

# UNSUPERVISED LEARNING IN COMPUTER AIDED MACRO ELECTROMYOGRAPHY

C.N. Schizas <sup>1&2</sup>, C.S. Pattichis <sup>2</sup>, R.R. Livesay <sup>1</sup>  
I.S. Schofield <sup>3</sup>, K.X. Lazarou <sup>2</sup>, L.T. Middleton <sup>2</sup>

<sup>1</sup> University of Indianapolis, 1400 East Hanna Avenue, Indianapolis, IN 46227, USA

<sup>2</sup> The Cyprus Institute of Neurology and Genetics, P.O. Box 3462, Nicosia, Cyprus

<sup>3</sup> Department of Neurophysiology, Newcastle Upon Tyne General Hospital, U.K.

## Introduction

Neuromuscular disorders, although relatively uncommon, constitute a significant cause of disability in both the young and old. Of the many disorders that have been clinically identified, two basic pathological processes have been found. Either muscle fibers may be lost through a degenerative process or there is loss of the motor neurones and their axons. The first instance is termed a myopathy, the second a neuropathy or neurogenic process. A large number of clinically distinct muscle disorders are not recognized. For this study normals and patients from three disorders have been selected for investigation.

1. *Motor Neurone Disease (MND)*: A rapidly progressive neurogenic disorder due to loss of cells associated with the voluntary motor system including the anterior horn cells. This process is typically seen in the older age groups and is in general not a hereditary disorder, but a familial form is recognized. The diagnosis can be made on clinical grounds.

2. *Becker Muscular Dystrophy (BMD)*: A slowly progressive myopathy due to a hereditary biochemical disorder of muscle fibers which is typically seen in younger age groups. Although the clinical features are well known there is also a diagnostic test for this disorder by demonstrating quantitative change of the protein Dystrophin in muscle fibers.

3. *Spinal Muscular Atrophy (SMA)*: A slowly progressive neurogenic disorder due to patchy loss of anterior horn cells. In general SMA is seen in the younger age groups and may be hereditary. Both the clinical features and muscle biopsy appearance are highly variable making diagnosis of this disorder difficult.

Changes in the motor units which occur in myopathies and neuropathies may be detected by the Macro (electromyography) EMG technique [1],[2]. Macro EMG is a valuable diagnostic technique in the investigation of these conditions. To date standard statistical procedures have been used to classify the results of a Macro EMG study. Recent developments in the field of Artificial Neural Networks (ANN) have raised the possibility of using these powerful techniques to analyze the Macro EMG data [3],[4] in a way that makes no assumptions upon the relationships between the parameters and without recourse to conventional modelling methods; [3] and [4] have employed supervised learning procedures. In particular, unsupervised learning procedures are used on the data which makes no apriori assumptions as to the clinical diagnosis which was obtained by means other than the neurophysiological findings.

Learning in ANN is basically achieved through systematic training. Training an artificial neural network is a matter of adjusting weights, either manually or automatically. Irrespective of weight adjustment, neural network training can take place in one of three ways: supervised, unsupervised, and self-supervised.

In supervised training, input and output data are supplied to the network in an effort to teach it to produce the desired output vector. In unsupervised training, data is simply entered into the network without any human intervention or side information. Training is achieved through the formation of internal constructions that capture regularities in their input vectors. In self-supervised training, the network monitors itself and corrects errors in the interpretation of data by feedback through the network. In each training method, learning is achieved when the network reaches a certain relative stability. In practice, after the initial connecting weights are set to small random values, input data is supplied to the network, thus causing it to pass through state changes which subsequently introduce changes to the initial values of the weights, until stability is reached when no further weight changes are caused. In effect, a neural network learns through adaptation. The learning rule, based on the adopted methodology, is the very heart of a neural network.

Recent studies [3],[4] have demonstrated how the supervised paradigm can be adopted through back propagation in building ANN models for classifying electromyographic signals. Even though these studies have produced much better results than any other known method, they were somewhat limited by the criteria which were used for producing the training and evaluation sets of data. It was stated earlier that the diagnosis in the pathological groups was used on the basis of clinical and/or muscle biopsy grounds giving considerable weight in certain cases to the clinical opinion of the diagnosis. Additional complication to the problem are introduced by the heterogeneous nature of the SMA group and the chronic or acute nature of the MND and BMD groups. EMG data from normal subjects can have, also, high degrees of variability. These inherited complications make the grouping of subjects of the same disease very difficult. In an earlier study [3] the problem of grouping was reduced by introducing subgroups among diseases.

## Macro EMG method

The Macro EMG method uses a special needle electrode with a 15mm by 0.8mm diameter cannula with a 25 $\mu$ m diameter side port electrode 7.5mm from the tip. The side port electrode records the activity of one or more single muscle fibers from an individual motor unit recruited by gentle voluntary contraction of the muscle. The single fiber action potentials (SFAP) are used to trigger a signal averaged into which the cannula signal is fed. In general about 200 discharges of the unit are averaged with an 80ms sweep to obtain a Macro motor unit potential (MMUP) [5]. At least 20 MMUPs are measured from a single muscle to obtain a reasonable estimate of the parameters of an average motor unit potential.

The MMUP data is analyzed by means of the peak-to-peak amplitude and the area and average power of the central 60ms of the signal, as it is shown in figure 1. An additional parameter to estimate the duration of the potential has been introduced [6] which is the 90% power duration; i.e., the width of the region of the potential that contains 90% of the power. In this figure the short vertical lines show duration beginning and ending points. Normal ranges for the amplitude and area by age decade have also been obtained [7]. Following extensive simulation studies [8], the amplitude and area were found to be highly correlated and proportional to the size; i.e., the number of fibers in the motor unit. The duration parameter remains to be studied further, but initial indications are that this relates to the range of muscle fiber diameters and the geometric width of the end-plate zone of the motor unit.

## Self Organization through Unsupervised Learning

Kohonen's self-organizing feature maps [9] have been employed in this study as an unsupervised

learning paradigm for macro FMG classification. The ANN model has been implemented using the C programming language, and resembles a two dimensional array of neuron-like logic units which are weight-connected to an input pathway where the feature values, mentioned earlier, describing each patient, are supplied in a vector form. Continuous-valued input vectors are presented sequentially in time without specifying the desired output. The presentation of a number of input vectors and the adjustment of weights accordingly will lead to a two-dimensional map with weights specifying clusters that resemble the input space such that the point density function of the vector centers tends to approximate the probability density function of the input vector. One may say then, that this is an adaptive process similar to that of the sensory pathways in the brain, during which the neuron cells are tuned to certain features of the input signal.

The algorithm responsible for weight adjustments is described below [10]:

**Step 1:** Initialize weights from N input to M output nodes with small random values (0.4 - 0.6). Set the initial radius of the neighborhood to cover the whole array.

**Step 2:** Present new input.

**Step 3:** Compute the Euclidean distance, or an arbitrary norm  $d_j$  between the input and each output node  $j$  using:

$$d_j = \sum_{i=0}^{N-1} (x_i(t) - w_{ij}(t))^2$$

where  $x_i(t)$  is the input to node  $i$  at time  $t$  and  $w_{ij}(t)$  is the weight from input node  $i$  to output node  $j$  at time  $t$ .

**Step 4:** Select the output node C with the minimum distance  $d_j$ .

**Step 5:** Update weight to node C and neighborhood  $N_c(t)$  using:

$$w_{ij}(t+1) = A_1 x_i(t) + A_2 w_{ij}(t)$$

$$j \in N_c(t) \quad 0 \leq i \leq N-1 \quad A_1 = A(1 - \frac{T_1}{T}) \quad A_2 = 1 - A_1$$

with  $A$  being a monotonically decreasing gain factor  $0 < A < 1$ , and  $T_j$  being the period for updating the radius of the neighborhood.

**Step 6:** Go to Step 2.

After the completion of one training cycle (epoch), i.e., after all vectors in the training set are presented once at the input, the whole procedure is repeated with the input vectors presented in a different random order each time. This exercise will eventually organize the weights such that topologically close nodes become sensitive to inputs that are physically similar. Output nodes will then be ordered in a natural manner.

## Procedure/Results

The data from thirty-six Macro EMG studies recorded from the biceps brachii were used for analysis. The study groups are as follows: Seven normal subjects (class 1), nine MND (class 2), fourteen BMD (class 3), and six SMA (class 4), patients.

This set of data was used for forming a training group of 23 subjects (5 normals, 7 MND, 7 BMD, and 4 SMA), and an evaluation group of the remaining 13 subjects (2 normals, 2 MND, 7 BMD, and 2 SMA).

The diagnosis of patients in the three pathological groups was made on the basis of clinical grounds for the MND group, clinical and muscle biopsy data for the BMD group including Dystrophin assay (with the exception of 3 patients), and clinical and biopsy data for the SMA group. The SMA group caused a number of problems from the clinical diagnostic viewpoint. This particular pathology is only slowly progressive and is characterized by rather patchy involvement of the muscles, and is sometimes evident in only one limb. As a result, the muscle biopsy can be normal in the muscle Vastus Lateralis usually studied for this purpose. Therefore, for the SMA group, considerable weight has been given to the clinical opinion of the diagnosis.

A number of different models were tested and the results obtained are summarized in Table 1. The various models were derived as follows: (1) For models 1-6 and 12-15 the mean and the standard deviation of the parameters for each subject formed an 8 input vector; (2) For models 7-11 and 16-20 only the mean of the parameters for each subject were used thus forming a 4 input vector; (3) For models 1-3, 7-9, 12-13 and 16-17, a 10x10 output grid was used, and for the rest of the models an 8x8 output grid was used; (4) Other parameters such as the gain factor  $A$  and the number of epochs were varied for determining their effect in the learning process; (5) the heterogeneous nature of the data obtained from the SMA group was the reason for building models 12-20 which were trained by excluding all the SMA patients.

For each subject in the training set there will be an output node for which the subject's data is causing the maximum response. On the map in figure 2, are shown the node and the subject which has caused the maximum response at that node, for model 1. Nodes with zero value are the less sensitive ones. If, for example, one selects a 10x10 output grid, then with a training set of 36 patients, at least 64 nodes will not be assigned to anyone of the input vectors, and remain as "zero" nodes. Unknown vectors then, falling on "zero" nodes during the evaluation process will not be diagnosed.

In an effort to assign as many output nodes as possible to subjects in the training set, thus enhancing the diagnostic power of the net, all the nodes which have not been assigned to any input vector in the training set, are reconsidered and assigned to the input vectors for which maximum response is obtained, when compared to the responses the other input vectors are producing. This criterion does not require that each node is necessarily assigned to an input vector. If the maximum response produced by a certain node is not high enough (determined by the confidence level), the node will be left as a "zero" node. The computer program offers to the trainer the facility for selecting confidence levels for assigning "zero" nodes to input vectors, during the training phase.

Likewise, during the evaluation phase the facility for selecting the confidence levels for determining the diagnostic yield for the evaluation set is provided. This flexibility adds another dimension to the diagnosis, thus making it more reliable and realistic.

The confidence level is related to the diagnostic yield as follows: A "zero" node is assigned to a certain input vector if the maximum response produced by the vector at the node is 50%,

75%, 95% or any other percentage selected, of the maximum response produced by that vector at any other node, and at the same time no other input vector of another class has produced a response at that node higher than 50% of the maximum response. This criterion will often result in a less diagnostic yield but certainly more reliable, as the percentage selected for the maximum response increases. For example, as shown in Table 1, model 10 for confidence levels at 50%, 75% and 95% resulted in 85%, 77%, and 69% diagnostic yields respectively.

During the evaluation phase the vectors describing the evaluation set were presented at the input of the model, one after the other and the output node producing maximum response for each subject was recorded. The diagnostic yield was then determined as the percentage of the correctly mapped subjects.

The sets of Figures 3,4,5 and 6,7,8 show the maps which have been obtained by models 1 and 10 respectively, after the completion of the training and evaluation phases for 50%, 75%, and 95% confidence levels for sensitivity analysis. The first number of each node indicates the class which has been assigned to the node after training. The second and third numbers indicate the class and the subject numbers respectively. (Ex. Node 4 4 2 indicate that the node is SMA, and that patient 2 in the SMA group has caused maximum response.) Nodes with an asterisk are nodes which have accommodated subjects from the evaluation set. It should be noted that during the evaluation process, one node might accommodate more than one subject. In any case, the total number of successfully diagnosed subjects are recorded by the system for calculating the diagnostic yield.

## Discussion

The results of this study, shown in Table 1, suggest that the unsupervised learning neural networks have generally produced better results than those which have been produced by the supervised learning neural networks of an earlier study [3].

Eight input models, i.e., those which include the standard deviation of the parameters in addition to the mean of the parameters, have produced better results than the four input models, which suggest that the standard deviation is contributing significantly to the learning process and classification.

No conclusion can be reached from the results, about the optimum size of the output grid, since the examined models for the 10x10 and 8x8 cases have produced similar results. It is expected, however, that an optimum grid size should exist, which will depend on the size and the variability of the training set.

More epochs can improve the performance of a model up to a certain level, beyond which the number of epochs will have no positive effect. Model 2, for example, after 100 epochs, reached a 77% diagnostic yield, at 50% confidence level. Increasing the number of epochs from 100 to 200 for this model, as illustrated by Model 3, increases the diagnostic yield to the 85% level. Increasing the number of epochs further did not improve the diagnostic yield.

By applying different gain factors (0.90 and 0.09), it was observed that the models with a 0.90 gain factor outperformed those with a 0.09 gain factor, in spite of the increase in the number of epochs for the latter.

Finally, when the SMA group was excluded from the system, the overall diagnostic yield improved, thus illustrating the negative effects due to the heterogeneous nature of the disease. The use of unsupervised learning in this study approaches the issue of separability which has

been an important aspect of clinical science for a long time. Separability is implied in the clinical processes of "lumping and splitting." There are a number of neuromuscular disorders which have a similar clinical appearance but different underlying pathologies. In general, the expectation is that coherent differences between pathological groups might be evident as clustered associations of findings. In practice this is not always the case. Pathology is not a discrete process, but forms a continuum from normality to disease. There is, however, some assistance with this problem by the fact that at the time of presentation of a patient, the degree of pathology is significantly different from normal. Given this, a patient group could exist as a separable group. Nonetheless, the inverse should also be considered. By the time of presentation, so much damage has been done to the muscles, that no effective treatment is possible. Therefore, sensitivity of diagnostic procedures is of paramount importance even if the specificity of a single test is low.

## References

- [1] Fawcett PRW, Dick DJ, Schofield IS, (1985). SFEMG and Macro EMG in Spinal Muscular Atrophy, EEG and Clin. Neurophysiol., 61, S76.
- [2] Fawcett PRW, Dick DJ, Schofield IS, (1986). SFEMG and Macro EMG in Becker Muscular Dystrophy, EEG and Clin. Neurophysiol., 63, 33P.
- [3] Schizas CN, Pattichis CS, Schofield IS, Fawcett PR, Middleton LT, (1990). Artificial Neural Nets in Computer-Aided Macro Motor Unit Potential Classification, IEEE Engineering in Medicine and Biology Magazine, Vol 9, 3, pp 31-38.
- [4] Schizas CN, Pattichis CS, Schofield IS, Fawcett PR, Middleton LT, (1989). Artificial Neural Net Algorithms in Classifying Electromyographic Signals, The First IEE International Conference on Artificial Neural Networks, IEE Pub. No. 313, pp 134-138.
- [5] Stalberg E, (1980). Macro EMG, a New Recording Technique, J. Neurol. Neurosurg. Psychiat., 43, 475-482.
- [6] Pattichis C, (1991). A Morphological Analysis of the Macro MUP, MSc Thesis, University of Newcastle Upon Tyne, Newcastle Upon Tyne, UK.
- [7] Stalberg E, Fawcett PRW, (1982). Macro EMG in Healthy Subjects of Different Ages, J. Neurol. Neurosurg. Psychiat., 45, 870-878.
- [8] Nandedkar S, Stalberg E, (1983). Simulation of Macro EMG Motor Unit Potentials, EEG Clin. Neurophysiol, 56, pp 52-62.
- [9] Kohonen T. (1984). Self-Organization and Associative Memory, Springer-Verlag, Berlin.
- [10] Lippmann RP, (1987). An Introduction to Computing with Neural Nets, IEEE ASSP Magazine, 4,2, pp 4-22.

## Acknowledgements

The authors would like to thank Dr C. A. Bonsett MD, Director of the Muscular Dystrophy Clinic - Riley Hospital, Indiana University for reviewing this paper, and Mr L. S. Walters, Technical Assistant.

Model of Classes	Number of Inputs	Output Grid	Gain Factor	Number of Epochs	Diagnostic Yield	Confidence Level
1	4	10x10	0.90	100	85	85
2	4	10x10	0.09	100	77	69
3	4	10x10	0.09	200	85	77
4	4	8x8	0.90	100	85	85
5	4	8x8	0.09	100	77	69
6	4	8x8	0.09	150	85	69
7	4	10x10	0.90	100	77	77
8	4	10x10	0.09	100	69	61
9	4	10x10	0.09	150	77	69
10	4	8x8	0.90	100	85	69
11	4	8x8	0.09	100	77	69
12	3	10x10	0.90	100	91	64
13	3	10x10	0.09	100	91	82
14	3	8x8	0.90	100	91	73
15	3	8x8	0.09	100	82	82
16	3	10x10	0.90	100	82	73
17	3	10x10	0.09	100	91	82
18	3	8x8	0.90	100	91	82
19	3	8x8	0.90	100	91	82
20	3	4	8x8	0.09	150	82

Table 1: Diagnostic Yield for the Evaluation Set at 50%, 75% and 95% Confidence Levels

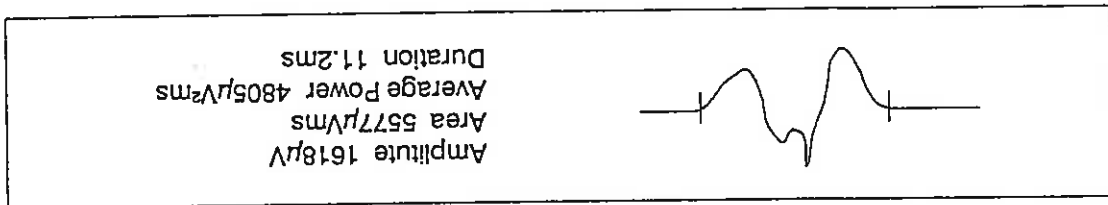


Figure 1: A Marco Motor Unit Potential (MMUP)

4.2	0.0	2.6	0.0	0.0	0.0	0.0	3.13	0.0	3.9	3.7
0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.11	0.0	1.1	3.2
0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.14	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.4	0.0	0.0	0.0
2.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.6
0.0	0.0	2.2	0.0	0.0	0.0	1.6	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.2
0.0	0.0	0.0	4.6	0.0	0.0	2.7	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.7

Figure 2: Maximum Response Map for the Training Set

Legend:  
1.5 Normal  
2.6 MND  
3.9 BMD  
4.2 SMA  
0.0 Zero Node

Legend: 1) 4 2 Means Class 4 (SMA) Node,  
SMA Patient Number 2  
2) 2 2 4 \* Means that the MND Patient  
Number 4 was Correctly Diagnosed  
1) 3 3 11 Means Class 3 (BMD) Node,  
BMD Patient Number 11  
2) 4 1 7 \* Means that the Normal,  
Number 7 was Diagnosed as SMA

444	442	228	226	000	441	3312	313	338	3311
442	442	224	228	118	441	441	000	111	332
000	000	000	229	115	341	334	441	3314	111
229	228	221	222	115	228	228	114	3314	112
228	228	222	227	112	112	115	114	336	336
228	228	222	222	118	118	118	112	115	117
228	444	446	446	000	000	118	118	112	112
448	448	448	448	446	446	227	227	118	112
225	448	448	448	227	227	118	447	447	447
225	225	223	223	223	227	000	447	447	447

Figure 3: Evaluation Map for Model 1, 50% Confidence Level

444	442	228	226	000	441	3312	313	338	3311
442	442	224	228	118	441	441	000	111	332
000	000	000	229	115	341	334	441	3314	111
229	228	221	222	115	228	228	114	3314	112
228	228	222	227	112	112	115	114	336	336
228	228	222	222	118	118	118	112	115	117
228	444	446	446	000	000	118	118	112	112
448	448	448	448	446	446	227	227	118	112
225	448	448	448	227	227	118	447	447	447
225	225	223	223	223	227	000	447	447	447

Figure 4: Evaluation Map for Model 1, 75% Confidence Level

444	442	228	226	000	441	3312	313	338	3311
442	442	224	228	118	441	441	000	111	332
000	000	000	229	115	341	334	441	3314	111
229	228	221	222	115	228	228	114	3314	112
228	228	222	227	112	112	115	114	336	336
228	228	222	222	118	118	118	112	115	117
228	444	446	446	000	000	118	118	112	112
448	448	448	448	446	446	227	227	118	112
225	448	448	448	227	227	118	447	447	447
225	225	223	223	223	227	000	447	447	447

Figure 5: Evaluation Map for Model 1, 95% Confidence Level

3311	339	338	441	3312	441	228	228
338	3314	338	441	441	441	228	228
3314	338	338	447	447	447	228	115
112	112	112	112	447	417	228	115
227	118	118	118	118	225	227	225
227	227	223	225	228	221	228	228
223	223	448	225	228	222	226	226
223	225	448	225	228	222	224	444

Figure 6: Evaluation Map for Model 10, 50% Confidence Level

3311	339	338	441	3312	441	228	228
338	3314	338	441	441	441	228	228
3314	338	338	447	447	447	228	115
112	112	112	112	447	417	228	115
227	118	118	118	118	225	227	225
227	227	223	225	228	221	000	226
223	223	448	225	228	222	000	228
223	225	448	225	228	222	024	444

Figure 7: Evaluation Map for Model 10, 75% Confidence Level

3311	339	338	441	3312	441	228	228
338	3314	338	441	441	441	228	228
3314	338	338	447	447	447	228	115
112	112	112	112	447	417	228	115
227	118	118	118	118	225	000	000
227	227	000	225	000	221	000	000
000	000	118	118	000	000	000	000
000	223	448	225	228	222	000	226
223	225	448	225	228	222	024	444

Figure 8: Evaluation Map for Model 10, 95% Confidence Level