



## **Caracterização e detecção automática de eventos epileptiformes em sinais de eletroencefalograma por dinâmica simbólica.**

**Lynwood Livi de Souza<sup>1</sup>**

<sup>1</sup>Mestrando do Programa de Pós-Graduação em Ciências Mecânicas - UnB. Bolsista da CAPES. Docente do Instituto Federal de Goiás – Câmpus Uruaçu. e-mail: lynwood@ifg.edu.br

**Resumo:** O presente trabalho apresenta uma nova metodologia para caracterização, detecção e classificação de sinais de exames de eletroencefalografia (EEG), visando o aprimoramento e agilidade na obtenção de diagnósticos com o objetivo de auxiliar o profissional de saúde, uma vez que os exames de EEG são caracterizados por longos trechos de registros temporais da atividade elétrica do cérebro, que devem ser minuciosamente examinados pelos médicos. Devido à dificuldade associada à caracterização e à detecção de sinais epileptiformes e à importância do diagnóstico, existem na literatura vários métodos desenvolvidos para abordar este problema, tais como: Métodos Auto-Regressivos (AR), Aproximação da Energia (TEO), Análise de Domínio de Freqüência, Análise de Tempo-Freqüência, Transformada de Wavelets e métodos não-lineares como os modelos estatísticos. Porém, nenhum modelo ainda não foi capaz de atingir uma performance inteiramente confiável em seus resultados, onde a média de acertos entre os trabalhos expostos na literatura fica em torno de 80,7%, um índice de acertos que apesar de significativo, ainda não é satisfatório para este tipo de exame. A fim de contribuir para o estudo do problema, propôs-se neste trabalho a aplicação da Dinâmica Simbólica para caracterização dos sinais, que se baseia na representação por seqüências de símbolos do estado do sistema e operadores de mudança de estado, e algoritmos genéticos para a otimização da representação, de forma que os sinais epileptiformes pudessem ser distinguidos dos sinais considerados normais por uma rede neural artificial, treinada para este fim. Os resultados obtidos demonstram que a metodologia fornece 92,4% de precisão e 96% de acerto.

**Palavras-chave:** ALGORITMOS GENÉTICOS, DINÂMICA SIMBÓLICA, EEG, REDES NEURAIS.

### **1. INTRODUÇÃO**

O grande desafio do homem nos últimos tempos vem sendo as descobertas da cura de patologias que levam à morte ou comprometem a qualidade de vida. Nesse sentido, juntamente com os grandes males, surgem as necessidades de inovações para detecção rápida e eficiente destes males, pois com agilidade e precocidade na definição do diagnóstico, certamente o tratamento se mostra mais eficaz levando à obtenção de resultados mais satisfatórios.

O avanço da medicina ocorre paralelo ao da tecnologia, que vem se desenvolvendo rapidamente, e suas inovações vem trazendo resultados surpreendentes, e ao mesmo tempo, modificando e ampliando as formas de prevenção, tratamento e cura das mais variadas doenças. Portanto, esta pesquisa propõe um novo método para auxiliar na identificação e tratamento de um mal que afeta a vida de milhares de pessoas em todo o mundo, denominada de epilepsia.

Nosso cérebro é constituído por neurônios, que estão interligados entre si trocando informações através de impulsos elétricos, eventualmente há uma atividade elétrica anormal entre esses neurônios, onde o resultado visível desta anormalidade são as crises intermitentes. Quando essas crises acontecem freqüentemente em uma pessoa, diz-se que ela possui epilepsia. Uma crise é uma situação temporária, onde no caso da crise epiléptica o cérebro é tomado por descargas elétricas rápidas e fortes. Estas crises são caracterizadas principalmente por ataques onde a pessoa fica inconsciente, não se lembrando do acontecimento e também por dormência temporária nos membros.

A epilepsia é a desordem neurológica mais comum na população mundial, podendo ser detectada e avaliada através do exame de eletroencefalografia (EEG), onde as características principais para existência de epilepsia são a presença de espículas agudas no registro dos sinais do exame.



Os resultados provenientes do exame de EEG para detecção de presença ou não de sinais epiléticos são feitos visualmente pelo profissional, tomando muito tempo para a avaliação dos resultados. Na maioria dos exames, que têm duração média de trinta minutos, podem conter até quarenta páginas de registros gráficos, não sendo às vezes suficientes para um diagnóstico conclusivo, podendo haver necessidade de se alongar a duração do exame em duas a três horas. Em casos mais complexos, pode-se haver necessidade de uma observação por meio de filmagem de dois a três dias para um laudo conclusivo.

Haja em vista a dificuldade de se obter um laudo conclusivo com agilidade devido à extensão do exame visual dos registros, esse trabalho propõe um software capaz de resolver esta questão de forma automática dos sinais de EEG. Utilizando-se o método de simbolização otimizada por algoritmos genéticos de séries temporais e o treinamento de uma rede neural artificial onde serão classificados sinais epiléticos e não epiléticos, podendo vislumbrar toda dinâmica das atividades cerebrais de pessoas com epilepsia ou não. Ressalta-se que o software proposto como resultado do estudo realizado conta com uma interface gráfica amigável para ser utilizada pelo usuário final. Pretende-se, dessa forma, contribuir para o diagnóstico mais ágil e preciso para o paciente, bem como, para a redução do seu custo.

O objetivo principal deste trabalho foi desenvolver um método capaz de avaliar os sinais de EEG, utilizando dinâmica simbólica para a descrição das séries temporais, classificando esses resultados em uma rede neural artificial e finalmente implementar a metodologia em um software capaz de gerar um resultado automático, que possa ser facilmente avaliado pelo usuário final, de modo a ser testado e validado experimentalmente por profissionais da área de eletroencefalografia.

## 2. MATERIAL E MÉTODOS

A obtenção dos dados de exames eletroencefalográficos foi definida pela separação entre os dados de resultados positivos e de resultados negativos. A partir daí realizou-se a simbolização das séries temporais para cada sinal a fim de avaliar os dados gerados para que seja montada uma base de dados com as características principais e eminentes de cada tipo de exame: positivo e negativo. A separação dos eventos epileptiformes dos sinais normais foi realizada pelo próprio software desenvolvido, a partir deste momento a base de dados com duas características distintas encontraram-se aptas a dar seqüência aos demais processos descritos neste trabalho. A parte selecionada será a parte do sinal a qual será salva na base de dados, notam-se claramente as características exploradas, após isso salvamos estes sinais para aplicarmos à simbolização, onde através da simbolização das séries temporais vamos abstrair características claras que diferenciem ambos os sinais.

Para análise e caracterização dos sinais advindos dos exames de EEG, utilizou-se o método de simbolização de séries temporais experimentais descritos por (DAW et al, 2003), o método consiste em transformar a série temporal em uma seqüência de  $n$  símbolos, baseados em um alfabeto. Por exemplo, utilizando um alfabeto binário ( $n=2$ ), a cada medida disponível nos dados são atribuídos os símbolos "0" ou "1", segundo o valor da medida esteja abaixo ou acima de um dado limiar, respectivamente, como mostrado na Figura 1, para o caso de uma série temporal periódica (senoidal). Em seguida, uma seqüência de símbolos de comprimento  $m$  é escolhida, e as freqüências relativas de todas as possíveis seqüências de símbolos de dimensão  $m$  são calculadas para a seqüência completa obtida da discretização dos sinais experimentais (ver Tabela 1).

Pode-se representar o comprimento  $m$  por seu equivalente decimal, por exemplo, a seqüência  $011_2$  (base binária) por  $3_{10}$  (base decimal), com isto, pode-se representar a distribuição estatística das seqüências simbólicas como um histograma, mostrado na Figura 2, no qual o eixo horizontal corresponde às seqüências simbólicas e o eixo vertical a freqüência relativa das seqüências.

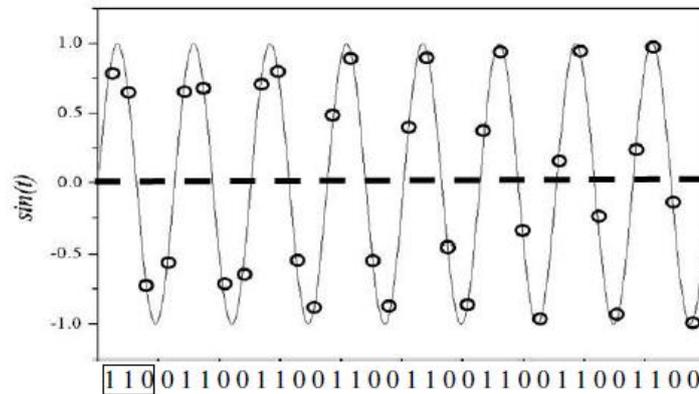


Figura 1 – Simbolização de série temporal

Para dados puramente aleatórios, todas as seqüências simbólicas de comprimento  $m$  são equiprováveis. Portanto, o distanciamento deste comportamento indica a evidência de estrutura determinística nos dados, podendo ser epiléptica ou não.

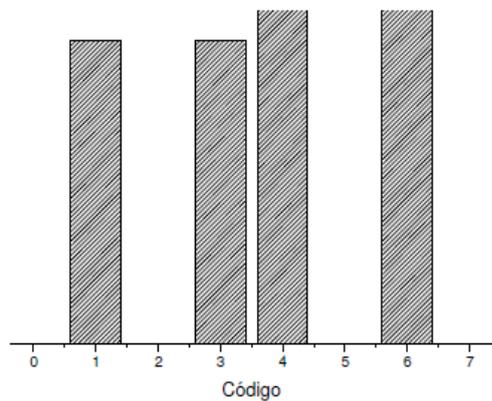


Figura 2 - Distribuição de frequência absoluta para uma série temporal periódica.

Seqüência	Código	Frequência
000	0	0
001	1	7
010	2	0
011	3	7
100	4	8
101	5	0
110	6	8
111	7	0

Tabela 1 - Frequência das seqüências obtidas na Figura 2.



No contexto da simbolização existem quatro parâmetros os quais são de suma importância para que a simbolização possa ser executada em uma ótima performance. Estes parâmetros são: Tamanho do Alfabeto (TA), Tamanho da Palavra (TP), Tamanho da Janela (TJ) e Passo da Janela (PJ). Onde o Tamanho do Alfabeto define quantos símbolos serão utilizados para simbolização das séries, o Tamanho da Palavra implica a quantidade de símbolos que formarão a palavra, o Tamanho da janela expressa a seção do sinal que será simbolizada e o Passo da Janela é intervalo de tempo entre uma posição da janela temporal e a posição subsequente durante a simbolização do sinal.

Dependendo dos valores atribuídos aos parâmetros pode ocorrer de não ser possível uma distinção, no contexto da simbolização, entre os sinais com epilepsia e sem epilepsia. Para resolver esse tipo de problema, uma ferramenta de otimização é implementada, neste caso os Algoritmos Genéticos irão otimizar os parâmetros necessários para a simbolização das séries. O objetivo é encontrar os parâmetros que proporcionem a melhor distinção entre os sinais com eventos epiléticos e não epiléticos.

No algoritmo de otimização, implementou-se a Entropia de Shannon, uma medida estatística capaz de encontrar graus de incerteza entre os valores utilizados, assim, quanto menor o valor do índice de Shannon menor o grau de incerteza, portanto, comparando os dois tipos de sinais quanto maior a diferença do grau de incerteza entre eles, melhor é a diferenciação entre os sinais, conseqüentemente, mais adequados são os parâmetros utilizados para a simbolização das series temporais naquele instante.

$$H_s(L) = - \frac{1}{\log N_{obs}} \sum_i P_{i,L} \log_{P_{i,L}}$$

Onde  $N_{obs}$  é o total de número de colunas do histograma com valor diferente de zero e  $P_{i,L}$  é a frequência relativa dos valores de cada coluna do histograma.

Para cada geração de indivíduos criada pelo Algoritmo Genético será calculada a sua entropia onde encontra-se para cada um sua aptidão, assim somaremos as aptidões de todos os indivíduos de certa geração comprovadamente epiléticos e outra não epiléticos, calculando em seguida a diferença entre eles.

$$Dif = \left| \int i_n - \int i_{ep} \right|$$

Após esse processo o algoritmo genético irá encontrar a população com a maior aptidão, isso significa que a população possui uma maior diferença entre epiléticos e não epiléticos. Através desta otimização o Algoritmo Genético nos informará os valores dos parâmetros utilizados para simbolização a qual obteve uma melhor diferença entre os sinais.

Para classificação e detecção automática da presença de eventos epileptiformes nos sinais de exames de eletroencefalografia, utilizamos as Redes Neurais Artificiais (RNA), um dos métodos mais utilizados nas literaturas para o tipo de finalidade deste trabalho.

As avaliações podem ser representadas de acordo com as seguintes medidas especificadas abaixo:

- VERDADEIROS POSITIVOS (VP): eventos que foram identificados pelo sistema como eventos epiléticos e confirmados pela marcação de um especialista.
- FALSOS NEGATIVOS (FN): eventos que foram identificados pelo sistema como não sendo eventos epiléticos, porém, os especialistas marcaram como sendo epiléticos.
- VERDADEIROS NEGATIVOS (VN): eventos que foram identificados pelo sistema como não sendo eventos epiléticos e confirmados pelos especialistas.
- FALSOS POSITIVOS (FP): eventos que foram identificados pelo sistema como sendo eventos epiléticos, mas não sendo confirmados pelos especialistas.

Existem outras formas de medidas de desempenho para expressar a relação entre eventos epiléticos e não epiléticos (SUBASIA, *et al*, 2005), tais como: sensibilidade (relação verdadeira positiva) que avaliam a precisão do sistema em detectar eventos epiléticos e a especificidade (relação verdadeira negativa) avaliam a precisão do sistema em detectar eventos não epiléticos.

$$\text{Sensibilidade} = \frac{VP}{VP + FN} \times 100[\%] \quad \text{Especificidade} = \frac{VN}{VN + FP} \times 100[\%]$$

A seletividade positiva, ou valor preditivo positivo (VPP), é a capacidade do sistema em indicar corretamente os sinais epiléticos. A seletividade negativa, ou valor preditivo negativo (VPN), é a capacidade do sistema em indicar corretamente os sinais não epiléticos.

$$\text{Seletividade positiva} = \text{VPP} = \frac{VP}{VP + FP} \times 100[\%]$$

$$\text{Seletividade negativa} = \text{VPN} = \frac{VN}{VN + FN} \times 100[\%]$$

Para estabelecer capacidade do sistema em reconhecer os sinais epiléticos e os sinais não epiléticos é aplicado o quantificador para reconhecer o sucesso do sistema, esse quantificador é denominado Precisão (SUBASIA, *et al*, 2005).

$$\text{Precisão} = \frac{\text{Sensibilidade} + \text{Especificidade}}{2}$$

### 3. RESULTADOS E DISCUSSÃO

A princípio, o primeiro passo adotado para o início da simbolização foi a separação dos eventos epiléticos, onde separou-se para a simbolização, os eventos característicos para que o sistema possa utilizar posteriormente em uma comparação, formando a base de dados com duas características distintas.

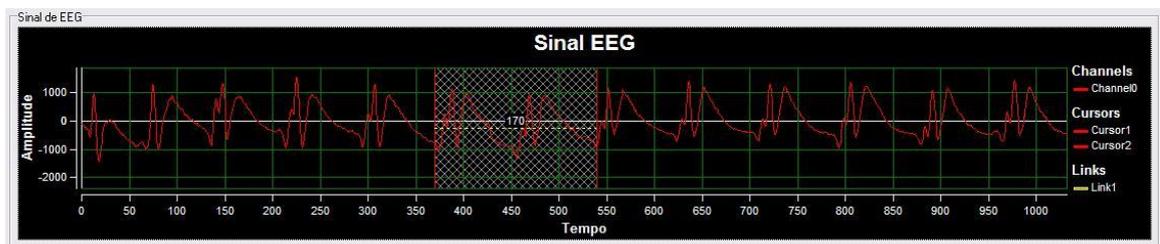


Figura 3: Separação de sinais com evidência de epilepsia.

A partir dos métodos descritos, realizamos a parte de construção e treinamento da rede, ao final do treinamento o erro total foi de 0.072321. Em seguida é realizada a sessão de teste, a rede foi carregada com o arquivo de conhecimento salvo para efetuar as classificações conforme padrões apresentados à sua entrada. O intervalo definido para os neurônios de saída permitem que trabalhem com um conceito aproximado de probabilidade.

Após a execução dos testes da rede, encontraram-se os seguintes resultados:

<b>REDE NEURAL MLP</b>	
<b>DESCRIÇÃO</b>	<b>VALORES</b>
ACERTOS	96%
ERROS	4%
VERDADEIROS POSITIVOS	96,8%
FALSOS NEGATIVOS	3,2%
VERDADEIROS NEGATIVOS	88%
FALSOS POSITIVOS	12%
ESPECIFICIDADE	88%
SENSIBILIDADE	96,8%
SELETIVIDADE POSITIVA	98,7%
SELETIVIDADE NEGATIVA	73,3%
PRECISÃO	92,4%

Tabela 2 – Resultados dos testes obtidos neste trabalho.

A base de dados utilizada no treinamento e nos testes deste projeto são oriundas do Departamento de Epileptologia da Universidade de Bonn – Alemanha (Universidade de Bonn) e também do grupo de estudos e desenvolvimento do software Wavelet Analysis of Epileptic Spikes (EEG SPIKES DETECTION). A Tabela 3 aborda sobre os dados utilizados:

<b>DESCRIÇÃO</b>	<b>VALOR</b>
NÚMERO DE SINAIS	550
TAXA DE AMOSTRAGEM	173,61 Hz e 240 Hz
LARGURA DE BANDA ESPECTRAL	0,5 Hz a 85 Hz
FILTRO PASSA-BAIXA	40 Hz
TAMANHO DA JANELA	4098 pontos

Tabela 3 – Informações dos arquivos utilizados para compor da base de dados

Observam-se dentro na literatura alguns estudos para chegarem ao mesmo propósito, como nos estudos de Zwe-Lee (2004) onde aplicou-se uma metodologia baseada na transformada de Wavelet discreta aliada com um classificador de redes neurais do modelo probabilístico, onde obteve em seus resultados um valor de 90% de acertos.

O pesquisador Pereira (2003) utilizou o modelo de redes neurais com entradas parametrizadas, onde desta forma obtêm-se os parâmetros observando diretamente o sinal, aplicando-se à rede neural quarenta e cinco parâmetros extraídos do sinal para o reconhecimento da rede que utiliza o modelo Perceptron Multicamadas treinada pelo algoritmo Backpropagation. Esta técnica esbarra em algumas dificuldades como por exemplo, uma profunda análise, capaz de eliminar equívocos relacionados à amplitude de eventos de epilepsia com ruídos e piscadas palpebrais. Contudo, esse método atingiu 85,2% de acertos, 89% de sensibilidade, 90% de especificidade e 89,5% de precisão.

No trabalho de Lui (2002) propôs-se a utilização de vários métodos para o processamento de sinais com diversos estágios como: filtros adaptativos integrados onde em primeira instância tinha o objetivo de separar a parte estacionária da parte não estacionária no sinal. Em seguida a transformada de wavelet para a extração de características; redes neurais artificiais para o reconhecimento de características onde utilizou a topologia de perceptrons com várias camadas; e por fim os sistemas especialistas para separar os artefatos que possam surgir no sinal utilizando regras heurísticas para decidir se o sinal possui eventos epileptogênicos ou não epileptogênicos. Com essas técnicas foi possível alcançar um total de acertos de 90,0%.

Em Argoud (et. al. 2004) foi proposto utilização da Transformada Wavelet para extrair características de eventos com a presença epileptogênica no sinal de exame de EEG, e um grupo de



redes neurais artificiais para classificação e distinção dos eventos epileptogênicos dos eventos normais de fundo. O sistema proposto atingiu um desempenho de 80% de acertos. No mesmo sentido, Patnaik (et al, 2008) utiliza wavelet para a extração de características e obtenção dos parâmetros estatísticos a partir dos coeficientes wavelet. Uma rede neural artificial com topologia feedforward aliada com o algoritmo de treinamento backpropagation para a classificação destas características. Utilizou-se também o algoritmo genético para a escolha do conjunto de treinamento e também para implementar um estágio de pós-classificação utilizando pesos harmônicos para aumentar a precisão. Enfim este trabalho alcançou uma especificidade média de 99,19%, sensibilidade de 91,29% e seletividade de 91,14%.

Observando algumas das principais literaturas sobre detecção e caracterização de eventos epileptiformes em sinais de EEG, estão em média divididas em duas partes: extração de características e classificação destas características. A extração de características o qual é um fator primordial para que haja uma ótima classificação, são utilizadas diversas técnicas, afim de que se encontre a que melhor consiga resolver a enorme dinâmica existente dentro de um sinal de EEG.

Para a classificação destas características a literatura é unânime em salientar a utilização das Redes Neurais Artificiais, devido a sua capacidade de classificar e aprender ao longo de sua aplicação.

Por se tratar de um desafio, tendo em vista que a proposta é resolver o grande problema da extração de características, sugere-se com esta pesquisa, um novo método capaz de dar resolutividade nesta área da ciência.

## **6. CONCLUSÕES**

A principal motivação para realizar este trabalho foi colaborar com o desenvolvimento das pesquisas na área da tecnologia, à medida que foi possível demonstrar uma nova técnica de identificação e caracterização de eventos epileptiformes em sinais de EEG, utilizando a dinâmica simbólica. Ou seja, através dos sinais dos exames de EEG observou-se toda a dinâmica existente entre os sinais, com presença de eventos de epilepsia, aplicando em seguida as redes neurais artificiais para auxiliar na classificação. Depois de observado a dinâmica destes sinais, aparecem dois resultados: sinal epiléptico ou sinal não epiléptico. Dentro das pesquisas realizadas, observou-se na literatura alguns resultados de trabalhos que utilizaram diversas técnicas para o mesmo propósito.

Observando algumas das principais literaturas sobre detecção e caracterização de eventos epileptiformes em sinais de EEG, estão em média divididas em duas partes: extração de características e classificação destas características. A extração de características o qual é um fator primordial para que haja uma ótima classificação, são utilizadas diversas técnicas, afim de que se encontre a que melhor consiga resolver a enorme dinâmica existente dentro de um sinal de EEG.

Para a classificação destas características a literatura é unânime em salientar a utilização das Redes Neurais Artificiais, devido a sua capacidade de classificar e aprender ao longo de sua aplicação.

Por se tratar de um desafio, tendo em vista que a proposta é resolver o grande problema da extração de características, sugere-se com esta pesquisa, um novo método capaz de dar resolutividade nesta área da ciência.

Neste trabalho foi aplicada a Dinâmica Simbólica para extração de características que evidenciem a presença de eventos epileptiformes, e de modo satisfatório foi possível alcançar os resultados com essa nova técnica. Para tanto, foi utilizada ferramentas de otimização como os Algoritmos Genéticos, a Dinâmica Simbólica para extração de características e as Redes Neurais Artificiais para a classificação destes eventos, chegando a resultados amplamente favoráveis, tendo em vista que alcançamos 96,8% de sensibilidade, 88% de especificidade, 92,4% de precisão e 96% de acertos. São números positivos para pesquisas que envolvem novos métodos de aplicação e que podem atingir parâmetros mais significativos, a partir da constatação deste trabalho.

## **AGRADECIMENTOS**

Ao meu orientador Tito Dias Junior por ser essa pessoa tão especial e ter proporcionado a infinita honra de estar integrando esse trabalho.



Aos professores e funcionários do Departamento de Engenharia Mecânica da Universidade de Brasília pelo apoio e formação acadêmica.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo auxílio financeiro ao longo deste trabalho.

Ao meu Saudoso e inesquecível pai, pela força, apoio e incentivo enquanto presente para a conclusão deste trabalho.

## REFERÊNCIAS

ARGOUD, F. I. M.; AZEVEDO, F. M.; NETO, J. M. Sistema de detecção automática de paroxismos epileptogênicos em sinais de eletroencefalograma. Revista Controle & Automação, Vol.15 no. 4, Outubro, Novembro e Dezembro 2004.

AZEVEDO, F. M. Algoritmos genéticos em redes neurais artificiais. V Escola de Redes Neurais. ITA. São José dos Campos – SP. 19 de Julho de 1999.

BARRETO, JORGE M. Introdução às Redes Neurais Artificiais. Anais V Escola Regional de Informática da SBC Regional Sul, 5 a 10 de maio de 1997. Páginas 41 - 71.

BORN, R. S. Filtros adaptativos aplicados a sinais biomédicos. Monografia, Universidade Federal de Pelotas, 2000.

EEG SPIKE DETECTION. Wavelet Analysis of Epileptic Spikes, banco de dados de EEG. Disponível em: <http://eegspike.republika.pl/signals.html>. Acessado em: 31 de janeiro de 2010.

GUYTON, A. C. Tratando de fisiologia médica. Rio de Janeiro: Guanabara Koogan, 1997.

HAYKIN, S. Neural Networks: a Comprehensive Foundation. 2nd edition. New Jersey, Prentice-Hall, 1999.

LUI, H. S.; ZHANG, T. & YANG, F. S. A Multistage, Multimethod Approach for Automatic Detection and Classification of Epileptiform EEG. IEEE-Transactons on biomedical engineering, vol. 49, no. 12, December 2002.

PARREIRA, F. J. Detecção de crises epilépticas a partir de sinais eletroencefalográficos. Tese de doutorado, Universidade Federal de Uberlândia, Uberlândia, 2006.

PEREIRA, M. C. V. Tratamento de Sinais Bioelétricos para Processamento por Redes Neurais Artificiais. Tese de Doutorado, Instituto de Engenharia Biomédica, Universidade Federal de Santa Catarina, Florianópolis, 2003.

UNIVERSIDADE DE BONN. Departamento de Epileptologia da Universidade de Bonn – Alemanha, banco de dados de EEG. Disponível em: [http://epileptologie-bonn.de/cms/front\\_content.php?idcat=193&lang=3&changelang=3](http://epileptologie-bonn.de/cms/front_content.php?idcat=193&lang=3&changelang=3). Acessado em: 02 de fevereiro 2010.

ZWE-LEE, G. Wavelet-based neural network for power disturbance recognition and classification. Power delivery, vol. 19, no. 4, Oct. 2004.