Journal of Medical Systems, Vol. 26, No. 4, August 2002 (© 2002)

Categorization and Analysis of Pain and Activity in Patients With Low Back Pain Using a Neural Network Technique

John J. Liszka-Hackzell[1,3] and David P. Martin[2]

Low back pain represents a significant medical problem, both in its prevalence and its cost to society. Most episodes of acute low back pain resolve without significant long-term functional impact. However, a minority of patients experience extended chronic pain and disability. In this paper, we have explored new techniques of patient assessment that may prospectively identify this minority of patients at risk of developing poor outcomes. We studied 15 patients with acute low back pain and 25 patients with chronic low back pain over 4 month's time. Patients monitored their pain and activity levels continuously over the first 3 weeks. Pain and functional status were assessed at baseline and at 3 weeks following enrollment. Follow-up assessment of functional status and progress were performed at 2 and 4 months. The pain and activity levels were categorized using a self-organizing-map neural network. A back-propagation neural network was trained with the categorization and outcome data. There was a good correlation between the true and predicted values for general health ($r = 0.96$, $p < 0.01$) and mental health ($r = 0.80$, $p < 0.01$). No significant correlation was found if activity and pain data were not entered into the analysis. Our results show that neural network techniques can be applied effectively to categorizing patients with acute and chronic low back pain. It is our hope that future research will allow these categorizations to be tied to prognostic and therapeutic decisions in patients who present with episodes of back pain.

KEY WORDS: low back pain; pain diary; artificial intelligence; pain measurement; neural networks.

INTRODUCTION

Low back pain represents a significant medical problem, both in its prevalence and its cost to society. Most episodes of back pain respond adequately to conservative therapy, but some patients will continue to suffer chronic back pain for long periods, incurring significant medical expenses and absence from work.[(1–3)] Since

[1]Department of Anesthesiology, Orebro Medical Center Hospital, S-701 85 Orebro, Sweden.
[2]Department of Anesthesiology, Mayo Clinic and Mayo Foundation, Rochester, Minnesota 55905.
[3]To whom correspondence should be addressed.

0148-5598/02/0800-0337/0 © 2002 Plenum Publishing Corporation

most patients will respond favorably to conservative measures, aggressive steps are not indicated early in an episode of back pain. However, it is speculated that early aggressive therapy may be beneficial for the minority of patients who do not improve with conservative therapy. This presents a dilemma because it is difficult to predict which patients are destined not to improve with conservative measures. In this paper, we have explored new techniques of patient assessment that may have prognostic value in this group of patients.

Pain is inherently a subjective experience and therefore cannot be objectively measured and analyzed. Nevertheless, we may be able to determine more information from patients by using a new technology in assessing symptoms continuously over time, and by using new mathematical techniques to analyze the data. In this paper, we have used an electronic symptom diary and activity meter to assess patients continuously. Neural network mathematical techniques have been applied to the data to categorize the history of the pain and activity levels.

Very little has been done on categorization of pain and activity levels in patients with back pain. Categorization of these variables may provide different pain and activity patterns that may contain information related to prognosis or outcome for the individual patient. Pain intensity level data are difficult to process and analyze since the measurements are highly subjective. It may therefore be more reasonable to study patterns rather than absolute levels. Also, by using artificial intelligence (AI) techniques, the chances of extracting important information may be facilitated.

Neural networks(4,5) and especially self-organizing map neural networks(6) are powerful tools when working with categorization problems. By using neural networks, a statistical model may be easily developed from experimental data. The knowledge in the model is a product of the learning process that the network undergoes while training data is presented to it and the network properties are adjusted.

In this study, we have used the wavelet transform to process the activity level data prior to subjecting it to the neural network categorization process. This is necessary since the characteristics of the two variables are very different. The pain levels are sampled on average every 90 min, and have predominantly low frequency content. The activity levels are sampled as counts (pulses) every 1 min. By using a wavelet filtering technique, the activity levels are processed to allow accurate analysis of the two variables together.

In addition to the pain and activity level data, we also collected pain status data using the McGill Pain Questionnaire (MPQ),(7) functional status from the SF-36 questionnaire(8) and a progress score at various times during the study period. The pain and activity level category data were then used together with the MPQ data, the SF-36 data and progress scores to train a back-propagation neural network. This back-propagation neural network was then used to make predictions about patient's progress at 2 months and 4 months as well as their functional status at 4 months.

METHODS

The protocol was reviewed and approved by the Institutional Review Board. We studied 15 patients with acute low back pain (<6 months) from both the Urgent Care Center and Emergency Department. Another 25 patients with chronic low back

Table I. Inclusion Criteria

Chronic low back pain	Acute low back pain
Low back pain >6 months Age 18–75 English speaking Able to consent Able to use pain diary	Acute onset low back pain No history of spine surgery No previous episode within the last year Age 18–75 English speaking Able to consent Able to use pain diary

pain (>6 months) were studied from the Spine Center. The patients were between the ages of 18 and 75. The inclusion criteria are listed in Table I.

All patients were interviewed by a nurse coordinator to confirm eligibility. During the initial meeting the data collection procedure was explained. An initial functional status was obtained using the SF-36 questionnaire (SF-36-1) as well as a pain status using the McGill Pain Questionnaire (MPQ1). Each patient was asked to monitor his or her pain and activity levels continuously during a period of 3 weeks. Pain levels were recorded every 90 min between 8 A.M. and 10 P.M. (0800–2200) using an electronic diary.

The Electronic Symptom Diary is a pager-sized device developed at Mayo Clinic for collecting symptoms from patients over time (Fig. 1). Patients were instructed to

Fig. 1. The Mayo Clinic Electronic Symptom Diary.

Table II. Description of the Progress Score (PS) Used in This Study[a]

Progress score	Description
3	Complete improvement
2	Significant improvement
1	Some improvement
0	No change
−1	Some deterioration
−2	Significant deterioration
−3	Complete deterioration

[a] Patients were asked to quantify how their back pain was compared to when they initially enrolled in the study.

quantify their pain using a numerical rating scale, and enter the value by pressing one of the buttons labeled from 0 to 10, where 0 is no pain and 10 is the worst possible pain. Their entry was stored in the device along with the corresponding date and time. Patients were prompted by the diary to enter the scheduled measurements, but could also enter measurements at any time. They were encouraged to enter additional measurements as often as they wanted, emphasizing the importance of recording symptoms at their peak and nadir as well as intermediate values. At the end of the session, the data were downloaded for analysis.

Activity levels were monitored using an AW-64 Actiwatch (Mini-Mitter Inc.). The Actiwatch passively measures activity,[9,10] and patients were asked to wear it continuously on their nondominant arm for the duration of the study. The activity was sampled as accumulated counts every minute. The patients were contacted by telephone on the second and tenth day of the study to provide an opportunity for further patient information. At the end of the 3-week pain and activity level data collection, the patients were again called over the telephone and asked to return their equipment. At that time, they were also asked to provide a progress score (PS1) as well as a second McGill Pain Questionnaire (MPQ2). The progress score was defined according to Table II.

Patients were again contacted at 2 months time when they were asked for a second progress score (PS2). Finally, patients were asked to complete a second functional status using the SF-36 (SF-36-2) at 4 months time. At this time the study was completed. Figure 2 shows the data collection sequence.

The time series were compared for each 0800–2200 period to assure that both time series were complete. Only the data collected between 0800 and 2200 each day were used for further analysis. Any incomplete time series were discarded. If the activity level time series for a certain day was incomplete, the corresponding pain level time series was also discarded and vice versa. Any patient who had not collected at least 14 complete time series (14 days) was dropped from the study. Also, the MPQ[7] and SF-36[8] questionnaires were processed according to their accompanying guidelines. The progress scores were processed according to Table II.

The activity level time series were sampled every 1 min. The pain level time series were sampled on average every 90 min. In order to further analyze the two time series, they had to be converted so that they would appear as if they had both been sampled at the same even intervals. The analysis was performed as outlined

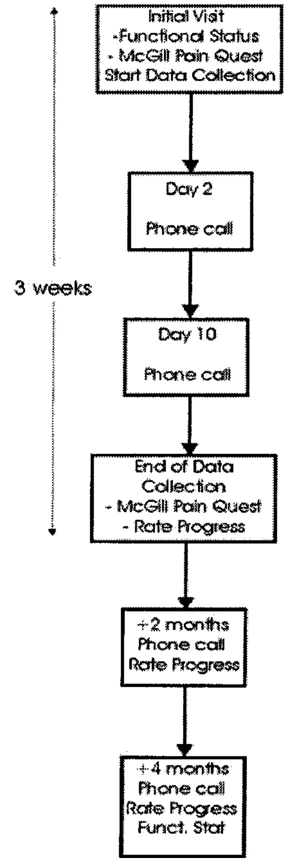

Fig. 2. The sequence of data collection during the study.

in Fig. 3. The two time series show very different frequency content, not only due to different sampling rates, but also due to the different processes producing each time series. Further analysis required the activity level time series to be threshold-filtered using the wavelet transform.[11,12] The activity level readings consist of two components: a stochastic component reflecting random activities in each patient, and a deterministic component corresponding to the general trend of activity in each

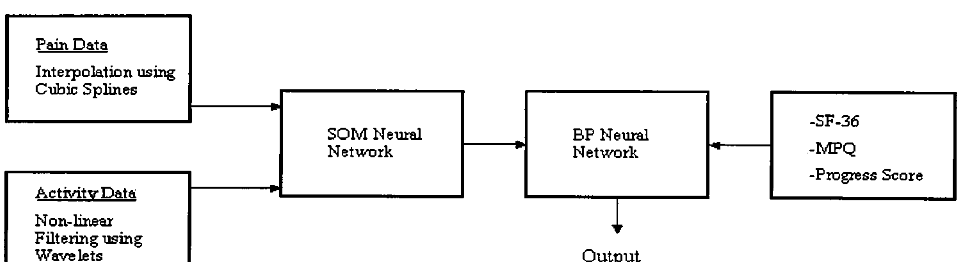

Fig. 3. Flowchart of the data analysis sequence.

patient. The separation of these two components may be done using threshold filtering. A percentage of the maximum wavelet coefficient magnitude was selected in a way such that the maximum kurtosis of the deterministic component was obtained. This way, the maximum deviation from a gaussian amplitude distribution was determined, which is the characteristic property of the deterministic component. It was found that a 7% threshold gave the best separation for these data. Simultaneously with the threshold filtering, the signal was also low-pass filtered in the frequency domain. This procedure is probably unique in the way that it reproduces the general trend of the activity data without influencing its level.

In the filtered activity level time series, every 10th measurement was kept and a new activity level time series was created. This new activity level time series appeared as if it was sampled every 10 min. The original pain level time series was interpolated using cubic splines,[13] to appear as if it would have been sampled every 10 min. The new time series now had equal length and sampling interval. Both time series now consisted of 84 measurements each corresponding to each 0800–2200 interval. We had to make the assumption that an interpolation of the pain level time series would reflect the change in pain levels outside the collected measurements. Collection of pain levels requires the patient to manually enter a measurement into the electronic diary when prompted. To prompt the patients more often than every 90 min over 3 weeks would likely have negatively affected compliance with the study. From our own clinical knowledge, we feel that an interpolation with cubic splines most accurately reflects this process.

Data from 20 randomly selected patients was used to train a self-organizing map neural network. The self-organizing map consisted of 168 processing elements (PE) in the input layer, a 14×14 PE Kohonen layer and a 2 PE output layer. The neural network software that was used in this study was NWORKS Professional 2 (NeuralWare Inc.). Recall was then done with the whole group of 40 patients. The output from the self-organizing map showed separation of patients into different patterns (Fig. 4). The self-organizing map output from 30 randomly selected patients together with their corresponding SF-36-1, MPQ1 and MPQ2 as well as PS1 was then used as input values for training a back-propagation neural network. The desired output values during the training sequence were PS2 and PS3 as well as the SF-36-2. Finally, recall was done with the trained back-propagation neural network using the remaining 10 patients. The back-propagation neural network consisted of 17 PE in the input layer, 27 PE in the hidden layer and 10 PE in the output layer. Correlations between the true and predicted outputs were calculated. Another back propagation neural network was trained using the above data, but without the input from the self-organizing map. This was done to investigate the predictive properties of the activity and pain information analyzed in the self-organizing map neural network. This information was used to compare the complete back-propagation neural network model with a back-propagation neural network model without the self-organizing map output.

To ensure that the group of patients selected for training and the group of patients selected for recall would accurately reflect the statistical properties of the entire group, a test was performed by comparing 10 different randomly selected groups. These 10 groups showed similar statistical properties and distributions and

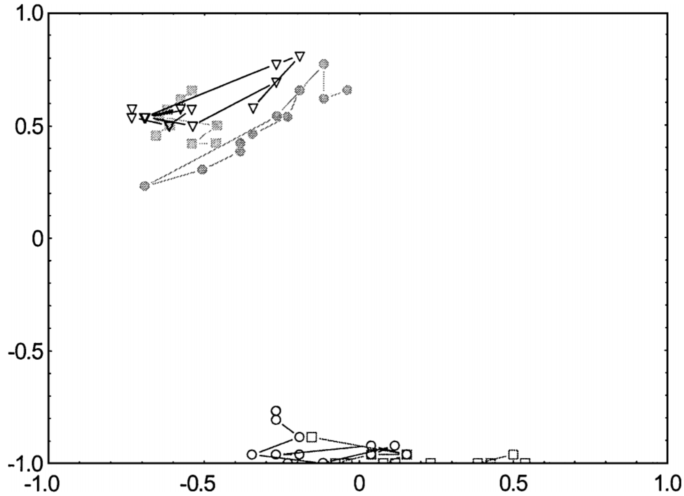

Fig. 4. Output from the self-organizing map Kohonen layer. The *X* and *Y* axes show the Kohonen coordinates as presented by the output layer. This example shows measurements of five different patients that map into two distinct parts of the Kohonen layer.

it was concluded that using one of these groups for recall and the remaining patients for training would be appropriate.

RESULTS

The average age of the patients that completed the study was 48.4 years. There were 24 males and 16 females. Of the 62 patients that were initially enrolled, 40 patients completed the study (15 patients with acute back pain and 25 patients with chronic back pain). The number of complete time series (activity and pain) of those who completed the study was on average 17 of 21. The main reason why patients did not complete the study was that it caused inconvenience by requiring too much time and patient effort. The output from the self-organizing map neural network showed separation of different patients into different areas of the Kohonen layer, with some patients showing a larger change during the 3-week activity and pain level measurements (Fig. 4).

The output from the back-propagation neural network in the 10-patient test group during recall showed significant correlations between the true and predicted values of the SF-36 scores describing general health and mental health. None of the other SF-36 or PS scores showed significant correlations. The correlation coefficient between the true and the predicted SF-36 score for mental health was 0.80 ($p < 0.01$) and between the true and the predicted SF-36 score for general health was 0.96 ($p < 0.01$). The scatter-plots are shown in Fig. 5.

In an attempt to make the same predictions from another back-propagation neural network without the activity and pain information from the self-organizing map, we found that correlations became nonsignificant (mental health $r = 0.33$ and general health $r = 0.40$).

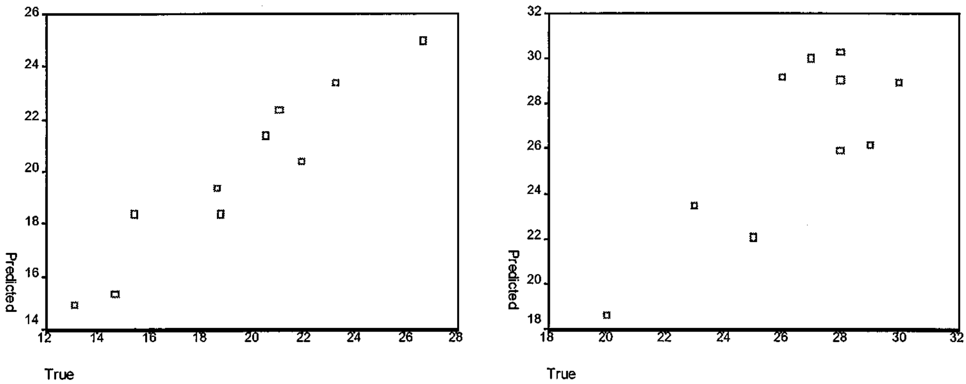

Fig. 5. Scatter-plots between the true and predicted values of the SF-36 scores describing mental health (left) and general health (right).

DISCUSSION

Neural networks are in wide use in many different areas of science and medicine. Within medicine, neural networks have been applied to analysis of medical images, medical signal processing, and clinical and laboratory data.[14] Especially problems concerning pattern recognition and classification have been successfully analyzed using these techniques. Some examples are character recognition, etiology of low back disorders, and interpretation of sonar signals from underwater objects.[15–17]

Neural networks have been shown to be superior to both conventional statistical methods and manual/specialist-based analysis in many studies. There are many benefits in using neural networks compared to conventional statistical methods.[14]

By using neural networks, more complex tasks can be learned from examples than by using conventional statistical techniques. Another benefit is that both qualitative and quantitative data can easily be included in the same model. Neural networks perform well in analysis of nonlinear multivariate data. Also, a fully trained neural network can be used for further analysis of new data in or close to real-time.

The disadvantages related to the use of neural networks includes difficulties of expressing their function/knowledge in a simple way. Validating the network results may be more difficult than with conventional statistics.

Hedén studied a neural network's capability to identify acute myocardial infarction in the 12-lead ECG, and compared it to conventional rule based analysis and analysis by an experienced cardiologist.[18] The neural network showed higher sensitivity and discriminant power than both the rule based interpretation program and the cardiologist.

Another study by Veselis *et al.* used neural networks to differentiate similar EEG spectra, and compared it to discriminant analysis, visual recognition by the experimenters, and analysis of variance (ANOVA).[19] Discriminant analysis performed as well as neural networks. Both these methods gave significantly better results than visual recognition by the examiners and ANOVA. Also, in this study, the authors showed that it was not possible to identify some patterns with conventional methods, but that it was possible to do so using neural networks or discriminant analysis.

Bounds used neural networks in the diagnosis of low back disorders.(17) The study of Bounds was based on a list of symptoms and physical findings obtained from each patient in the study at one point in time. The data were then classified as one of four defined diagnoses. Bounds also compared the neural network with other pattern recognition methods such as diagnosis by a group of experts, K-Nearest Neighbor and Closest Class Mean. The neural network approach outperformed the clinicians, and performed as well as or better than the other pattern recognition techniques.

In Fast-Fourier Transform (FFT), the analyzed signal is folded using a complex sine wave of constant amplitude and varying frequency. That way, the signal is decomposed into a frequency spectrum. With wavelets,(11) the analyzed signal is folded with a "mother wavelet," which may have different forms, but is limited in the time domain. A special form of wavelets is the Morlet wavelet, which is a locally periodic wave-train that is related to windowed Fourier analysis. It is obtained by taking a complex sine wave and localizing it with a Gaussian (bell-shaped) envelope. An advantage of wavelet analysis is that its resolution varies with the frequency of the signal. The largest difference between the Fourier transform and the Morlet wavelet is that there is a constant width window used in the Fourier transform, while the window width decreases with frequency in the Morlet wavelet.

In wavelet analysis as compared to Fourier analysis, the frequency in a dynamic frequency spectrum is referred to as the time scale. Rather than a frequency spectrum, a scalogram with a reversed Y-axis (due to the use of a time scale instead of a frequency scale) is produced by wavelet analysis. The spectral density in the Fourier spectrum is comparable to the wavelet coefficient magnitude, the frequency step is referred to as dilation and the time step with which the window is moving becomes the translation.

A valuable property of wavelet analysis is that in addition to conventional filtering in the frequency domain, it is possible to perform filtering in the domain of the wavelet coefficient magnitude.(12) The simplest form of this filtering technique is threshold filtering, where the signal may be decomposed into two parts. One part corresponds to wavelet coefficients above a certain magnitude (threshold) while the other part corresponds to wavelet coefficients below the threshold (strong and weak component). This procedure is important when working with signals of different character, for example, a signal with a stochastic and deterministic component.

In this study, the output from the self-organizing map indicated that the pain and activity patterns in these patients have different properties. Some patients showed a more stationary appearance while others showed a pattern that covered a larger area of the Kohonen layer. It appears that patients who experienced a small change in either their pain or activity levels were more stationary as indicated by the self-organizing map output. Different patients also mapped to different geographic areas of the Kohonen layer. It would be reasonable to assume that different parts of the Kohonen layer reflect different properties of the corresponding pain and activity levels. The output from the Kohonen layer was, together with the initial MPQ, SF-36, and progress score, used to make predictions about the patient's 2- and 4 months status.

We were able to accurately predict the 4-month SF-36 scores describing general and mental health using a neural network model trained with activity and pain level data collected during an initial 3-weeks period in combination with initial MPQ, SF-36, and Progress scores. Prediction of the general health score ($r = 0.96$) showed a higher correlation coefficient than the mental health score ($r = 0.80$). The general health score reflects the patient's perception of their overall health, while the mental health score describes the patient's mood. It is not entirely clear why the general health score was easier to predict than the mental health score. We were unable to make predictions of the remaining 4-months SF-36 scores or the progress scores at 2 and 4 months. It appears that the activity and pain information processed in the self-organizing map neural network is vital to the predictive performance of the back-propagation neural network, since a back-propagation neural network without input from the self-organizing map is unable to provide acceptable predictions.

One must consider that we have only investigated a small group of patients and that this should be considered a pilot study. At the same time, the results may indicate that there is prognostic information in the input data. It may indirectly suggest that back pain affects a patient's perception of their general or mental health. If one would expect a low predicted general- or mental health estimation in a patient, that patient may benefit from a more aggressive management.

The difficulty to predict the 2- and 4 months progress scores may be explained by the fact that the progress score was designed to reflect change. However, one can question how well a patient can quantify a change over a longer period of time due to recall bias. We must conclude that our inability to accurately predict the progress score should have been anticipated and may be impossible due to the characteristics of this variable.

In conclusion, our results show that neural network techniques can be applied effectively to categorizing patients with acute and chronic low back pain. It is our hope that future research will allow these categorizations to be tied to prognostic and therapeutic decisions in patients who present with episodes of back pain.

ACKNOWLEDGMENTS

This study was funded by a grant from Mayo Foundation. The authors thank Diane Maxson, RN, Greg Wilson, and Laurie Meade, RN, for their help in recruiting patients and collecting data.

REFERENCES

1. Chapman, C. R., and Syrjala, K. L., *Bonica's Management of Pain*, 3rd edn., Lippincott Williams and Wilkins, Baltimore, MD, 2001.
2. Tollison, C. D., *Handbook of Chronic Pain Management*, Williams and Wilkins, Baltimore, MD, 1989.
3. Deyo, R. A., and Weinstein, J. N., Low back pain. *N. Engl. J. Med.* 344:363–370, 2001.
4. Klimasauskas, K., *Neural Computing: A Technology Handbook*, NeuralWare Inc., 1993.
5. Pao, Y. H., *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley, Reading, MA, 1989.
6. Kohonen, T., *Self-Organization and Associative Memory*, Springer, New York, 1989.

7. Melzack, R., The McGill Pain Questionnaire: Major properties and scoring methods. *Pain* 1:277–299, 1975.
8. Ware, J. E., *SF-36 Physical and Mental Health Summary Scales: A User's Manual*, Quality Metric Inc., 1997.
9. Cole, R. J., Automatic sleep/wake identification from wrist activity. *Sleep* 15(5):461–469, 1992.
10. Patterson, S. M., Automated physical activity monitoring: Validation and comparison with physiological and self-report measures. *Psychophysiology* 30:296–305, 1993.
11. Chui, C. K., *An Introduction to Wavelets*, Academic Press, New York, 1992.
12. Liszka, L., Decomposition of time series. *Invited Lecture at the 2nd International Workshop on Artificial Intelligence Applications in Solar-Terrestial Physics*, Lund, Sweden, 1997.
13. Weisstein, E. W., *Concise Encyclopedia of Mathematics*, CD-ROM, Chapman and Hall/CRCnet Base, 1999.
14. Forsstrom, J. J., Artificial neural networks for decision support in clinical medicine. *Ann. Med.* 27:509–517, 1995.
15. Gorma, R. P., Analysis of hidden units in a layered network trained to classify sonar targets. *Neural Netw.* 1:75–79, 1988.
16. Rajavelu, A., A neural network approach to character recognition. *Neural Netw.* 2:387–393, 1989.
17. Bounds, D. G., A comparison of neural networks and other pattern recognition approaches to the diagnosis of low back disorders. *Neural Netw.* 3:583–591, 1990.
18. Hedén, B., Acute myocardial infarction detected in the 12-lead ECG by neural networks. *Circulation* 96:1798–1802, 1997.
19. Veselis, R. A., Veselis, R. A., Reinsel, R., and Wronski, M., Analytical methods to differentiate similar EEG spectra: Neural networks and discriminant analysis. *J. Clin. Monit.* 9:257–267, 1993.