

Pré-Processamento do Espectro Raman para Detecção de Placas Ateromatosas através de Redes Neurais

Alderico R. de Paula Jr¹; Claudia E. da Matta¹; Érika R. M. Teixeira¹;
Landulfo Silveira Jr¹; Paulo R. Galhanone¹; Roberto K. H. Galvão²

1. Instituto de Pesquisa e Desenvolvimento - IP&D - UNIVAP

Av Shishima Hifumi 2911, São José dos Campos, SP

E-mail: alderico/erika/landulfo/galhanon@univap.br

2. Instituto Tecnológico de Aeronáutica - ITA

Centro Técnico Aeroespacial, São José dos Campos, SP

E-mail:kawakami@ele.ita.cta.br; matta@bibl.ita.cta.br

Abstract

This article presents the results of a study to reduce the number of variables necessary to represent the Raman spectrum acquired from the artery tissue when irradiated by a low power infrared laser. The analyzed methods were: peak detection, genetic algorithm, wavelets and principal components. The reduced vectors generated by these methods were applied to an artificial neural network to classify the target artery tissue into a non-pathological tissue, atherosclerotic lesion or calcified tissue. For each method, the neural network was trained by a set of 15 reduced spectra of samples of the three types of tissues and validated by another set of 10 sample reduced spectra. Among the analyzed methods, the spectrum compression by wavelets and the variable number reduction by principal components presented the best results and they were selected for the next phase of the project.

1. Introdução

A aterosclerose é causada pela deposição de placas de gordura, denominadas de ateromas, na superfície interna de artérias, diminuindo seu calibre e podendo até bloquear completamente a circulação do sangue. Alguns ateromas evoluem para placas calcificadas diminuindo significativamente a elasticidade das paredes das artérias. Uma das técnicas [1] que vem sendo pesquisada para a detecção da presença de ateromas é a espectroscopia Raman. O diagnóstico [2] pode ser feito *in vivo*, introduzindo-se um cateter óptico na artéria do paciente. A Figura 1 apresenta um equipamento em desenvolvimento no Instituto de Pesquisa e Desenvolvimento (IP&D) da UNIVAP para detecção de ateromas.

Neste equipamento uma das fibras ópticas conduz a radiação do laser de baixa potência, operando no infravermelho (830 nm), até a região da artéria que se quer examinar. A radiação do laser é espalhada elástica e inelasticamente pelo tecido, além de induzir as células a gerarem fluorescência. Um conjunto de fibras ópticas capta a radiação retroespalhada e a fluorescência do

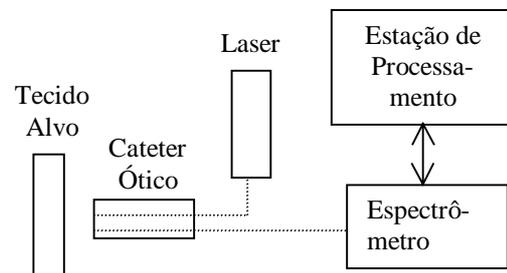


Figura 1 - Equipamento para Detecção de Ateromas

tecido alvo e conduz a radiação coletada até o espectrômetro, que utiliza uma grade de difração para espalhar espectralmente a radiação coletada sobre uma câmera CCD de 1024 elementos. A região espectral de interesse está na faixa entre 600 a 1800 cm^{-1} . O tempo de exposição da CCD é controlado pela estação de processamento e o sinal gerado na CCD é convertido de analógico para digital e enviado para a estação de processamento.

O sinal recebido na estação é formatado em um vetor de 1024 elementos sendo que cada elemento representa a radiação captada pela CCD em uma faixa do espectro. Em seguida, o sinal é filtrado para reduzir o ruído gerado pela CCD, calibrado espectralmente e removida a parte referente à fluorescência restando, assim, apenas o sinal gerado relativo ao espalhamento inelástico da radiação laser denominado de radiação Raman. Este sinal fornece informações das substâncias que constituem os tecidos alvos. Portanto, analisando-se o sinal adquirido é possível classificar o tecido alvo em sadio, ateromatoso ou calcificado. A Figura 2 apresenta os espectros de uma amostra de tecido sadio, uma de ateromatoso e outra de calcificado, após serem filtrados e calibrados.

Atualmente o método utilizado para classificar o tecido alvo é uma máquina de inferência [3] desenvolvida com base em conhecimentos de especialistas e que utiliza a localização e amplitude dos principais picos Raman. Este método opera satisfatoriamente quando o tempo de exposição do CCD é maior que 500 ms. Contudo, ao reduzir o tempo de

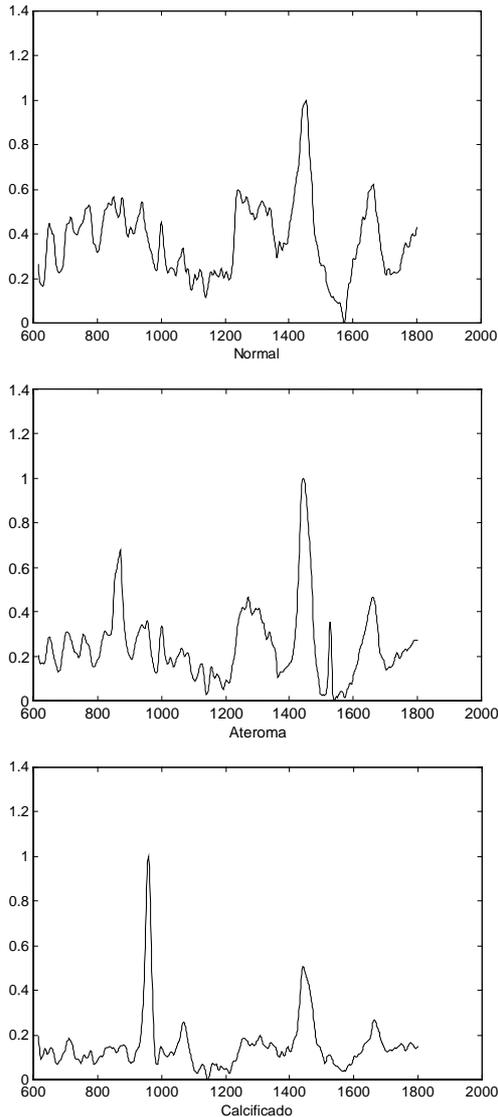


Figura 2 Espectro Raman típico de um tecido sadio, ateromatoso e calcificado

exposição do CCD a taxa de acerto diminui. Como se pretende utilizar uma técnica de classificação do tecido em um aparelho para remoção de ateromas [3], o tempo de exposição deverá ser menor que 100 ms.

Este artigo apresenta o resultado da utilização de um novo método de classificação do tecido alvo baseado em redes neurais artificiais. O vetor do espectro da amostra após a calibração foi reduzido a 754 variáveis e foram analisados diversos métodos de pré-processamento do vetor calibrado visando reduzir o número de variáveis. Os resultados serão apresentados nas seções seguintes. Os métodos de pré-processamento selecionados foram os seguintes:

- Seleção dos picos Raman de maior amplitude por faixa;
- Seleção de algumas frequências utilizando-se algoritmos genéticos;
- Compressão do sinal utilizando-se wavelets;

- Redução do número de variáveis através de componentes principais.

Para treinar a rede neural foram utilizadas quinze amostras sendo seis de tecido ateromatoso, seis de tecido sadio e três de tecido calcificado, organizados em uma matriz de treinamento denominada X, formada de 15 colunas de 754 elementos. Para validar a rede neural foram utilizadas dez amostras sendo quatro de tecido ateromatoso, três de tecido normal e três de tecido calcificado, organizadas em uma matriz de validação, denominada Y, contendo dez colunas de 754 elementos.

2. Redes Neurais Artificiais

As redes neurais artificiais (RNA) [4] são algoritmos computacionais apropriados para soluções de problemas complexos de classificação e reconhecimento de padrões. De forma semelhante ao cérebro humano, as RNAs são organizadas através de elementos de processamento simples, denominados neurônios que se interligam uns com os outros formando redes capazes de processar, armazenar e transmitir informações vindas do exterior. As redes podem ser treinadas para resolver determinadas classes de problemas através da alteração dos padrões das interligações entre os neurônios. Existem diversos tipos de redes neurais. Este trabalho limitar-se-á às redes neurais multicamadas sem realimentação, com uma ou duas camadas escondidas e três neurônios na camada de saída conforme apresentado na Figura 3.

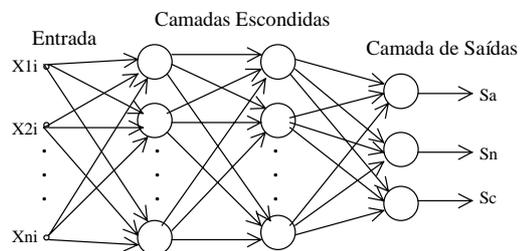


Figura 3 - Rede neural selecionada

A função de transferência escolhida para os neurônios das camadas escondidas foi a signóide e na camada de saída a linear. Os algoritmos de aprendizagem são do tipo retropropagação utilizando-se o método do gradiente.

Para facilitar as discussões, nas sessões posteriores as redes neurais serão representadas na seguinte forma: RNA-E1-E2-S, onde E1 representa o número de neurônios na primeira camada escondida, E2 o número de neurônios na segunda camada escondida quando esta for utilizada e S o número de neurônios na camada de saída que neste trabalho foi fixado em três. Os três neurônios de saída são denominados de Sa, Sn e Sc. Visando-se estabelecer uma distância razoável entre os valores classificados, as saídas desejadas para o treinamento foram estabelecidas de acordo com a Tabela 1.

Tabela 1 Valores esperados para as saídas

	Sa	Sn	Sc
Tecido Normal	-1	1	-1
Tecido Ateromatoso	1	-1	-1
Tecido Calcificado	-1	-1	1

As redes neurais foram treinadas até atingirem um erro médio quadrático (EMQ) entre as saídas e os valores esperados de 0,005 ou 5000 ciclos de treinamento. Na fase de validação da rede neural os seguintes critérios foram utilizados para a classificação:

- A amostra foi classificada como correta (C) quando os sinais das saídas foram os mesmos que as saídas esperadas.
- Incorreta (E) quando os sinais das saídas foram diferentes das saídas desejadas, entretanto iguais às saídas esperadas de um outro tipo de tecido.
- Não classificados (N) quando os sinais das saídas foram diferentes dos sinais desejados para os três tipos de amostras.

Para comparar os diversos métodos foi calculado o EMQ de cada amostra em relação às saídas esperadas e o EMQ de todas as amostras de validação processadas pela rede neural. Como a rede gerada no treinamento é função dos valores iniciais das conexões, o algoritmo de treinamento foi executado 10 vezes e foi escolhida a rede que melhor classificou as amostras de validação ou a que gerou o menor EMQ.

3. Detecção de Picos por Faixa

Como a principal informação do espectro Raman está na localização em frequência e na amplitude dos picos, o método de detecção de picos por faixa consiste, em sua fase inicial, em localizar os principais picos do espectro e medir suas amplitudes. A Figura 4 apresenta o resultado desta etapa para o espectro de uma amostra de ateroma.

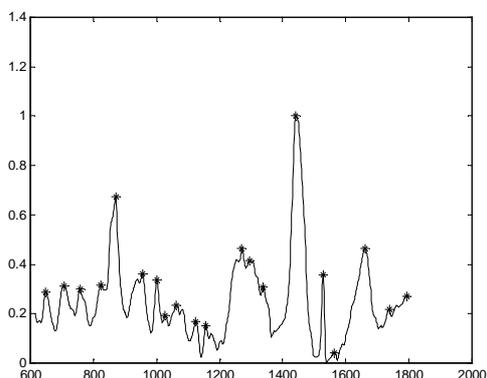


Figura 4 Localização dos principais picos

A segunda fase deste método consiste em gerar um novo vetor com um número menor de variáveis ou seja, uma variável por cada faixa de n pontos. O valor da variável de cada faixa é igual ao valor do maior pico da

faixa. Caso nenhum pico seja encontrado na faixa, a variável recebe o valor zero. Desta forma, o número de variáveis pode ser reduzido para $754/n$.

Utilizando-se este método foram realizados os três testes seguintes:

- PPF5 - $n=5$ gerando 150 pontos por amostra e uma RNA-64-16-3.
- PPF10 - $n=10$ gerando 75 pontos por amostra e uma RNA-64-16-3.
- PPF15 - $n=15$ gerando 50 pontos por amostra e uma RNA-32-3

A Tabela 2 apresenta os resultados dos testes realizados.

Tabela 2 - Resultados dos teste picos por faixa

Teste	% de acertos	% de erros	% de não classif.	EMQ
CP2	70	30	0	0,83
CP4	60	0	40	0,51
CP8	60	0	40	0,94

4. Algoritmos Genéticos

Os algoritmos genéticos (AG) [5] são algoritmos computacionais apropriados para busca de soluções ótimas em problemas complexos de difícil formulação algébrica. Estes algoritmos imitam a seleção natural dos seres vivos onde apenas os mais aptos sobrevivem. Os AGs são estruturados de forma que as informações de interesse do sistema em estudo são codificadas de forma análoga aos cromossomas biológicos. Cada gene representa um determinado parâmetro que será otimizado.

Inicialmente cria-se uma população onde os genes dos indivíduos da população são gerados de forma aleatória. Define-se em seguida um critério de avaliação (aptidão) dos indivíduos, os quais são selecionados para o cruzamento de forma que os mais aptos têm maiores chances de serem escolhidos. O passo seguinte consiste no cruzamento entre os genes de diversos indivíduos criando uma nova geração que herda características dos pais. Algumas das características são melhores que as dos pais e outras piores. Visando a geração de indivíduos diferentes do grupo anterior o algoritmo permite a mutação de genes de alguns indivíduos. Algumas mutações podem gerar indivíduos mais aptos e outras menos aptos. Os novos indivíduos são avaliados juntamente com os pais e um novo grupo é selecionado para novos cruzamentos. Este procedimento é repetido até que o erro mínimo desejado seja alcançado ou até atingir um número máximo de gerações.

O algoritmo genético acima apresentado foi adaptado para selecionar as n melhores frequências do vetor X que, ao treinar a RNA, fornecessem a melhor classificação das amostras de validação (vetor Y) com o menor erro médio quadrático entre as saídas fornecidas pelo RNA e as saídas esperadas. O cromossoma selecionado foi definido com apenas um gene de 754 bits. Neste cromossoma o bit 1 indicava a presença da

6. Componente Principal

O método da componente principal [8] ao ser aplicado na matriz de treinamento X fornece um novo conjunto de variáveis, denominadas de componentes principais, através de uma transformação linear das variáveis originais, ou seja, cada componente principal é formado pela combinação linear das variáveis originais. Os componentes principais são ortogonais entre si e, portanto, não contêm informações redundantes.

As coordenadas dos valores das amostras no novo sistema são denominadas “scores”. Os coeficientes da combinação linear dos componentes principais são denominados “loadings” e representam os cossenos dos ângulos entre os eixos originais e os novos eixos.

O primeiro componente principal é um eixo no qual ao se projetar o valor das amostras forma-se uma nova variável cuja variância é a máxima possível. O segundo componente principal é um novo eixo perpendicular ao primeiro. Ao se projetar o valor das amostras neste novo eixo forma-se uma nova variável cuja variância é a máxima possível. Os demais eixos são formados de forma similar. O número de componentes principais é igual ao número de variáveis originais, entretanto, as primeiras variáveis principais representam, em geral, a maior parte da variância e podem representar de forma adequada os valores das amostras.

Uma forma de determinar os componentes principais consiste na decomposição da matriz original X de dimensões n por m , em valores singulares, formando três novas matrizes denominadas U , S e V de tal forma que $X=U S V^t$. A matriz S é diagonal e contém os valores singulares de X . As matrizes U e V são ortogonais de dimensão n por n e m por m , respectivamente. A matriz T das novas coordenadas pode ser calculada das seguintes formas:

$$T=U*S=X*V.$$

O número de componentes principais a ser considerado depende da variância dos primeiros componentes principais. A Tabela 5 apresenta os cinco primeiros valores singulares da matriz X e sua variância percentual.

Tabela 5-Valores Singulares e Variância

Parâm.	Componente Principal					
	CP1	CP2	CP3	CP4	CP5	Demais
VS	34,81	6,08	5,02	3,72	3,16	7,95
VP	92,79	2,83	1,93	1,06	0,77	0,61

Onde VS significa Valor Singular e VP, Variância Percentual

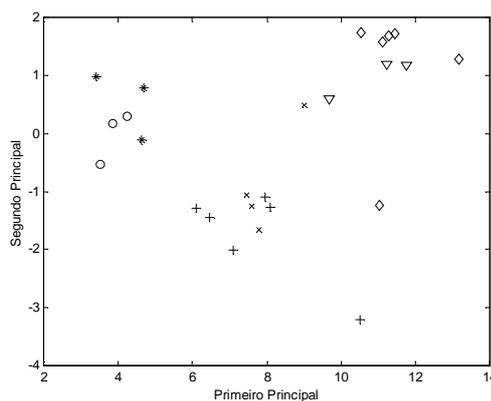
Verifica-se na Tabela 5 que o primeiro componente principal, CP1, contém 92,79% da variância, enquanto o segundo 2,83%. Portanto, os dois componentes principais podem representar de forma adequada a matriz X .

As coordenadas referentes aos primeiros k componentes principais da matriz de treinamento X , PX_k , e da matriz de validação Y , PY_k , podem ser calculados através das seguintes equações:

$PX_k=X*V(:,1:k)$ e $PY_k=Y*V(:,1:k)$, onde $V(:,1:k)$ representa as k primeiras colunas da matriz V .

A Figura 7 representa as coordenadas das 15 amostras de treinamento e as 10 de validação em relação aos dois primeiros componentes principais CP1 e CP2.

Nota-se na Figura 7 que houve uma boa separação das amostras, com exceção da quarta amostra do conjunto de validação de tecido ateromatoso que se aproximou do conjunto dos tecidos normais. Este fato pode indicar que o ateroma nesta amostra estava em uma fase inicial.



+ ≡ ateroma (treinamento); x ≡ ateroma (validação);
 ≡ normal (treinamento) ∇ ≡ normal (validação);
 * ≡ calcificado (treinamento); o ≡ calcificado (validação)

Figura 7 - Coordenadas das amostras em relação aos dois primeiros componentes principais.

Os seguintes testes foram realizados e são apresentados na Tabela 6:

- CP2 – Os dois primeiros componentes principais e uma RNA-10-3.
- CP4 – Os quatro primeiros componentes principais e uma RNA-10-3.
- CP8 – Os oito primeiros componentes principais e uma RNA-10-3.

Tabela 6 - Resultados do Método Componentes Principais

Teste	% de acertos	% de erros	% de não classif.	EMQ
CP2	90	0	10	0,11
CP4	90	10	0	0,26
CP8	90	10	0	0,19

7. Conclusões

O método de detecção de picos por faixa, apesar de apreender a principal informação da radiação Raman que é a localização em frequência e amplitude dos principais picos do espectro, não apresentou um bom resultado tendo apenas 70% de acertos. A principal causa da baixa taxa de acertos deveu-se ao processo de calibração e filtragem do sinal que desloca a posição

dos picos em até 5 pontos para a direita ou para a esquerda. A utilização do Algoritmo Genético, apesar de apresentar uma boa taxa de acerto, depende também do processo de calibração e filtragem. Neste estudo o Algoritmo Genético foi otimizado para o conjunto de amostras de validação, entretanto, para outras amostras de validação poderá apresentar resultados diferentes.

Os dois métodos que apresentaram melhores resultados foram o da redução dos números de variáveis utilizando-se componentes principais e o da compressão dos espectros utilizando-se wavelets. O método do componente principal teve uma taxa de acertos de 90% quando os primeiros dois, quatro e oito componentes principais foram selecionados. O método de compressão do espectro por transformada discreta de wavelets apresentou a taxa de acertos de 100% para a taxa de compressão de 8:1 e 32:1 e 90% para a taxa de compressão para 16:1. A amostra classificada incorretamente pelo método da wavelets foi a mesma incorretamente classificada pelo método do componente principal.

Como estes dois últimos métodos apresentaram uma alta taxa de acertos e uma grande redução do número de variáveis para alimentar a rede neural, reduzindo desta forma a sua complexidade, eles foram selecionados para a próxima fase do projeto. Esta nova fase consistirá em determinar o menor tempo de exposição do CCD que resulte em uma taxa de acertos acima de 80%. Como estes dois métodos são pouco sensíveis ao ruído será também analisada a possibilidade de aplicá-los diretamente no espectro original sem filtragem, o que reduziria ainda mais o tempo de processamento.

Referências

- [1] Sathaiiah, S.; Silveira, L.; Zangaro, R. A.; Pacheco, M. T.; *Diagnosis of Human Coronary Artery with Near Infrared Raman Spectroscopy*, Proc. Int. Conf. on Raman Spectroscopy XV, USA, 1996.
- [2] Kortum, R.; Mehta, A.; Hayes, G.; Cothren, R.; *Spectral Analysis of Atherosclerosis Using an Optical Fiber Laser Catheter*, American Heart Journal, Vol. 118, pp. 381-391, Aug. 1989
- [3] Paula Jr., A. R.; Peris, C. M. F.; Sidaoui, H.; Sathaiiah, S.; *Digital Processing of Raman Spectra for Diagnosis of Atherosclerosis*, Third International Caracas Conference on Device, Circuits and Systems, Cancun, Mexico, Março de 2000, pg S75-1 to S75-6.
- [4] Haykin, S.; *Neural Networks*, Prentice Hall, New Jersey, USA, 1999, ISBN 0-13-273350-1
- [5] Goldberg, D. E.; *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison Wesley, 1989.
- [6] Kaiser, G., A., *A Friendly Guide to Wavelets*, Birkhäuser, Berlin 1994; ISB 0-8176-3711-7.
- [7] Peris, C. M. F.; Paula Jr., A. R.; *Análise Comparativa de Filtros Wavelets para Processamento de Sinais Raman*; Encontro Regional de Matemática Aplicada e Computacional, São José dos Campos, SP, Março de 2000, pg. 16.
- [8] Jackson, J. E., *A User's Guide to Principal Components*, John Wiley & Sons, 1991.