

G. Winterer^a
M. Ziller^a
B. Klöppel^b
A. Heinz^a
L.G. Schmidt^a
W.M. Herrmann^a

Analysis of Quantitative EEG with Artificial Neural Networks and Discriminant Analysis – A Methodological Comparison

^a Department of Psychiatry of the Free University of Berlin,

^b Institute for Computer Sciences and Mathematics, University of Kassel, Germany

.....
Abstract

Artificial neural networks (ANN) are widely used to solve problems of differentiating between groups. However, serious comparisons of this method with the traditional procedure for such tasks (discriminant analysis) are rare. Discussing the results of both methods with the example of highly topical data, we try to demonstrate advantages and drawbacks of both methods. For this purpose, quantitative EEGs of 78 alcoholics were investigated in order to determine whether it is possible to predict relapse of these patients at the beginning of treatment. ANN software is available in Kassel (Institute for Computer Sciences and Mathematics).
.....

.....
Key Words

EEG
Artificial neural networks
Discriminant analysis

Introduction

This article intends to provide a short, basic introduction to artificial neural networks (ANN) and a comparison with conventional multivariate discriminant analysis to scientists who are engaged in neurophysiological research. This is done by giving a brief explanation of the most popular neural network tool at present, a feedforward multilayered network trained using backpropagation, and by contrasting discriminant analysis with ANN on a practical example. Thus, we investigated quantitative EEGs (QEEG) of 78 chronic alcoholics in order to find out whether relapse prediction is possible at the beginning of treatment. Major benefits and drawbacks of neural networks compared to conventional statistical methods with reference to neurophysiological research will be discussed.

ANN have received a great deal of attention during the last few years. They have been applied to a wide range of classification problems where statistical methods are tra-

ditionally employed, including speech recognition [1], prediction of stock market performance [2], or identification of underwater sonar contacts [3]. In addition, neural networks have been used successfully in biomedical research, such as predicting heart problems in patients [4, 5], predicting the mechanisms of action of cancer drugs [6], detecting errors in anesthesia breathing circuit [7], medical signal processing [8], medical imaging [9, 10] and analysis of laboratory data [11–13]. There are also a number of good introductory articles on neural networks. Warner and Misra [14] provide an excellent understanding of neural networks as a statistical tool for the reader with a mathematical or statistical background, Forström and Dalton [15] give a comprehensible overview of ANN for decision support in clinical medicine, and Selker et al. [16] present an impressive comparison of ANN with logistic regression, to mention just a few articles.

Neural networks have their origin in using computers to simulate the function of the brain. Theories on the physiology of the brain have set the scene for the develop-

Fig. 1. This figure shows the principal constituents of a neuron within an artificial neural network. It receives input from other neurons (x_1 – x_3) firing at different rates (w_1 – w_3). Different firing rates are weighted and summed up $\sum w_i$ which results in a function y that expresses the firing rate of the neuron.

Fig. 2. A typical multilayer perceptron architecture.

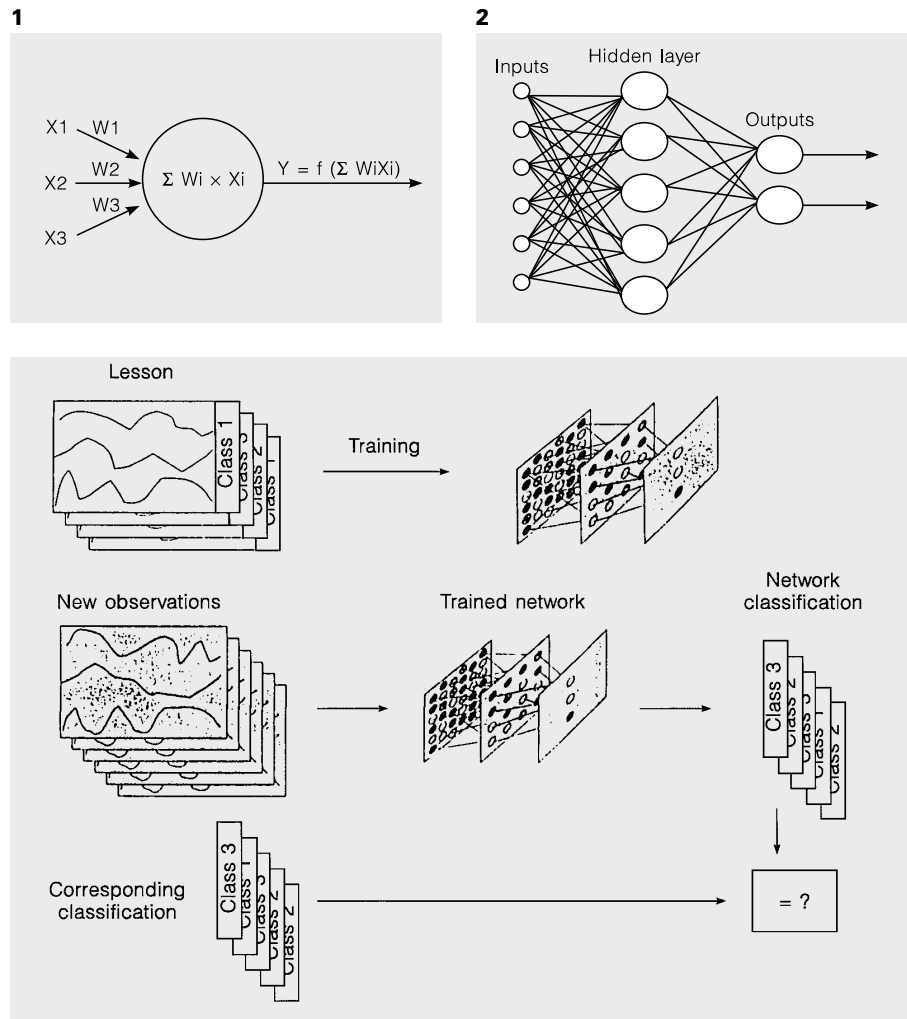


Fig. 3. Way of working of ANN.

ment of neural network algorithms. Thus, during the last decade a new branch of research developed, now known as *computational neuroscience* [17–19]. While research in this area essentially strove for modelling basic neuronal processes (e.g. Hebb’s learning rule), a byproduct of this research has been the application of ANN as a statistical tool to feature detection in complex data sets. This development was furthered by the argument that conventional statistical programs running on currently available computers are very fast at numerical computations, far exceeding human capabilities, however, lack abilities of the human brain such as quickly identifying features, even in the presence of noise; to understand, interpret, and act on probabilistic or fuzzy notions, and to make inferences and judgements based on past experience and relate them to situations that have never been accounted before. Therefore, it seemed attractive to include properties of the human brain in computer programs, when prob-

lems of feature detection (e.g. classification tasks) are addressed.

Simulating basic brain architecture, ANN consist of neurones as the principal computational units (fig. 1). The general structure of a feedforward multilayer neural network (*multilayer perceptron*) is made up of an input layer, one or several hidden layers and an output layer; each layer consisting of several neurones, which are interconnected as depicted in figure 2. Usually, the number of neurones in the input layer is determined by the number of attributes in the data (e.g. different laboratory tests). The number of output neurones is equal to the number of predefined classes (e.g. number of treatment effects to be extracted from a given data input). The number of neurones of the hidden layer is essentially dependent on the complexity of the relationships between the input variables and has to be established experimentally, as it is frequently not known in advance. In regard to the function

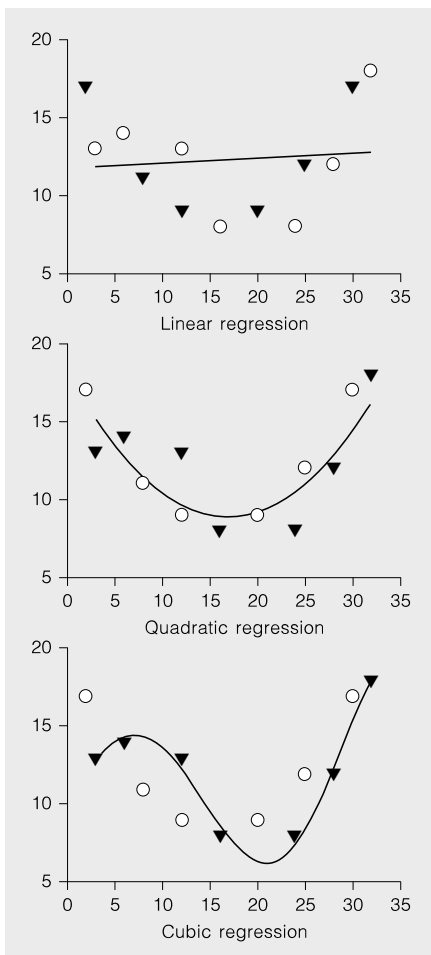


Fig. 4. Elasticity-plasticity dilemma in finding the optimal model, which both describes data and allows generalization. The training set data (filled triangles) were fit by either a linear, quadratic or cubic regression. A linear approximation is poor to both sets of data. The quadratic function is best fit in this case, while a cubic one is too specific.

of a net, one has to distinguish between the learning and test phase (fig. 3). In its learning phase, the net is given predefined input-output example pairs (training set), and it is taught to calculate the desired output from the input data, which is called *supervised learning method*. This is similar to the regression problem where we have a set of independent variables (inputs) and dependent variables (output), and we want to find the relationship between the two. After an input pattern has been applied as a stimulus to the input-layer, it is then propagated through each upper layer until an output is generated (*feedforward net*). This output pattern is then compared to the desired output, and an error signal is calculated for each output unit. This error signal is then transmitted backwards across the

net (*backpropagation algorithm*), which results in an adjustment of the *connection weights* between the neurones. By this way, the patterns which the network learns are coded into the distributed connection weights between neurones as internal representations. Typically, example cases must be presented to the net many thousands of times, until the knowledge inherent in the input data set is learned adequately. Once a network has learned its knowledge, it is able to classify new cases very quickly, that is, to generalize its knowledge to independent test data (test set).

Although a detailed discussion of technical aspects (benefits and drawbacks) inherent to the application of neural networks is beyond the scope of this article, several general remarks have to be made. First, as knowledge is distributed across the connections, it is impossible to interpret knowledge, e.g. to characterize the functional relationship between the input variables in relation to a given output. In other words, knowledge cannot be expressed in rules. Therefore, neural networks are often called 'black boxes' [15]. Second, the topographical structure of a net has a considerable influence on the learning of the net. Actually, the capacity of a net, that is, the number of connections, should be kept as small as possible. If the capacity is too large, this will result in a bad generalization performance (fig. 4), because the net remembers every single example presented during the learning phase and only recognizes exactly identical examples during the test phase [20, 21]. Other related difficulties with neural networks involve e.g. choosing the initial starting weights of the connections and the decision when to stop training [14]; the latter being directly related to the necessity for large data sets [22] which is a frequent problem in medical research and the problem of so-called overlearning after too many iterations. Here, overlearning means that the network has adapted its weights too closely to the set of learning cases, which again results in a bad generalization performance. Finally, it should be mentioned here, that feedforward nets and backpropagation algorithm – although most frequently applied – represent only one of numerous possible network approaches and it is currently far from being clear which network type optimally suits a given classification task.

Despite all these yet unsolved problems, there is now general agreement that ANN offer some advantages when compared to conventional statistics. Thus, neural networks allow much more complex tasks to be learned from examples and are particularly suited to the analysis of nonlinear, multivariate data [15], although, in practice, this advantage again presupposes large data sets for net-

work training. Additionally, no a priori assumptions about the character of a functional relationship between input and output variables (e.g. type of nonlinear function) are necessary [14]. Finally, it is noteworthy that variables with different scalars can be fed simultaneously to the net, which for instance, allows the admixture of clinical data and data from laboratory investigations for classification purposes [15].

Patients and Methods

Patients. 78 alcohol-dependent patients who were treated for detoxification in our hospital were diagnosed according to the criteria of the International Classification of Diseases (ICD 10). Patients had been included in the study only when detoxicated, unmedicated for at least 3 days and if they no longer showed any clinical signs of withdrawal. After release from the hospital, patients were followed up in our outpatient unit for 3 months. During this time, patients were seen on a weekly basis within the first month and biweekly until the end of observation to determine whether they abstained or relapsed. Relapse was defined as minimum consumption of 50 g ethanol. To be able to compare distinguishable groups, we assessed patients for clinical outcome and included: (1) 49 patients who relapsed within 3 months after detoxification and (2) 29 patients who remained abstinent for at least 3 months.

EEG Recording. All EEG recordings were performed between 9 and 11 a.m. 7 days after the beginning of detoxification. Each EEG lasted 10 min and was recorded with Ag/AgCl electrodes under resting condition (semisupine position, eyes closed and instruction to relax but not to fall asleep). Electrode positions were defined according to the international 10/20 system, impedance was kept below 5 k Ω . The recordings were carried out with a conventional 32-channel electroencephalograph (Walter Graphtek). Amplifiers were calibrated using a 0.5-Hz rectangle generator, notch filter: 50 Hz., time constant: 0.3 s. The signals were digitized with a sampling frequency of 166.6 Hz. Antialias was performed in two steps. Analogue antialias at 1,000-Hz sampling rate: Butterworth 4th order, 140 Hz and digital antialias: FIR 64 taps, 75 Hz. Data were stored on optical disks. Of each patient one uninterrupted, 'representative' artefact-free EEG segment of 30 s was further analyzed. A segment was considered 'representative', when it contained exclusively basic EEG activity. Additionally, EEG activity 10 s after the intermittent instruction to open and close the eyes for 2 s during the running recording were excluded.

Parameters. For quantitative analysis the recorded signals (30-second segments) were split in 9 consecutive epochs of 3.072 s duration and submitted to spectral analysis using the fast Fourier transform. The square root of absolute power (magnitude), expressed in μ V as well as the relative power was computed for the following frequency bands according to the recommendations of Kubicki et al. [23] which were based on factor analysis: δ F = 0.5–5.5 Hz, θ F = 6.0–8.0 Hz, α F = 8.5–12 Hz, β 1F = 12.5–18 Hz, β 3F = 21.0–30.0 Hz. Power values were computed for all electrode positions with ears as reference (A1 + A2)/2. Additionally, we considered the correlation dimension of Grassberger-Procaccia [24–26], which is an upper bound of the fractal dimension of the signal. The correlation dimension was calculated for a 27-second EEG segment for the following

pairs of electrodes: F3–F7, F4–F8, P3–O1, P4–O2, T3–P3, T4–P4. The latter parameter was applied to our data although we are aware of their high sensitivity to nonstationarity of the signal. However, from visual inspection we knew that EEGs of alcoholics are generally highly stationary.

Discriminant Analysis. Further analyzing our data, we computed a discriminant function analysis with stepwise introduction of spectral variables (approximated F test for Wilk's lambda; weighted cases) [27]. For this purpose, we first compared randomly chosen 14 abstainers and 24 relapsers (training group), with the aim to achieve proper results with as few variables as possible. Then we applied the resulting discriminant function to the remaining 15 abstainers and 25 relapsers (test group), thus testing the resulting discriminant function, independently. For better comparison with nonlinear network tools, we extended the linear model by interaction terms (polynomial model of second order). Although there are not enough cases to estimate much more independent model parameters, tendencies can be described.

Artificial Neural Network. A detailed technical description of our network approach is submitted elsewhere [28]. For our investigation, improved, multilayer perceptron (feed-forward) networks (transfer function $1/(1 + e^{-x})$) [21, 29, 30] had been used with one hidden layer (2–20 neurons). Improvement was carried out by enhancing the learning speed: stepsize control and a layer-specific overrelaxation factor [31–33]. Additionally, for a formal improvement of generalization, a quantitative reduction of synaptic weights (weight decay) was carried out [34–36]. However, as the large number of input neurons (due to the large number of variables of the original data set) enforces the large amount of synaptic weights (degrees of freedom), only minor improvements were achieved. Consequently, two input selection methods had been applied: a. *knowledge-based input sets* referring to the results of the discriminant analysis and b. *genetic algorithms* [37–39] with stepwise input pruning, depending on nonlinear sensitivities.

The network input had been selected from the major domains relative spectral power and absolute spectral power. Spectral values had been applied either directly to the ANN or after a linear z-transformation. Comparable to the discriminant analysis, training of the net was performed with the data of the training group, independent testing was done with the data of the test group.

In a first knowledge-based approach, we reduced the number of input neurons in the subsequent way: First, we built a training set (EEG data of the training group) with subsets of possible network inputs (unscaled values) based on our knowledge from the discriminant analysis. Second, a reduction using singular value decomposition was applied [34]. So, the original lesson was substituted by a lesson with fewer inputs and maximum rank. For the subsequent independent test (EEG data of the test group) the same transformation was applied.

As an alternative to the knowledge-based approach, we applied genetic algorithms for the reduction of variables. In general, the genetic algorithm (GA) [37–39] as a mathematical model mimics the natural survival of the fittest individual of a given population (selection) together with mutation and recombination within the genom of an individual. A population is defined as a set of input variables. Each individual represents a subset of variables (e.g. some spectral parameters) which act as input for an appropriate neural network. Initially, the population contains randomly chosen individuals. During the genetic algorithm, for each individual of the actual population the corresponding network will be built and trained on a subset L_i of

the complete lesson L . Then, the generalization of the trained network will be evaluated using a subset L_2 of L , which is disjoint from L_1 . According to this generalization, the GA calculates the fitness of each individual and the set of best individuals will remain within the population. All other individuals will be substituted by new ones which are built by recombination or mutation of those individuals which have 'survived': e.g. exchange input with an existing input of another surviving individual (recombination) or substitute it with an arbitrary new input (mutation). The advantage of this procedure is that it uses no a priori knowledge about the given data. Therefore, it is possible that unusual subsets will be selected and can yield good results for the corresponding neural network. However, this fact makes theory-driven considerations and interpretation of the selection hard or rather impossible.

Furthermore, genetic algorithms are global optimization algorithms, thus avoiding problems of local minima. Clear disadvantages are a usually very slow convergence compared with mathematical optimization methods and the fact that one reduces the effective length of the original lesson by introducing the disjoint subsets L_1 and L_2 where only L_1 will be used to train the network and L_2 is used by the genetic algorithm. In our case, L_2 is equivalent to the EEG data of the second split half (test group) of the linear approach.

In combination with the GA approach, a *neural sensitivity analysis* [40] was performed, based on a simplified network modularization approach (SNMA). This approach allows a uniform combination and concatenation of probably different network modules. Here we applied it to the concatenation of a linear preprocessing algorithm S with the above mentioned network N giving a system $N \circ S$. The preprocessing $S(x) = y$ calculates $y_i s_i$, so it depends on scaling parameter vector s with one component for each possible input component.

Using SNMA, the combined system $N \circ S$ was trained according to the desired task. After successful training, the scaling parameters s_i of the preprocessing S can be evaluated. For uniform scaled input vectors x , absolute small values of s_i indicate a small relevance for all network inputs of the component i . So, for network inputs i with $|s_i| < \delta$ for a given threshold δ , the input can be eliminated and the procedure can continue.

To improve the estimation for the significance, the training of the combined system will be repeated, starting with random synaptic weights for the network component N . The final elimination uses the mean values of s_i for the elimination condition.

Results

Discriminant Function Analysis

QEEG clearly shows differences between abstainers and relapsers in group 1 (training group resp. first split half) and group 2 (test group resp. second split half) as indicated in table 1. By means of QEEG, applying linear discriminant function analysis, relapsers can be classified correctly with high sensitivity (training group = 87.5%; test group = 88.0%), however, specificity (correct classification of abstainers) is far from being satisfying in the test set (training group = 92.9%; test group = 53.3%). This resulted in an overall correct classification rate of 90.2%

Table 1. Linear discrimination of relapsers and abstainers with five EEG variables

Group	n	Predicted group membership, %	
		abstainers	relapsers
Abstainers	14 (15)	92.9 (53.3)	7.1 (46.7)
Relapsers	24 (25)	12.5 (12.0)	87.5 (88.0)
Overall pooled correct classification		90.2 (70.7)	
Test of different group means at discriminant function: Degrees of freedom: 5 (p = 0.0007)			

Numbers refer to the training group, numbers in parentheses refer to the independent test group.

Table 2. Quadratic discrimination of relapsers and abstainers with four variables of second order

Group	n	Predicted group membership, %	
		abstainers	relapsers
Abstainers	14 (15)	92.9 (53.3)	7.1 (46.7)
Relapsers	24 (25)	4.2 (16.0)	95.8 (84.0)
Overall pooled correct classification		94.4 (68.7)	
Test of different group means at discriminant function: Degrees of freedom: 4 (p = 0.0001)			

Numbers refer to the training group, numbers in parentheses refer to the independent test group.

(training group) and 70.7% (test group). The result was achieved with a 5-variable model: compared with abstainers, relapsers showed especially less frontocentral alpha-activity and less left frontal relative theta-activity, a higher left frontal correlation dimension and, to a lesser extent, also an increased left temporal (relative) beta3-power, and increased left frontal (relative) beta3-power. Additional inclusion of variables did not improve classification results. Taken together, these results point to a frontally pronounced desynchronization (low-voltage EEG) in our relapsing patients.

Very similar results produced a discriminant function analysis using a polynomial model of second order (table 2). To the five variables chosen in the linear case we added all of their paired interactions. A stepwise procedure took only four interaction variables into account, yielding a better sensitivity for the training group (95.8%), but for the test group it was slightly worse (68.7%).

Table 3. Discrimination of relapsers and abstainers using ANN (genetic algorithm and pruning)

Training data OK, % (overall)	Test data OK, % (overall)	Hidden neurons	Electrodes/frequency bands
100	82.5	8	O1-delta O1-beta1 O1-beta3 O2-delta O2-theta O2-beta1 O2-beta3 F7-total F8-total T3-theta T3-beta1 T3-beta3 T4-beta1 T4-beta3 Fz-theta Fz-beta1 Fz-beta3 Cz-delta Cz-theta Cz-beta1 Cz-beta3

Table 4. Classification of relapsers (sensitivity) and abstainers (specificity) of the test group

Group	n	Predicted group membership, %	
		abstainers	relapsers
Abstainers	15	73.3	26.7
Relapsers	25	8.0	92.0

Table 5. Discrimination of relapsers and abstainers using ANN (knowledge-based)

Training group OK, % (overall)	Test group OK, % (overall)	Hidden neurons	Electrode/frequency bands
86.8	85.0	3	F7, F8, Fz (total, beta1, beta3)

Table 6. Classification of relapsers (sensitivity) and of abstainers (specificity) of the test group

Group	n	Predicted group membership, %	
		abstainers	relapsers
Abstainers	15	60.0	40.0
Relapsers	25	0.0	100.0

ANN Approach

In order to improve our results, we additionally applied ANN (supervised multilayer perceptron). Thus, we were able to include complex, nonlinear interactions between EEG variables in our analysis without any a priori

assumptions about the character of their functional relationship, which is not possible with classical linear statistics. The best results were obtained by absolute power spectra; relative power values and correlation dimension are significantly worse. In contrast to the discriminant analysis, inclusion of more than five variables improved the classification results.

Actually, the overall classification of ANNs, using absolute power spectra, could be increased up to 80.0% for the independent test group. The input data were z-transformed using means and standard deviations from the training and test sets, respectively. The actual input variables for the ANN were obtained by a genetic algorithm, working only on the data of the training group. Further input reduction (pruning) based on nonlinear sensitivity analysis, allowed an increase of the overall classification rate up to 82.5% (tables 3, 4).

Clearly, the number and selection of the electrodes and frequency bands show no similarity to the selection of the linear statistics (discriminant function analysis, DFA), indicating that the brain as a whole is involved and not only the frontal and temporal cortex. Hence, the very poor specificity (correct detection of abstainers) of the discriminant analysis could be raised from 53.3 to 73.3%. However, due to the unfavorable ratio of input dimension and number of observations, this result should be regarded with caution. Actually, only 3 more abstainers have been correctly classified. Nevertheless, this result was achieved with an independent test set.

A second investigation used predefined variable sets of unscaled spectral values, based on our knowledge from the DFA in order to obtain a more realistic comparison between stepwise discriminant analysis and neural network classification techniques and in order to be able to show that, based on frontal variables only, a classification is already possible. Indeed, an overall classification of 85% could be achieved, with a considerably smaller number of variables. However, this approach shows a similar

sensitivity/specificity ratio as the discriminant analysis, although on a significantly better performance level (tables 5, 6).

Discussion

The conclusion that can be drawn from our comparative analysis is that the application of ANN obviously offers the possibility to extract more relevant information from the EEG signal when compared with discriminant analysis, which resulted in a better classification of patients. Especially correct detection of abstainers was improved compared with discriminant analysis. However, it has to be taken into account that the network approach based on the genetic algorithm was only achieved with a large amount of variables which – despite independent classification of test data – may have resulted in so-called overfitting of the network function. This means that due to the large number of variables the possibility to improve classification power by chance is certainly increased. Actually, this restriction points to an important and major weakness of networks. Thus, the application of network usually needs large sample sizes to be able to take advantage of the possibility to create complex functions. Unfortunately, sample size of clinical trials is usually relatively small as for instance in our case. This problem can only be managed by the a priori selection of variables, which we have done with our knowledge-based approach and which also resulted in better classification results when compared with the discriminant analysis, but was less successful in comparison to the approach based on the genetic algorithm. Nevertheless, taking into account the current

discussion about advantages and drawbacks of ANN in comparison with conventional statistics, our investigation exemplifies that it may be reasonable from a practical point of view to apply both tools, especially when there is a paucity of a priori knowledge about the data structure and the nature of the discriminating function, respectively. Thus, in addition to the improvement of classification results, only by the application of both statistical tools were we able to localize all pathophysiologically relevant brain areas which would have remained undetected if only discriminant analysis had been applied. However, in the clinical context it is presumably more practical only to apply one of the applied tools.

Our results suggest that the function of the whole cortex, with emphasis on the prefrontal cortex, is affected in our relapsers. The application of only one of both statistical strategies would not have been sufficient to allow this conclusion. However, here we have to be aware of another major weakness of neural networks. Although classification results could be improved, pathophysiological interpretation of the network results is difficult because it only gives us an impression about the involved variables, but does not distinguish between more or less important variables.

Summarizing, our results suggest that neural network analysis of QEEG measures from patients with chronic alcoholism predicts abstinence or relapse with high sensitivity and satisfying specificity. In other words, almost all relapsers are classified correctly but some abstainers are misclassified as relapsers. Finally, our example demonstrates advantages and drawbacks of neural network technology which have to be taken into account in future studies.

References

- 1 Lippmann RP: An introduction to computing with neural nets. *IEEE ASSP Mag* 1987;4–22.
- 2 Hutchinson JM: A radial basis function approach to financial time series analysis; diss, Massachusetts Institute of Technology, 1994.
- 3 Gorman RP, Sejnowski TJ: Analysis of hidden units in a layered network to classify sonar targets. *Neural Networks* 1988;1:75–89.
- 4 Baxt WG: Use of an artificial neural network for data analysis in clinical decision-making: The diagnosis of acute coronary occlusion. *Neural Comput* 1991;2:480–489.
- 5 Fujita H, Katafuchi T, Uehara T, Nishimura T: Application of artificial neural network to computer-aided diagnosis of coronary artery disease in myocardial SPECT bull's-eye images. *J Nucl Med* 1992;33:272–276.
- 6 Weinstein JN, Kohn KW, Grever MR: Neural computing in cancer drug development: Predicting mechanism of action. *Science* 1992; 158:447–451.
- 7 Sittig DF, Orr JA: A parallel implementation of the backward error propagation neural network training algorithm: Experiments in event identification. *Comput Biomed Res* 1992;25:547–561.
- 8 Ebenbrandt L, Divine B, MacFarlane PW: Neural networks for classification of ECT ST-T segments. *J Electrocardiol* 1992;25:167–173.
- 9 Kippenhan JS, Barker WW, Pascal S, Nagel J, Duara R: Evaluation of a neural-network classifier for PET scans of normal and Alzheimer's disease. *J Nucl Med* 1992;33:1459–1467.
- 10 Floyd CE, Tourassi GD: An artificial neural network for lesion detection on single-photon emission computed tomography images. *Invest Radiol* 1992;27:667–672.
- 11 Furlong JW, Dupuy DE, Heinsimer JA: Neural network analysis of serial cardiac enzyme data. A clinical application of artificial machine intelligence. *Am J Clin Pathol* 1991;96:134–141.
- 12 Reibnegger G, Weiss G, Werner-Felmayer G, Judmaier G, Wachter H: Neural networks as a tool for utilizing laboratory information: comparison with linear discriminant analysis and with classification and regression trees. *Proc Natl Acad Sci USA* 1991;88:11426–11430.

- 13 Bugliosi R, Tribalto M, Avisati G: Classification of patients affected by multiple myeloma using a neural network software. *Eur J Haematol* 1994;52:182–183.
- 14 Warner B, Misra M: Understanding neural networks as statistical tools. *Am Statistician* 1996; 50:284–293.
- 15 Forström JJ, Dalton KJ: Artificial neural networks for decision support in clinical medicine. *Ann Med* 1995;27:509–517.
- 16 Selker HP, Griffith JL, Patil S, Long WJ, D'Agostino RB: A comparison of performance of mathematical predictive methods for medical diagnosis: Identifying acute cardiac ischemia among emergency department patients. *J Invest Med* 1995;43:468–476.
- 17 Rumelhart D, McClelland J (eds): *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, MIT Press, 1986, vol 1.
- 18 Churchland PS, Sejnowski TJ (eds): *The Computational Neuroscience*. Cambridge, MIT Press, 1992.
- 19 Spitzer M (ed): *Geist im Netz*. Heidelberg, Spektrum Verlag, 1996.
- 20 Hutton L, Sigillito V, Johannes RS: An interaction between auxiliary knowledge and hidden nodes on time to convergence. *Proc SCAMC*, 1989, pp 302–306.
- 21 Klöppel B: Application of neural networks for EEG analysis. Considerations of first results. *Neuropsychobiology* 1994;29:39–46.
- 22 Rappelsberger P, Magdolen J, Winterer G, Dorffner G, Flexer A: Classification of EEG of schizophrenics and depressives with artificial neural networks. *J Neurosurg Anesth*, in press.
- 23 Kubicki St, Herrmann WM, Laudahn G: *Factor Analysis and EEG Variables*. Stuttgart, Fischer, 1980.
- 24 Grassberger P, Proccacia I: Characterization of strange attractors. *Phys Rev Lett* 1983;50:346–349.
- 25 Grassberger P, Proccacia I: Measuring the strangeness of strange attractors. *Physica* 1983; 9D:189–208.
- 26 Winterer G, Postnikov N, Ziller M, Niestroj E, Frick K, Heinz A, Marquardt S, Lewinsky M, Herrmann WM: Signalkomplexität versus Spektralparameter in EEG-Zeitreihen von psychiatrischen Patienten: Eine retrospektive Klassifikationsstudie. *Z EEG EMG* 1995;26: 61–71.
- 27 Schuemer R, Ströhlein G, Gogolok J: *Datenverarbeitung und statistische Auswertung mit SAS, vol 2: Komplexe statistische Analyseverfahren*. Stuttgart, Fischer, 1990.
- 28 Klöppel B, Winterer G, Werner H: Detection of relapse in chronic alcoholism with neural networks and genetic algorithms. *Int J Intell Systems*, submitted.
- 29 Klöppel B: Neural networks as a new method for EEG analysis. *Neuropsychobiology* 1994; 29:33–38.
- 30 Rumelhart DE, Hinton GE, Williams RJ: Learning internal representations by error backpropagation; in McClelland JL, Rumelhart DE (eds): *Parallel Distributed Processing*. Cambridge, MIT Press, pp 318–362, 1986, vol 1.
- 31 Vogl TP, Mangis JK, Rigler AK, Zink WT, Alkon DL: Accelerating the convergence of the back-propagation method. *Biol Cybern* 1988; 59:257–263.
- 32 Dietrich C: *Einsatz von Schrittweitensteuerungen beim Training Neuronaler Netze (am Beispiel des Backpropagating Algorithmus unter Verwendung von Fuzzy Logik)*; Master-Thesis, Universität Kassel, 1996.
- 33 Rigler AK, Irvine JM, Vogl TP: Rescaling of variables in back propagation learning. *Neural Networks* 1991;4:225–229.
- 34 Omlin CW, Giles CL: *Pruning Recurrent Neural Networks for Improved Generalization Performance*. Troy, Rensselaer Polytechnic Institute, 1993.
- 35 Hinton GE: Connectionist learning procedures. *Artif Intell* 1989;40:185–234.
- 36 Finnoff W, Hegert F, Zimmermann HG: Improving model selection by nonconvergent methods. *Neural Networks* 1993;6:771–783.
- 37 Ruhland J: *Topologie veränderten Lernens in Neuronalen Netzwerken mittels Genetischer Algorithmen*; Diplom-Arbeit, Universität Kassel, 1995.
- 38 Harp SA, Samad T, Guha A: Towards the genetic synthesis of neural networks. *Proc 3rd Int Conf Genetic Algorithms*, Morgan Kaufmann, 1989, pp 360–369.
- 39 Booker S, Lashon B: Representing Attribute-Based Concepts in a Classifier System. *Foundations of Genetic Algorithms*, Morgan Kaufmann, 1991, pp 115–127.
- 40 Klöppel B: *Stabilität und Kapazität neuronaler Netzwerke am Beispiel der EEG-Analyse*. Shaker, 1994.