

An MLP-based model for identifying qEEG in depression

Sushmita Mitra, Suptendra Nath Sarbadhikari, Sankar K. Pal*

Machine Intelligence Unit, Indian Statistical Institute, Calcutta 700035, India

Received 4 August 1995; revised 29 May 1996; accepted 4 June 1996

Abstract

Manual differentiation of electroencephalography (EEG) paper recordings in cases of depression is not very helpful. So, a Multilayer Perceptron (MLP) has been used to differentiate the EEG power density spectra (qEEG) in the wakeful state from animals (control, exercised and depressed). The qEEG ranging from 1 to 30 Hz, at 1 Hz increments (30 input features) and also as slow, medium and fast activity (represented by three ranges of frequencies at the input) were used. After training with depressed and control qEEG only, the MLP has been found to distinguish successfully between the normal and the depressed rats in more than 80% of the cases, identifying, in the process, most of the exercised groups' EEG as normal. The reduction in the dimension of input features from 30 individual frequencies to 3 frequency bands has produced similar results. The rules generated for making such distinctions have been found to be similar to the clinical views.

Keywords: EEG power spectra (qEEG); Depression; Rats; Chronic physical exercise; Multilayer perceptron; Rule generation

1. Introduction

Depression is one of the common underdiagnosed diseases in clinical psychiatry. Electroencephalography (EEG) in the form presently applied contributes very little diagnostically or prognostically in determining the extent of depression. EEG reflects the

electrical activity of the brain during the various states of sleep and wakefulness. However, instead of analyzing only the frequency and amplitude changes, obtaining power spectra by fast Fourier transform (FFT) of EEG signals often conveys more information. This is also known as quantitative EEG or quantified EEG or qEEG. FFT is not superior to time-domain transforms, rather it is more informative than the conventional analog recordings which are still preferred by

* Corresponding author. e-mail: sankar@isical.ernet.in

clinicians. Any transform, being derived from the original signal, is always likely to lose some information. However, frequency domain processing makes the findings more objective (quantified) and hence, observer independent. Power density spectra measurement is certainly a form of qEEG, albeit not the only one.

In most cases, before experimenting with human beings, it is customary to test the validity of results in animals. Animal models of depression are now available [1]. Chronic physical exercise has been found to be effective in combating stress, in both human beings and animals [2,3]. The preventive or prophylactic beneficial role of exercise in an animal model of depression has also been proven through a neural net [4].

Artificial neural networks (ANN) [5–7] can be formally defined as ‘massively parallel interconnections of simple (usually adaptive) processing elements that interact with objects of the real world in a manner similar to biological systems’. The benefit of neural nets lies in the high computation rate provided by their inherent massive parallelism, thereby enabling real-time processing of huge data sets with proper hardware backing. All information is stored distributed among the various connection weights. The networks can be trained by examples (as is often required in real life) and sometimes generalize well for unknown test cases. Recently, ANNs have also been used for inferring and rule generation in expert systems [8]. Medical experts rely on linguistic labels of any symptom or sign for reaching a proper decision. So, if an expert system can provide linguistically the rules for its inferences, it can be of great help to the clinician. For EEG analysis, the use of connectionist approaches has also been reported [9–12]. For further review on automated EEG analysis, one may consult [13–15].

Recently, an ANN (Madaline) has been successfully used for distinguishing the EEG power spectra (qEEG) in normal, depressed and exercised rats [4]. In that case, Madaline Rule I [16] had been used with a first layer of 30 hard-limited (signum) adalines whose outputs purveyed inputs to a second layer of majority-vote-taker. In that work, rules were not generated. There the results in the wakeful state were not as good as in the sleep states. However, assessing the wakeful state is more important for long term monitoring, especially in humans.

In the present article, we have three objectives. Firstly, to investigate whether a better architecture (MLP) can classify qEEG more efficiently in the wakeful state. Secondly, using 30 features as input requires much greater computation time. So, whether the dimensions can be reduced, without sacrificing a significant amount of information, has also to be verified. Apart from that, the rule generation capability for inferring will be investigated, followed by comparison with clinical views. A description of the MLP-based algorithm for rule generation is given in Section 2. In Section 3 we provide the results on the qEEG data for both classification and rule generation. In Section 4, the conclusions and discussion are presented.

2. Rule Generation in multilayer perceptron (MLP) based model

Consider the MLP of Fig. 1. The input and output of node j at layer h are denoted by x_j^h and y_j^h , respectively, while the connection weight between nodes i and j in layers h and $h-1$ is given by w_{ji}^{h-1} . The network is used for rule generation in ‘If-Then’ form [8]. These rules describe the extent to which a test pattern belongs or does not belong to one of the classes in terms of antecedent and conse-

quent clauses. For this, we backtrack along maximal weighted paths using the trained net and utilize its input and output activations.

After training, the connection weights of an MLP encode among themselves (in a distributed fashion) all the information learned about the input-output mapping. Therefore, any link weight with a large magnitude reflects a strong correlation between the connecting neurons. We utilise this property in evaluating the importance of an input node (feature) on an output layer mode (decision). For this we need to compute the path weights from each output node to each input node through the various hidden nodes. Each output-input path that is maximum (through any hidden node) denotes the importance of that input feature for arriving at the corresponding output decision. Note that one or more hidden layers may be considered (if necessary) for evaluating the path weights. An increase in the number of hidden layers simply leads to an increase in the possible variety of paths being generated through the various hidden nodes in the different layers.

An input pattern F_p from the training set is presented to the input of the trained network and its output is computed. The consequent part of the corresponding 'If-Then' rule is generated from the output node activation.

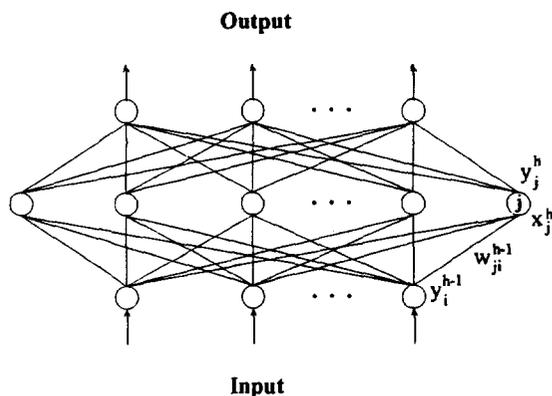


Fig. 1. The three-layered MLP model.

To find the antecedent clauses of the rule, we backtrack from the output layer to the input through the maximal weighted links.

2.1. Output layer

Let the user ask for the justification about a conclusion regarding class j . Starting from the output layer $h = H$, the process continues in a 'top-down' manner until the input layer ($h = 0$) is reached. In the first step, for layer H , we select those neurons i in the preceding layer that have a positive impact on the conclusion at output neuron j . Hence we choose neuron i of layer $H - 1$ if $w_{ji}^{H-1} > 0$. Let the set of m_{H-1} neurons of layer $H - 1$, so selected, be $\{a_1^{H-1}, a_2^{H-1}, \dots, a_{m_{H-1}}^{H-1}\}$ and let their connection weights to neuron j in layer H be given as $\{\text{wet}_{a_i^{H-1}} = w_{ja_1}^{H-1}, \dots, \text{wet}_{a_{m_{H-1}}^{H-1}} = w_{ja_{m_{H-1}}}^{H-1}\}$. For the remaining layers we obtain the 'maximum weighted' paths through these neurons down to the input layer.

2.2. Intermediate layers

We select neuron i in layer $0 < h < H - 1$ if $y_i^h > 0.5$,

and

$$\text{wet}_{i,h} = \max_{a_k^{h-1}} [\text{wet}_{a_k^{h-1}} + w_{a_k^h}^h] \quad (1)$$

such that $\text{wet}_{i,h} > 0$. Let the set of m_h neurons so chosen be $\{a_1^h, a_2^h, \dots, a_{m_h}^h\}$ and their cumulative link weights to neuron j in layer H be $\{\text{wet}_{a_1^h}, \text{wet}_{a_2^h}, \dots, \text{wet}_{a_{m_h}^h}\}$ respectively, by Eq. (1). Note that this heuristic ensures that each of the selected m_h neurons have a significant output response y_i^h . This implies choosing a path with neurons that are currently active for deciding the conclusion that is being justified. It also enables each neuron i to lie along one of the 'maximum weighted' paths from the input layer ($h = 0$) to the output

node j in $h = H$, by choosing only one of the m_{h+1} previously selected paths that provides the largest net weight wet_{jh} .

2.3. Input layer

Let the process of Eq. (1) result in m_0 chosen neurons(paths) in(from) the input layer ($h = 0$). These neurons indicate inputs that are 'known' and have contributed to the ultimate positivity of the conclusion at neuron j in the output layer H . It may happen that $m_0 = 0$, such that no clear justification may be provided for a particular input–output case. This implies that no suitable path can be selected by Eq. (1) and the process terminates.

Let the set of the selected m_0 input neurons be $\{a_1^0, a_2^0, \dots, a_{m_0}^0\}$ and their corresponding path weights to neuron j in layer H be $\{wet_{a_1^0}, wet_{a_2^0}, \dots, wet_{a_{m_0}^0}\}$. We arrange these neurons in decreasing order of their net impacts, where we define the net impact for neuron i as

$$\text{net impact}_i = y_i^0 \times wet_i^0$$

Then we generate clauses for an If-Then rule from this ordered list until

$$\sum_{i_s} wet_{i_s^0} > 2 \sum_{i_n} wet_{i_n^0} \quad (2)$$

where i_s indicates the input neurons selected for the clauses and i_n denotes the input neurons remaining from the set $\{a_1^0, a_2^0, \dots, a_{m_0}^0\}$ such that

$$|i_s| + |i_n| = m_0$$

and $|i_s|$, $|i_n|$ refer, respectively, to the number of neurons selected and remaining from the said set. This heuristic allows selection of those 'currently active' input neurons contributing 'the most' to the final conclusion (among those lying along the maximum weighted paths to the output node j) as the

clauses of the antecedent part of a rule. Hence, it enables the 'currently active' test pattern inputs (current evidence) to influence the generated 'knowledge base' (connection weights learned during training) in producing a rule to justify the 'current' inference.

The complete 'If' part of the rule is found by ANDing the clauses corresponding to each of the features.

3. Implementation and results

The data was collected from male rats (*Rattus norvegicus*) of Charles Foster strains. They were divided into four groups:

- (1) Depressed by applying chronic intermittent stress for two weeks.
- (2) Exercised for eight weeks on treadmill at 20 m/min for six days a week.
- (3) Exercised as above for six weeks, followed by concomitant stress and exercise for two weeks.
- (4) Control (handled and operated only).

The EEG was recorded through an eight channel polygraph (Medicare, India) on paper and sampled at 256 Hz through a 12-bit analog-digital converter (Micronics, India) into the hard-disk of a PC-AT (HCL, India). The digitized data was filtered through a second order, twice-cascaded Butterworth (Infinite Impulse Response) filter. The EEG power spectra was calculated by an FFT routine from the digitized and filtered signals. Three successive four s artifact-free epochs were averaged [4]. The input features were percentages of FFT of EEG at each integer frequency from 1 to 30 Hz. All 30 input features, as in the previous work [4], were used at first. Then the input features were reduced to three: slow activity (1–4 Hz), medium activity (5–15 Hz) and fast activity (16–30 Hz).

Table 1
Recognition score (%) on qEEG data with training set (30+30)

Model	MLP							Bayes		k-NN				
	30		3		3		3		3		3		3	
Input features	30		3		3		3		3		3		3	
No. of layers	3	4	5	3	3	4	5	—	$k=1$	$k=3$	$k=5$	$k=7$		
Hidden nodes	30	5:5	5:5:5	5	15	5:5	5:5:5	—	—	—	—	—		
Training (overall)	100	100	100	96.67	98.33	96.67	95	—	—	—	—	—		
Testing (overall)	82.14	82.14	83.93	80.36	80.36	80.36	80.36	69.6	82.14	80.36	76.79	78.57		
Depressed	100	100	100	96.67	96.67	96.67	96.67	80	100	96.7	93.3	90.1		
Normal	61.54	61.54	65.38	61.54	61.54	61.54	61.54	57.69	61.54	61.54	57.69	65.38		

For training, 30 samples from the depressed group and 30 from the control group were used. In two other cases, 20 and 25 samples respectively, from each of the groups were used. For testing, 30 from the depressed group, 6 from the control group and 10 each from the exercised only and exercised + stressed groups were used. All the data files were chosen randomly. During training, the files from the exercised groups (b and c) were not used, but the testing phase was conducted on all four (a to d) groups. All the samples were chosen randomly. The data from the training sets were not used for testing. Therefore, truly independent training and testing sets were obtained.

3.1. Classification

The recognition scores obtained using MLP corresponding to the training samples' size of 30 + 30 (Table 1), 25 + 25 (Table 2) and 20 + 20 (Table 3) from the depressed and normal groups, respectively, with different number of layers, the k-NN classifier and the Bayes classifier.

Only those hidden node configurations yielding the best results have been shown in the table. Note that $n_1:n_2:n_3$ nodes, in the

third row, refer to n_1, n_2, n_3 nodes in the first, second and third hidden layers, respectively. In the tables, the number of layers = 3, denote the input, the output and one hidden layer. Similarly, layers = 4 and 5 denote 2 and 3 hidden layers, respectively.

Here, the Bayes classifier has been used for multivariate normal patterns with a priori probabilities $p_i = C_i/N$, where C_i indicates the number of patterns in the i th class and N is the total number of pattern points. The covariance matrices are different for each pattern class.

The k -nearest neighbors (k -NN) classifier [17], with $k = 1, 3, 5, 7$, has also been implemented. The k -NN classifier is reputed to be able to generate piecewise linear decision boundaries and, thereby, is quite efficient in handling concave and linearly nonseparable pattern classes. Therefore, a comparison of the performance of the neural net model with that of the k -NN classifier is highly desirable.

It may be noted that in the earlier work [4], with Madaline Rule I, with 30 input features and 30 adalines, the overall percentage recognition of the test set was 81. In case of depressed group, it could correctly recognize only 86%, but it recognized 79.33% of the other groups as normal. However, the better

Table 2
Recognition score (%) on qEEG data with training set (25+25)

Model	MLP								Bayes	k-NN			
	30		3		3		3			3	3	3	3
Input features	30		3		3		3		3	3			
No. of layers	3	4	5	3	3	4	5	—	$k = 1$	$k = 3$	$k = 5$	$k = 7$	
Hidden nodes	30	5:5	5:5:5	5	15	5:5	5:5:5	—	—	—	—	—	
Training (overall)	100	100	100	100	100	100	100	—	—	—	—	—	
Testing (overall)	76.8	76.8	75	82.14	82.14	80.36	80.36	69.64	80.36	80.36	78.57	80.36	
Depressed	93.33	90	93.33	96.67	93.33	90	90	80	96.67	96.67	93.3	93.3	
Normal	57.69	61.54	53.85	65.38	69.1	69.1	69.1	57.69	61.54	61.54	61.54	65.38	

recognition for the normal groups was due to the fact that the training set had a hundred samples from each group.

The overall recognition scores for the training and test sets are provided along with the individual classwise performance for the test set only. It is observed that reducing the dimension of input features from 30 to 3 leads to no significant degradation in performance. Using more than one hidden layer yields better results and this is seen from the tables. It has been pointed out [7] that two hidden layers are necessary to model concave/disjoint decision regions. Often more hidden layers enable the network to discriminate more efficiently between linearly non separable classes due to the higher degrees of freedom available.

The results are provided for 20 + 20, 25 + 25 and 30 + 30 training samples. It is observed that 3 features (ranges) yield better scores than 30 features. This indicates that the selection of the three ranges of frequencies is reasonably good. Therefore, besides faster computation there is the added advantage of increased efficiency by reduction in input feature dimension. Only the salient points are selected from the three ranges and this is also justifiable clinically since the slow

and the fast EEG activities are indeed altered in depression.

Bayes classifier gives the best decision surface for normal densities. Here we observe that, the recognition scores are poorer compared to that of the MLP. Comparing with the k -NN is naturally a better option. It is found that the performance of the MLP is often better and sometimes comparable to that of the k -NN. Here also, three features are found to be sufficient to model the classification problem.

Another interesting feature observed is that this model is able to classify most of the exercised rats' (with or without concomitant stress) qEEG as normal, as in the previous study [4].

3.2. Rule generation

Rules have been generated for the three input case according to Eq. (1). Let the three features be represented as F_1 , F_2 and F_3 corresponding to EEG activity levels 'slow', 'medium' and 'fast,' respectively. The rules obtained are enumerated as follows: If F_1 is large and $F_2 > 0$ and F_3 is small then class 1 (Depressed) and if F_1 is small and $F_2 < 0$ and F_3 is large then class 2 (Normal).

Table 3
Recognition score (%) on qEEG data with training set (20+20)

Model	MLP							Bayes		k-NN				
	30		3		3		3		3		3		3	
Input features	30		3		3		3		3		3		3	
No. of layers	3	4	5	3	3	4	5	—	$k=1$	$k=3$	$k=5$	$k=7$		
Hidden nodes	30	5:5	5:5:5	5	15	5:5	5:5:5	—	—	—	—	—		
Training (overall)	100	100	100	100	100	100	100	—	—	—	—	—		
Testing (overall)	78.6	73.2	76.8	80.36	80.36	80.36	80.36	71.43	76.79	80.36	80.36	75		
Depressed	86.67	80	86.67	93.33	96.67	93.33	90	83.33	93.3	96.67	100	86.67		
Normal	69.1	65.38	65.38	65.38	61.54	65.38	69.1	57.69	57.69	61.54	57.69	61.54		

Note that by Eq. (2) we descend at input nodes having activation greater than 0.5, along the maximally weighted paths [8].

The neuronal activation is the product of the weight and signal value traveling through the path. We have actually calculated the weights, as well as, the products of weights and respective signal values for deducing a particular rule. In both the cases the results have been similar in the present investigation.

In this article we modify the approach slightly to suit the requirement of the data domain. Here we use the conventional approach of n input nodes instead of the $3n$ linguistic input nodes as in [8]. Therefore, we compare the product $w_{ji}^0 y_{iA}^0$ corresponding to F_1 and F_3 , such that the feature providing a relatively lower value is assigned a linguistic label 'small' while the other is defined as 'large'.

It is observed that the two rules match partly with the statistical analysis of the data [4,18]. In depression, the slow wave (δ) EEG activity is increased and the fast wave (β) EEG activity is reduced. Incidentally, the role of F_2 cannot be explained from the existing clinical ideas and the experimental findings. The reasons for this may lie in the fact that the medium EEG activity defined here en-

compasses two bands: θ and α , and a sub-band β_1 (beta-1) or slow beta activity.

4. Conclusions and discussion

Manual interpretation of EEG paper recordings in depression is almost impossible. So, a multilayer perceptron (MLP) has been used for automated distinction of the EEG power density spectra (qEEG), in the wakeful state, in animals (control, exercised and depressed). The model has also been used for rule generation on the qEEG data and comparison with the clinical expertise.

Since the results with 3 reduced features (3 ranges of frequencies) have been found to be comparable to those with 30 original input features (individual frequencies), one can exploit this dimension reduction capability. It will naturally make a neural net converge much faster even while dealing with huge databases and enable a diagnosis to be reached more quickly. Note that, more than one hidden layer yields better results.

The generated rules, if comprehensible clinically, will add to the diagnostic validity of automated qEEG detection in various neuropsychiatric conditions. In general, doctors

deal with linguistic variables to reach a proper diagnostic decision. If the expert systems can produce such linguistic rules for discriminating between various clinical conditions, it will be of great help to the clinician.

In future, such automated methodologies may also be applied to human beings. ANN can become a very useful tool in this field, leading to a greater concurrence between the diagnosticians of various laboratories, unlike at present. Building up a proper database for psychiatric conditions, like depression, based on the qEEG findings and the rules generated by the ANN, may help in realizing reliable connectionist expert diagnostic systems.

In schizophrenia type 2 (with predominantly negative symptoms), there are behavioural similarities with depression. These may be reflected in the qEEG which can then be compared with the qEEG in depression.

However, for psychiatric assessment by computers, an accompanying human clinician is essential [19]. Computerized assessments and clinical acumen are not mutually exclusive, rather they reinforce each other [20].

Acknowledgements

This work was carried out while Prof. S.K. Pal held a Jawaharlal Nehru Fellowship and Dr. S.N. Sarbadhikari held a Research Associateship in a Project (No. 22(235)/93/EMR-II) of the Council for Scientific and Industrial Research, India. The data was collected by Dr. Sarbadhikari while he was a Senior Research Fellow in the School of Biomedical Engineering, Institute of Technology, Banaras Hindu University.

References

- [1] Thiébot MH, Martin P and Puech AJ: Animal behavioural studies in the evaluation of antidepressant drugs, *Brit J Psychiatry*, 160 (Suppl.15) (1992) 44–50.
- [2] Harris SS, Caspersen CJ, DeFries GH and Estes Jr. EH: Physical activity counseling for healthy adults as a primary preventive intervention in the clinical setting report for the US preventive services task force, *J Am Med Assoc*, 261 (1989) 3590–3598.
- [3] Overton JM, Kregel KC, Davis-Gorman G, Seals DR, Tipton CM and Fisher LA: Effects of exercise training on responses to central injection of CRF and noise stress, *Physiol Behav*, 49, (1991) 93–98.
- [4] Sarbadhikari SN: A neural network confirms that physical exercise reverses EEG changes in depressed rats, *Med Eng Physics*, 17(8) (1995) 579–582.
- [5] Rumelhart DE and McClelland JL: *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, Vol. 1, MIT Press, Cambridge, MA, 1986.
- [6] Hertz J, Krogh A and Palmer RG: *Introduction to the Theory of Neural Computation*, Addison-Wesley, Reading, MA, 1994.
- [7] Lippmann RP: An introduction to computing with neural nets, *IEEE Acoustics, Speech Signal Processing Magazine*, 4 (1987) 4–22.
- [8] Mitra S and Pal SK: Fuzzy multi-layer perceptron, inferencing and rule generation, *IEEE Trans Neural Networks*, 6 (1995) 51–63.
- [9] O'Boyle DJ, Choi EWK, Conroy G and Turega M: Learned classification of EEG power spectra using a neural network. In *J Physiol, Proceedings of the Physiological Society (Sheffield Meeting, U.K., 19–20 April, 1991)* 1991, pp. 345P.
- [10] Bruha I and Madhavan GP: Need for a knowledge-based subsystem in evoked potential neural net recognition system. In *Proceedings of 11th Annual International Conference of IEEE-EMBS*, IEEE Press, New York, 1989, pp. 2042–2043.
- [11] Hiraiwa A, Shimohara K and Tokunaga Y: EEG topography recognition by neural networks, *IEEE Eng Med Biol*, 9 (1990) 39–42.
- [12] Klöppel B: Neural networks as a new method for EEG analysis: a basic introduction, *Neuropsychobiology*, 29 (1994) pp. 33–38.
- [13] Duffy FH: The role of quantified EEG in psychological research. In *Human Behavior and the Developing Brain* (Eds: G Dawson and KW Fischer), Guilford Press, New York, 1994, pp. 93–132.

- [14] Jansen BH and Cheng WK: Structural EEG analysis: An explorative study, *Int J Biomed Comput*, 23 (1988) 221–237.
- [15] Jansen BH: Quantitative analysis of electroencephalograms: Is there chaos in future?, *Int J Biomed Comput*, 27 (1991) 95–123.
- [16] Widrow B and Lehr MA: 30 years of adaptive neural networks: Perceptron, Madaline and Back-propagation, *Proc. IEEE*, 78 (1990) pp. 1415–1440.
- [17] Tou JT and Gonzalez RC: *Pattern Recognition Principles*, Addison-Wesley, London, 1974.
- [18] Sarbadhikari SN, Dey S and Ray AK: Chronic exercise alters EEG power spectra in an animal model of depression, *Indian J Physiol Pharmacol*, 40(1) (1996) 47–57.
- [19] First MB: Computer-assisted assessment of DSM-III-R diagnoses, *Psychiatr Ann*, 24 (1994) 25–29.
- [20] Sarbadhikari SN: Medical informatics—are the doctors ready?, *J Indian Med Assoc*, 93 (1995) 165–166.