

Diagnóstico de Câncer de Cabeça e Pescoço Baseado em Redes Neurais e Dados de Impedanciometria

¹Carlos Eduardo Ferrante do Amaral, ¹Marcos Santos Hara, ¹Heitor Silvério Lopes,
¹Lúcia Valéria Ramos de Arruda, ²Antonio José Gonçalves, ²Adilson Alves Dias

¹Laboratório de Bioinformática / CPGEI / CEFET-PR
Av. 7 de setembro, 3165 – 80230-901 – Curitiba (PR)

²Faculdade de Ciências Médicas da Santa Casa de São Paulo
R. Cesário Motta Júnior, 112 - Vila Buarque – São Paulo (SP)
e-mail: hslopes@cpgei.cefetpr.br

Abstract

This work reports the use of Neural Networks for classification of healthy and cancerous patients (head and neck cancer), based on anthropometrical and bioimpedance data. Using the proposed system, an accuracy rate of 85% was achieved, demonstrating the viability of this method. This result encourages the utilization of Neural Networks as classifiers using patient's bioimpedance data as input. This technique allows early diagnose of cancer, when the prognosis can be better, therefore changing the course of disease or suggesting a better treatment option.

1. Introdução

Apesar da técnica de medida de bioimpedância já ser conhecida há algum tempo, as aplicações deste método ainda não foram completamente estudadas e suas bases biofísicas ainda não são claramente compreendidas [1] e [2]. Sabe-se, no entanto, que em certas doenças, como câncer, a impedância do paciente sofre grandes distúrbios [3] e [4]. Em alguns casos pode-se detectar mudanças neste parâmetro antes mesmo que a pessoa apresente sintomas fisiológicos, porém ainda não se conseguiu determinar uma lei que estabeleça uma relação entre doenças e a bioimpedância do paciente.

Atualmente, a principal aplicação da bioimpedância é a determinação da composição corporal com a ajuda de *softwares* especializados. Porém, observou-se que as alterações nos parâmetros de bioimpedância do corpo humano, causadas por muitas doenças levavam a uma falha na determinação da composição corporal. Isto levou à hipótese de que os valores de bioimpedância poderiam ser também utilizados como fator de diagnóstico de determinadas doenças.

A falta de modelo matemático e o grande número de variáveis que influenciam este sistema foram fatores que levaram à escolha da utilização de redes neurais. A proposta do trabalho é procurar estabelecer uma metodologia para diferenciar pacientes portadores de

tumores e pacientes saudáveis, com base nos valores de bioimpedância.

2. Impedanciometria

Quando uma diferença de potencial é aplicada a um material condutor haverá a circulação uma corrente elétrica cuja intensidade depende da resistividade do material. Em circuitos elétricos formados pela associação de resistores, capacitores e indutores, a oposição circulação de corrente é dada por uma impedância (Z), cujo valor é uma função da frequência. Matematicamente, impedância é uma grandeza complexa formada por um módulo e um ângulo de defasagem entre a tensão e a corrente aplicadas. Também é possível representá-la na forma retangular através de sua componente resistiva (R) no eixo real, e de suas componentes reativas (X_C ou X_L) no eixo imaginário [5]. A relação entre a impedância de um circuito com um capacitor e um resistor em série na forma polar e retangular é mostrada nas equações (1) e (2) e sua representação gráfica é dada na figura 1:

$$Z = \sqrt{R^2 + X_C^2} \quad (1)$$

$$\theta = \arctan \frac{X_C}{R} \quad (2)$$

onde: Z é a impedância (Ω), X_C é a reatância capacitiva (Ω), R é a resistência (Ω), θ é o ângulo de defasagem ($^\circ$).

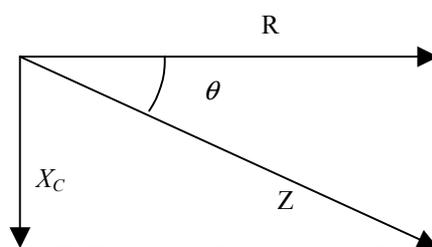


Figura 1: Exemplo de representação de uma impedância na forma retangular e polar

O módulo da impedância (Z) pode ser calculado pela razão da tensão pela corrente que percorre certo elemento. O ângulo θ pode ser obtido pela defasagem que a corrente apresenta em relação à tensão.

Considerando o corpo humano como um material condutor, a bioimpedanciometria consiste em submeter pacientes a baixas correntes elétricas (da ordem de centenas de microampères) e calcular o valor de impedância. Esta impedância deve variar de acordo com o estado clínico do paciente, sendo um reflexo direto da quantidade de água intra e extracelulares, bem como do estado das membranas celulares. Neste trabalho desenvolve-se um modelo por Redes Neurais que relaciona a bioimpedância humana à presença de tumores cancerígenos.

2.1. Técnica de Medição

No caso de tecidos vivos a condução elétrica não é feita por elétrons, mas através do movimento de íons. Para medida da bioimpedância, utiliza-se então corrente alternada para evitar o efeito da polarização comum em substâncias iônicas [6].

Neste trabalho, para a obtenção dos valores de impedância em pacientes utilizou-se dois aparelhos BIA-101A (RJL Systems Inc.), que geram uma corrente de 800 μ A a 50 KHz. Foram utilizados eletrodos descartáveis e a técnica tetrapolar de colocação dos eletrodos. O aparelho mede diretamente R e X_C , enquanto que θ foi calculado diretamente.

2.2. Coleta de Dados

As medidas de impedância foram realizadas em pessoas saudáveis e em pessoas com tumores. Além da impedância foram registrados os valores da idade, peso, altura e sexo. O índice de massa corporal (IMC) foi calculado dividindo-se o peso pelo quadrado da altura [7].

A Tabela 1 mostra alguns exemplos de valores medidos para pessoas saudáveis. No total, foram coletados dados de 463 indivíduos saudáveis. Os dados foram obtidos de indivíduos normais da população brasileira. Os valores do peso e da altura foram substituídos pelo IMC.

Tabela 1: Exemplos dos dados de pessoas saudáveis

Sexo *	Idade (anos)	Resistência (Ω)	Reatância (Ω)	Ângulo de fase (graus)	IMC (kg/m^2)
0	31	558	66	6,745	22,031
0	20	725	65	5,123	18,314
1	21	544	73	7,642	21,604
0	32	619	82	7,546	20,202
1	22	488	61	7,125	26,023
1	57	414	52	7,159	25,529

* 0 = feminino, 1 = masculino

A Tabela 2 mostra alguns valores obtidos para pessoas portadoras de câncer de cabeça e pescoço. Ao todo foram obtidos dados de 80 pacientes, do Serviço de Cirurgia de Cabeça e Pescoço da Santa Casa de São Paulo. Para este trabalho, optou-se por não diferenciar os vários tipos de tumores, como câncer de traquéia, tireóide, laringe, lábios e outros órgãos. A coluna diagnóstica é classificada da seguinte forma:

- (1) Pacientes que foram submetidos à cirurgia, mas não apresentaram complicações;
- (2) Pacientes que foram submetidos à cirurgia e apresentaram complicações leves;
- (3) Pacientes que foram submetidos à cirurgia e apresentaram complicações graves;
- (4) Pacientes que foram submetidos à cirurgia e vieram a óbito;
- (5) Pacientes com tumores irresssecáveis;
- (6) Pacientes com câncer terminal.

Tabela 2: Exemplos dos dados para pessoas com tumores, com o respectivo tipo de diagnóstico

Sexo *	Idade (anos)	Resistência (Ω)	Reatância (Ω)	Ângulo de fase (graus)	IMC (kg/m^2)	Diagnóstico
0	70	537	33	21,333	25,171	2
1	59	395	27	20,199	19,554	3
0	30	634	73	23,140	27,397	1
1	78	210	11	28,719	7,312	4
1	65	534	48	17,578	30,378	6
0	64	735	61	20,415	36,002	5

* 0 = feminino, 1 = masculino

3. Redes Neurais Artificiais

Redes Neurais Artificiais (RNA) são sistemas paralelos distribuídos que tentam modelar funções matemáticas, em geral não-lineares, a partir da conexão massiva de unidades de processamento simples denominadas de neurônios. Estes são dispostos em camadas e interligados por um grande número de conexões.

A solução de um problema via redes neurais pressupõe inicialmente uma fase de aprendizado em que um conjunto de dados é apresentado à rede, a qual extrai automaticamente as características necessárias para representar a informação fornecida. Entre as aplicações de redes neurais, a classificação de padrões é uma das tarefas mais usuais. Em problemas desta natureza, cada vetor de entrada pertence ou não a uma determinada categoria [8].

Neste trabalho, deseja-se usar uma Rede Neural para diferenciar pessoas saudáveis de pessoas portadoras de câncer. Já se comprovou que o valor da impedância total de um indivíduo pode variar com parâmetros como a umidade, estado psicológico, idade, sexo, atividade hormonal, peso, altura, etc... O objetivo deste trabalho é mostrar que o valor da impedância é também influenciado pela presença ou não de tumores

possibilitando, assim, seu uso como ferramenta de diagnóstico. Os parâmetros de entrada disponíveis para a rede neural estudada foram: sexo, idade, IMC, resistência, capacitância e ângulo de fase.

Dentre os diversos tipos e topologias de Redes Neurais existentes, foi utilizado a rede Perceptron de Múltiplas Camadas (MLP) com treinamento tipo *Backpropagation*. Nesta forma de treinamento aplica-se a técnica do gradiente descendente para minimizar o erro médio quadrático total (MSE) da saída. O treinamento envolve três fases: a resposta na saída da rede a sinais de entrada, o cálculo do erro associado e o ajuste dos pesos. Após o treinamento, a aplicação da rede envolve somente o processamento das entradas, com tempo de resposta bastante rápido.

4. Análise Estatística dos Dados de Entrada

Redes Neurais são ferramentas eficientes para a construção de modelos a partir de dados que possuam alguma correlação entre si. Elas são capazes de definir *clusters* para certas classes associadas aos dados. Por este motivo realizou-se uma análise estatística dos dados a fim de verificar a existência de eventuais *clusters* e o grau de colinearidade.

A análise descritiva das variáveis de entrada é mostrada no diagrama da figura 2, para cada uma das duas classes. O diagrama evidencia que a análise isolada de qualquer um dos parâmetros de entrada é incapaz de classificar inequivocamente um determinado paciente, sendo que os dados das duas classes de diagnósticos se sobrepõem.

Para que não haja sobreposição da importância de nenhuma variável de entrada no treinamento da rede, todos eles foram normalizados com relação ao valor máximo de todas as amostras. Seguindo o comportamento previsto de uma rede neural, caso exista uma relação matemática entre as seis variáveis de entrada e as duas classes de saída, a Rede Neural deverá encontrá-la.

5. Simulações e Resultados

Devido à pequena quantidade de dados de pacientes com câncer, não foi possível realizar uma análise discriminada de cada tipo de diagnóstico. A separação dos diagnósticos possíveis em apenas duas classes simplifica o problema de classificação. Foram utilizados os 80 casos disponíveis da classe de pacientes portadores de câncer (classe 1). Para a classe de pessoas saudáveis (classe 2), também foram utilizados 80 casos, extraídos aleatoriamente do conjunto de dados disponível. O uso de classes balanceadas (mesmo número de casos) torna o treinamento da Rede Neural menos tendencioso. A partir destes 160 casos, foi definido o conjunto de dados de treinamento (120 casos) e de teste (40 casos), sendo 50% de cada classe.

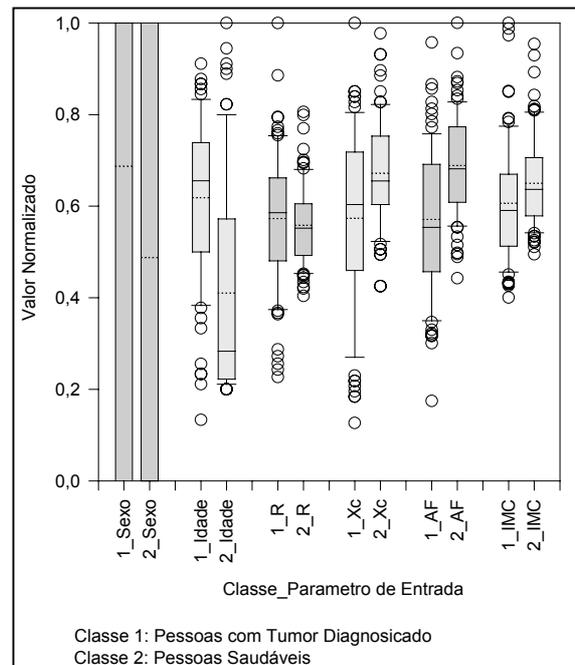


Figura 2: Análise descritiva das variáveis de entrada. BoxPlots evidenciando as distribuições dos percentis P10, P50, P90, a média, e os dados acima e abaixo destas distribuições

Para o treinamento e teste da Rede Neural MLP, foi utilizado o *software* SNNS (*Stuttgart Neural Network Simulator*). Duas topologias foram testadas, a 6-13-2 (seis neurônios na camada de entrada, treze na camada intermediária e dois na camada de saída) e 6-13-1. Mesmo tendo apenas um neurônio na camada de saída, esta última rede pode ser capaz de gerar respostas que podem ser classificadas em diferentes de classes, dependendo da maneira como se interpreta o valor contínuo da saída. No treinamento de ambas as topologias, os pesos iniciais foram valores aleatórios na faixa de $[-1,+1]$.

Para a rede MLP 6-13-1, a classificação é dada pelo valor contínuo da ativação do neurônio de saída. A classe “1” é identificada por ativações com valor menor do que 0,5 e a classe “2” por valores maiores ou iguais a “0,5”. No treinamento as duas classes foram representadas pelos valores “0” e “1”, respectivamente. Para esta rede foram utilizados os seguintes parâmetros: 10000 épocas de treinamento, coeficiente de aprendizado $\alpha=0,2$, MSE obtido no treinamento 0,102 e MSE obtido no teste 0,172.

O percentual de acerto simples obtido pela rede MLP 6-13-1 foi de 83% para o conjunto de teste (a rede classificou erroneamente seis casos da classe 2 como sendo da classe 1 e um caso da classe 1 como sendo da classe 2).

Para a rede MLP 6-13-2, a classificação é obtida diretamente pela ativação do neurônio na camada de saída correspondente, pois para o treinamento, os vetores de resposta desejada eram compostos por um valor “1” no neurônio que representava o diagnóstico correto, e valores “0” no outro. Uma vez que a saída é um valor contínuo no intervalo $[0,1]$, o neurônio da

camada de saída com o maior valor de ativação representa a classe associada ao vetor de entrada. Para esta rede foram utilizados os seguintes parâmetros: 12500 épocas de treinamento, coeficiente de aprendizado $\alpha=0,2$, MSE obtido no treinamento 0,12 e MSE obtido no teste 0,38.

O percentual de acerto simples obtido pela rede MLP 6-13-2 foi de 85% para o conjunto de teste (a rede classificou erroneamente 2 casos da classe 2 como sendo da classe 1 e quatro casos da classe 1 como sendo da classe 2).

A título de ilustração, a figura 3 mostra a ativação da rede para um determinado vetor de teste apresentado em sua entrada, onde a classe 1 foi indicada como diagnóstico.

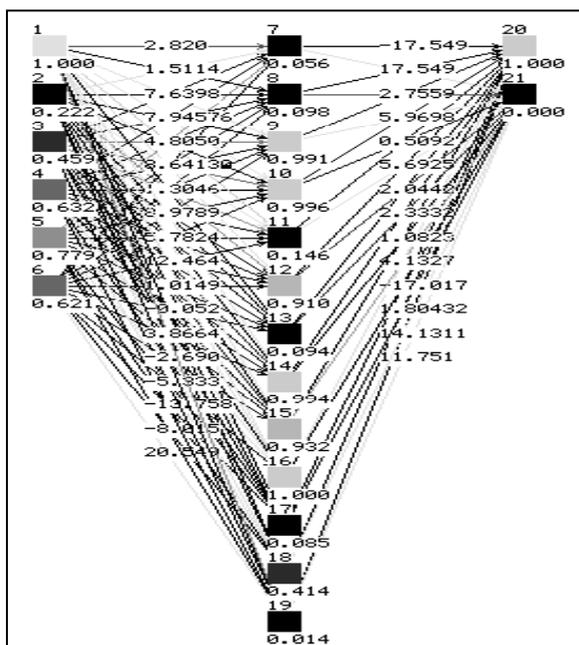


Figura 3: Ativação dos neurônios da rede 6-13-2 para um determinado vetor de entrada de teste

6. Conclusões

Para classificação em apenas dois possíveis diagnósticos, as simulações realizadas para as redes propostas apresentaram um índice de acerto de cerca de 85%. Considerando a pequena quantidade de dados disponível para treinamento, a aplicação de redes neurais MLP com *Backpropagation* neste tipo de problema pode ser considerada bem sucedida. O índice de acerto obtido é satisfatório, considerando-se a variabilidade biológica e a margem de erro das medidas de bioimpedância.

Uma classificação mais detalhada, na qual a saída da rede apontaria qual o prognóstico da evolução do quadro clínico do paciente (seção 2.2), parece também ser possível, desde que haja uma quantidade significativa de casos de treinamento/teste para cada classe. Este trabalho mostra que a utilização de Redes Neurais como classificadores, utilizando dados de

bioimpedância, pode permitir diagnósticos e prognósticos precoces de tumores aumentando assim a possibilidade de cura ou fornecendo uma melhor opção de tratamento ao paciente.

Referências

- [1] J.R. Pere, J. Rosell, R. Bragós, O. Casas. *Electrical Bioimpedance Methods: Applications to Medicine and Biotechnology*. New York Academy of Sciences, New York, 1999.
- [2] C. Ren, H. Wang, Y. An, S. Hong, G. Lin. Development of electrical bioimpedance technology in the future. In *Proceedings of IEEE/EMBS 20th Annual International Conference*, vol. 2, pages 1052 – 1054, 1998.
- [3] B. Blad, E.B. Baldetorp. Impedance spectra of tumor tissue in comparison with normal tissue, a possible clinical application for electrical impedance tomography. *Physiological Measurements*, 17:A105–A115, 1996.
- [4] B. Blad, P. Wendel, M. Jönsson, K. Lindström. An electrical impedance index to distinguish between normal and cancerous tissues. *Journal of Medical Engineering & Technology*, 23(2):57–62, 1999.
- [5] J. L. Rudolph. Principles of bioelectrical impedance analysis. Abril, 1997. Disponível em: <http://www.rjlsystems.com/research/bia-principles.html>. Acessado em 22/07/2002.
- [6] L. A. Geddes, L. E. Baker. *Principles of Applied Biomedical Instrumentation*. 3rd ed. John Wiley & Sons, New York, 1989.
- [7] H. H. Bendixen. Morbid obesity. *Refreshment Courses on Anesthesiology - ASA*, 6, 1978.
- [8] L. Fausett. *Fundamentals of Neural Networks, Architectures, Algorithms and Applications*. Prentice Hall, Englewood Cliffs, 1994.

Rev. 10/Abr/2003