

PROCESSAMENTO DE ESPECTROS RAMAN DE MAMA HUMANA UTILIZANDO REDES NEURAIS ARTIFICIAIS

Walter Andrés Avendaño Jara¹, Ana Carolina de Miranda Marzullo¹, Renata Andrade Bitar¹, Alderico R. de Paula², Ailton Abrahão Martin¹

¹Laboratório de Espectroscopia Vibracional Biomédica/ IP&D-UNIVAP, Av. Shishima Hifumi, 2.911 – Urbanova, São José dos Campos – SP, Brasil, 12444-000. Fone: +55 3947- 1165, andres@univap.br, rabc@univap.br, acmarzullo@hotmail.com, amartin@univap.br

²Grupo de Instrumentação e Processamento de Sinais e Imagens – GIPSI / IP&D-UNIVAP, Av. Shishima Hifumi, 2.911 – Urbanova, São José dos Campos – SP, Brasil, 12444-000. Fone: +55 3947- 1140, alderico@univap.br

Resumo - Atualmente a espectroscopia Raman é uma das áreas mais ativas de pesquisa, podendo vir a ser utilizada na concepção de equipamentos para análise de tecidos biológicos fornecendo diagnósticos precisos e em tempo real. O objetivo deste trabalho foi processar espectros Raman, de mama humana, utilizando redes neurais artificiais, com a finalidade de proporcionar em um futuro próximo, o diagnóstico desta lesão, em tempo real. Através do sistema de espectroscopia FT-Raman RFS 100 (Bruker), que utiliza como energia de excitação, o laser de Nd:YAG, com o comprimento de onda de 1064 nm; 34 espectros de tecido de mama normal e 34 espectros de tecido de com doença fibrocística foram adquiridos. A banda espectral que apresentou maior diferença entre as classes, encontra-se na frequência Raman entre 800 a 1800 cm^{-1} totalizando 519 variáveis. Como resultado do trabalho obtivemos uma taxa classificatória de acerto de 100% entre as classes estudadas. Concluímos que espectroscopia Raman associada as redes neurais artificiais, possibilita o diagnóstico de câncer de mama com precessão.

Palavras-chave: Câncer de Mama, Espectroscopia Raman, Análise de Componentes Principais, Redes Neurais Artificiais.

Área do Conhecimento: III - Engenharias

Introdução

A espectroscopia Raman fornece informações dos níveis vibracionais moleculares. Na espectroscopia Raman a amostra é irradiada com laser. A energia pode ser então absorvida ou espalhada. Grande parte da luz espalhada tem a mesma frequência da luz incidente, no entanto, uma pequena fração da luz incidente interage com as moléculas, proporcionando informações, através de um espectro de intensidade em função da variação de frequência, consistindo em uma série de picos, que caracterizam a composição química da amostra [1-2].

Os avanços na área de computação permitem criar através de linguagens de programação, redes neurais artificiais, dando origem a um novo segmento na área de Inteligência Artificial; estas redes são capazes de reconhecer informações e produzir respostas seguindo a teoria de aprendizagem cognitiva, se adaptando a mudanças de controle, classificação e processamento ao longo do tempo [3].

O neurônio artificial é a peça chave para a construção das redes neurais artificiais, sendo possível identificar três elementos básicos no modelo de um neurônio. A região de conexão

entre as sinapses, que pode ter tanto valores positivos quanto negativos; região de junção das sinapses e a saída do sinal do neurônio, a qual varia em amplitude com relação à somatória dos sinais de entrada [4]. A Figura 1 ilustra, um neurônio artificial, com valores de entrada $[x_1, x_2, \dots, x_i]$, conjunto de pesos $[w_1, w_2, \dots, w_i]$, função de ativação f e saída y .

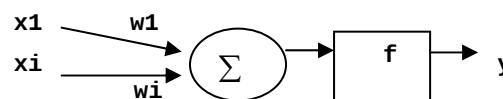


Figura 1: Neurônio Artificial.

A rede neural perceptron de camadas múltiplas (MLP - multilayer perceptron) é uma rede de múltiplos neurônios, do tipo discriminadores lineares e não lineares, organizados em camadas [5]. A Figura 2, visualiza uma rede perceptron de camadas múltiplas. Na camada de entrada (CE) os padrões são apresentados à rede. Nas camadas intermediárias (CI), a maior parte do processamento é realizado, elas podem ser consideradas como extratoras de características.

Na camada de saída (CS) as respostas dos padrões são apresentadas.

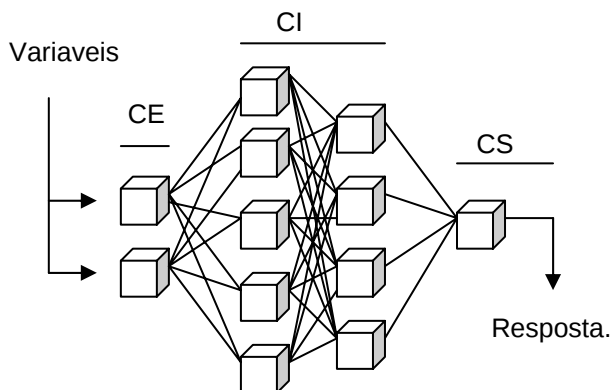


Figura 2: Rede Perceptron de Camadas Múltiplas.

Nossa proposta é processar espectros Raman, de mama humana normais e neoplásicas utilizando redes neurais artificiais, com a finalidade de proporcionar, em um futuro próximo, o diagnóstico de câncer de mama, em tempo real.

O câncer de mama foi escolhido neste estudo, por ser a neoplasia maligna de maior incidência entre as mulheres, apresentando a maior taxa de mortalidade no Brasil, entre todos os outros cânceres [6].

Para o processamento dos espectros Raman, três etapas foram desenvolvidas [7]:

- Um pré - processamento dos espectros.
- Redução do número de variáveis.
- Extração e classificação de características.

Na etapa de pré-processamento, os espectros foram organizados em duas matrizes; uma matriz de treinamento e uma matriz de validação. Foi definida a banda espectral e uma linha de base [8-9].

Para redução do número de variáveis foi utilizada a técnica de análise de componentes principais (PCA - Principal Components Analysis) [10].

A etapa de extração e classificação de características, foi realizada utilizando uma rede neural artificial, perceptron de camadas múltiplas.

Materiais e Métodos

Para realizar o processamento dos sinais, 34 espectros de tecido de mama normal do paciente A e 34 espectros de tecido com doença fibrocística do paciente B foram adquiridos, o procedimento foi aprovado pela Comissão de Ética e Pesquisa da Universidade do Vale do Paraíba (017/2000/CEP) e Hospital do Câncer A. C. Camargo (Nº 268/00).

O sistema, utilizado para aquisição dos espectros é o sistema FT-Raman RFS 100

(Bruker). A Tabela 1, descreve os parâmetros do equipamento, na aquisição dos espectros.

Tabela 1: Parâmetros do equipamento FT-Raman spectrometer RFS 100 (Bruker).

FT-Raman RFS 100	Parâmetros.
Resolução	4 cm ⁻¹
Laser Nd: YAG	1064 nm
Abertura do feixe	7 mm
Potência na amostra	132 mW

As amostras foram identificadas e conservadas em formol a 10% ao final da aquisição, sendo enviadas para análises histopatológicas.

Os espectros adquiridos foram organizados, na matriz de treinamento, composta por 2/3 do total de espectros por classe; e na matriz de validação, composta por 1/3 do total de espectros por classe. A banda espectral que apresentou maior diferença entre os tipos de tecido encontra-se na frequência Raman, entre 800 a 1800 cm⁻¹, totalizando 519 variáveis.

Uma linha de base foi definida, a partir da reta obtida pela equação $y - y_1 = (y_2 - y_1 / x_2 - x_1) (x - x_1)$. Na qual, primeira variável cartesiana do espectro (x_1 , y_1), correspondente ao deslocamento Raman 800 cm⁻¹; e a última variável cartesiana do espectro Raman (x_2 , y_2), correspondente ao deslocamento Raman 1800 cm⁻¹. A Figura 3 visualiza, a banda espectral e o processo de aquisição da linha de base, de um espectro de tecido de mama normal.

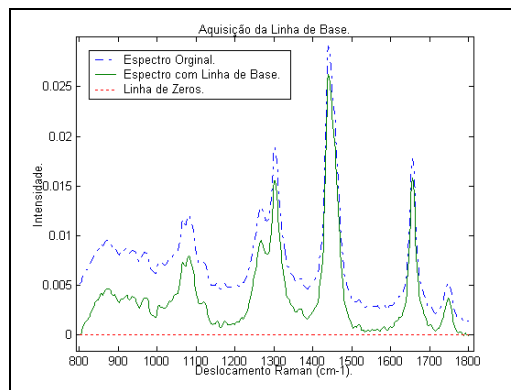


Figura 3: Banda Espectral e Aquisição da Linha de Base.

A técnica de PCA foi aplicada, aos espectros das matrizes de treinamento e de validação, para reduzir o número de variáveis, evitando redundâncias no processo de classificação. Na transformação linear representada pela matriz $V_k \in \mathcal{R}^{N \times K}$, onde N é o número das possíveis variáveis correlatas; a transformação compacta o vetor de entrada $x \in \mathcal{R}^{N \times 1}$ em um vetor

$y \in \mathbb{R}^{K \times 1}$, $K \leq N$. As novas variáveis K , não são correlacionadas, e são denotadas como componentes principais (PC). A Figura 4, visualiza a primeira componente principal, em função da segunda componente principal, da matriz de treinamento.

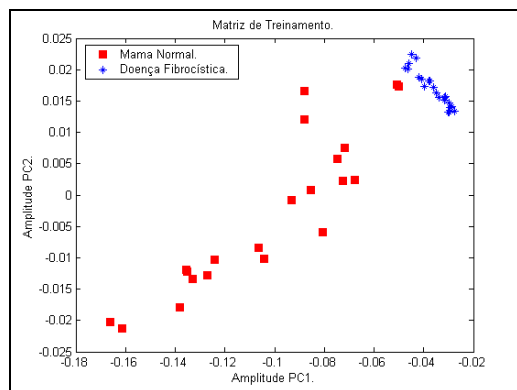


Figura 4: Matriz de Treinamento.

A Figura 5 visualiza, a primeira componente principal em função da segunda componente principal da matriz de validação.

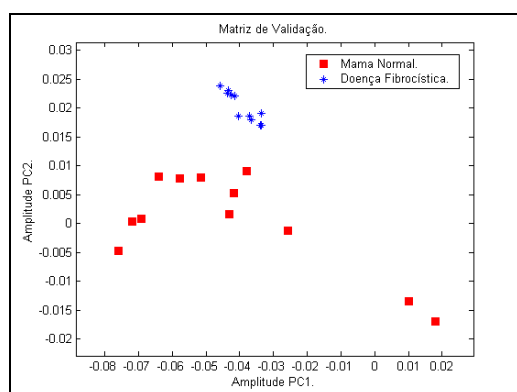


Figura 5: Matriz de Validação.

Para treinar a rede neural artificial, foram utilizadas as duas primeiras componentes principais da matriz de treinamento. Após o treinamento a rede neural, foi validada utilizando as duas primeiras componentes principais, da matriz de validação.

Resultados

A Tabela 2 e a Tabela 3 apresentam, a taxa de acerto dos grupos de treinamento e validação, respectivamente.

Tabela 2: Taxa de Acerto do Grupo de Treinamento.

Grupo de Treinamento.	Taxa de Acerto (%).
Mama Normal	100
Doença Fibrocística.	100

Tabela 3: Taxa de Acerto do Grupo de Validação.

Grupo de Validação.	Taxa de Acerto (%).
Mama Normal	100
Doença Fibrocística.	100

Discussão

Tendo em vista os resultados deste trabalho, temos como objetivo seguinte, aumentar o número de pacientes por classe, obtendo segurança estatística. Além de incluir as demais classes; classificando através das redes neurais artificiais, espectros Raman de tecido de mama humana com carcinoma ductal infiltrante, carcinoma ductal in situ, carcinoma ductal in situ comedo, carcinoma lobular infiltrante, carcinoma ductal infiltrante inflamatório, carcinoma ductal infiltrante medular, carcinoma ductal infiltrante musinoso, incluindo os espectros de tecido de mama humana já classificados. Assim, a análise qualitativa das nove fases que compõem o estudo do câncer de mama poderá ser concluída, partindo para a fase de análise *in vivo*.

Conclusão

O câncer de mama é um mal que atinge muitas mulheres em todo mundo, ocasionando óbitos, traumas físicos e psicológicos. Atualmente, a forma existente, para diagnosticar o câncer de mama é a biópsia. Esta tem um papel importante no processo de diagnóstico, porém existem muitos inconvenientes, tais como, o tempo de espera e a precisão do exame, observado de forma crítica, por depender de diversos componentes químicos e da experiência de observação do profissional que examina as lâminas histopatológicas. A questão burocrática também deve ser salientada, pois o exame passa por diversos processos antes do resultado ser obtido. Cada processo exige documentações específicas, ocorrendo desgastes físicos, emocionais e econômicos para o paciente, familiares e amigos. Diante da atual realidade, nossa proposta de trabalho, mostra-se como uma solução inovadora, que possibilita a classificação em tempo real, permitindo ao cirurgião, determinar com precisão os tecidos cancerosos, que serão retirados. Este método poderá ser estendido para diversos outros tipos de cânceres, tais como, câncer de colo de útero, câncer de pele, entre outros.

Referências

- [1] H. Kuzmany, Solid-state spectroscopy, Springer, Berlin, 1998.
- [2] SCHRADER B., DIPPEL, B., NIR Raman spectroscopy in medicine and biology: results and aspects, J. Mol. Struct. 480/481: 21-32. 1999.
- [3] MALCON A. MARCOS DE XEREZ. Redes Neurais Artificiais Introdução e Princípios de Neurocomputação, Editora Eko, 1996.
- [4] ALEKSANDER,I.,MORTON H. An Introduction to Neural Computing. Inglaterra, Londres: Chapman Hall,1990.
- [5]S. HAYKIN, Neural Network: A Comprehensive Foundation, Prentice Hall, Second Edition, New Jersey, 1999.
- [6] INCA - Instituto Nacional de Câncer – disponível em <http://www.inca.gov.br> . Acesso em 08/11/2004.
- [7] M. FRIEDMAN, and A. KANDEL, Introduction to Pattern Recognition, Kandel-Word Scientific, 1999.
- [8] BITAR, R. A. C.; MARTIN, A. A.; MOURÃO NETO, M.; SOARES, F.A.; “FT-Raman Spectroscopy Study of Human Breast Tissue. Biomedical Vibrational Spectroscopy, Estados Unidos, v. 5321, n. 7, p. 190-195, 2004.
- [9] SIMMONS, G. F., Cálculo com Geometria Analítica, vol. 1, Makron Books do Brasil Editora Ltda, S. P., 1987.
- [10] J.E. JACKSON, A User Guide to Principal Components, John Wiley & Sons, 1991.