification of Raman Spectra of Human Coronary Artery Atherosclerosis Using Principal Component Analysis”, *Applied Spectroscopy*, Vol. 53, Number 8, pp. 938-942.

A. Baraldi, P. Blonda, (1998). “A Survey of fuzzy clustering algorithms for pattern recognition”. *International Computer Science Institute, Berkeley*.  
[http://cne.gmu.edu/modules/dau/stat\\_idx\\_frm.html](http://cne.gmu.edu/modules/dau/stat_idx_frm.html)