NEURAL AND TRADITIONAL TECHNIQUES IN DIAGNOSTIC ECG CLASSIFICATION.

Rosaria Silipo, Giovanni Bortolan †

Dip. Sistemi e Informatica, University of Florence, Italy.
†LADSEB-CNR, Italy.
rosy@hp700.ing.unifi.it

ABSTRACT

Neural and traditional techniques have been compared for the particular task of automatic ECG analysis. A large validated ECG database has been used. Statistical methods, neural architectures with supervised and unsupervised learning, and a neuro-fuzzy architecture have been considered. The results from the connectionist approach are always at least comparable with those coming from more traditional classification methods. But the best performances have been obtained by the combination of the connectionist with the fuzzy approach.

1. INTRODUCTION

The medical knowledge is usually not very deterministically structured, since the same signs can point out very different pathologies and a slight change in the data can mean a serious change in the patient status. In addition, the available patient data are often not complete.

Among all the biomedical techniques, the electrocardiogram (ECG) signal analysis remains a very helpful diagnostic tool in many circumstances, since it still represents the most economic and the least invasive test for assessing cardiac diagnosis.

In the past years, several ECG analyzers — based on statistical methods [1], clustering methods [2], expert systems [3] and Markov models [4] — have been developed and implemented to solve the problem of the time consuming ECG analysis. Their not sufficient reliability, together with their high sensitivity to noise and their failure to deal with new or ambiguous patterns, leads the research towards investigation of new analysis techniques.

Artificial Neural Networks (NN) have been often proposed as good classifiers when non linear separation borders are required and incomplete or ambiguous input patterns can be found. In the last years the connectionist approach has been applied to the ECG analysis with promising results [5, 6, 7].

A crucial point in the design of any ECG analyzer is the clinical validation. In order to be meaningful, the test of the new system has to be performed on a database sufficiently representative, annotated and clinically validated.

In this study a systematic comparison among several neural paradigms and architectures and between neural and statistical approach has been performed on a large validated database, in order to investigate the behaviour and the features of each approach in the particular problem of the short term ECG analysis.

2. THE ECG DATABASE

In order to have a significant validation procedure, a large database developed at the Medical Department of the University of Leuven and already tested with other classical classification methods has been used [1, 5, 8, 9].

Seven diagnostic classes have been taken into account: *normal*, Left (LVH), Right (RVH), and Bilateral Ventricular Hypertrophy (BVH), Inferior (IMI), Anterior (AMI), and Mixed Myocardial Infarction (MIX).

The database is composed of 12 lead rest ECG records from 3266 patients, 2158 males and 1108 females. A random set of 2446 patients, has been selected for the learning phase, and the remaining 820 cases have been used for the testing.

Each ECG signal is characterized by 540 ECG primary measurements (45 x 12 leads). The most significant 39 parameters have been chosen to describe each record, through a clinical and statistical selection.

Given the generality of the database, its clinical validation, and the signifiance of the chosen parameters, a meaningful comparison among different classification techniques can be obtained.

The performance of each system is characterized by means of the average sensitivity and specificity, calculated over all the diagnostic classes. $sensitivity(i)$ indicates the rate of true positive events for class i. $specificity(i)$ measures the rate of true negative events for class i. Let:

N_C = number of diagnostic classes;

$E_x(i)$= number of events X referred to the class i by the process x, with x being:

- $A = X$ is assigned to class i;
- $B = X$ belongs to class i;

then:

$$average\ sensitivity = \frac{1}{N_C}\sum_{i=1}^{N_C}\frac{E_{AB}(i)}{E_B(i)} \qquad (1)$$

$$average\ specificity = \frac{1}{N_C}\sum_{i=1}^{N_C}\frac{E_{\overline{AB}}(i)}{E_{\overline{B}}(i)} \qquad (2)$$

The reference traditional algorithms were assessed from statistical procedures and applied to the short term ECG

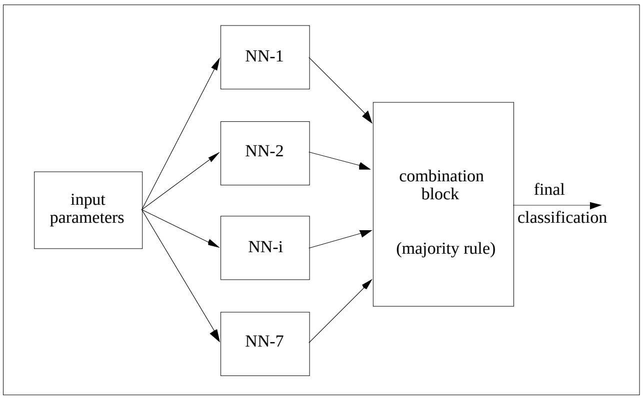

Figure 1. Combined Neural Networks (cNN) for ECG analysis.

analysis [1, 8]. They referred to a LOGistic discriminant analysis, reported in the result table as LOG, and to a Linear Discriminant Analysis, reported as LDA, and evaluated on the same ECG database.

3. THE NEURAL ECG CLASSIFIERS

3.1. Supervised neural architectures

The basic connectionist structure has been assessed by a feed forward NN, having 39 inputs, 7 hidden nodes, and 5 output nodes (*sNN*), and trained with Back Propagation (BP) algorithm.

This architecture copes with multiple diagnosis: BVH is considered as the combination of LVH and RVH, and MIX as the combination of AMI and IMI. In this way a more compact structure is obtained and ambiguous diagnosis with overlapping of single and composite diagnostic classes are avoided [9].

Possible groups of similar patterns with homogeneous features have been examined, for the additional advantage of isolating outliers. For this purpose a cluster analysis based on Euclidean distance in the input parameter space has been applied to the original training set. Three clusters have been obtained containing respectively 601, 1209, and 636 ECG records, and each one of them has been used for the training of a sNN.

A set of sNN is considered simultaneously, the obtained output classes are combined together and the final classification is produced. Several combination rules have been evaluated and the majority rule has been adopted. The combined classes, BVH and MIX, are still managed as the combination of RVH and LVH, and of AMI and IMI respectively. The resulting architecture *cNN* is reported in Fig.1 [9].

3.2. Neuro-fuzzy approach

Because of the crucial role played by the pre-processing of the input space in the connectionist approach, a neuro-fuzzy architecture was designed, by combining the linguistic description of the fuzzy approach with the classification performed by a neural classifier.

Each input parameter has been represented by means of a set of membership-functions, related with the output classes. A neural structure feed-forward, fully connected, with sigmoidal units and 7 output nodes has been analyzing the produced linguistic description of the input space.

The gaussian shape functions, or *Normalized Radial Basis Functions (NRBF)* , have been chosen as membership-functions, because of their mathematical properties of continuity and differentiability [11].

In order to obtain a more dynamical architecture, the NRBF parameters have been determined in an automatic way and not from the "expert knowledge", by developing the NRBF layer and the connectionist approach in the same framework. For this purpose the Back-Propagation algorithm has been modified to allow the training of the matrix weight and of the NRBF parameters as well.

Several neuro-fuzzy structures have been evaluated, varying the number of NRBF units (n_{NRBF}) from 1 to 7, where 7 is the number of diagnostic classes considered (Fig. 2).

The first choice, $n_{NRBF} = 1$, fuzzy describes each input parameter by means of only one NRBF unit (*1-NRBF*). The critical point of this approach is the selection of the appropriate NRBF shape. In this case the statistics of the normal diagnostic class from the training set has been considered for the computation of the initial μ and σ for each input parameter. This choice is supported by the fact that the pathological classes can be described as a strong or light deviation of the input parameters from the normal class.

An other fuzzy description of the input space has been obtained by defining the NRBF with respect to all the 7 diagnostic classes (*7-NRBF*). The initial values of mean and standard deviation of the NRBF are here derived by the statistics of each class in the training set.

The choice of considering 7 NRBF unit for each input parameter has the effect of increasing consistently the dimension of the pre-processed input space by a factor 7, which can be prohibitive from a computational point of view. Thus the problem of the optimal choice of the "most characteristic" NRBF has been studied. The Euclidean distance has been calculated among the 7 NRBFs, considering the two-dimensional space $\{\mu, \sigma\}$. The p NRBFs with maximum distance are selected to pre-process each input parameter. These can be considered as the most representative initial values of the NRBFs. The optimal number of NRBFs was found to be $p = 2$ [12].

To perform the classification task, in all the described neuro-fuzzy architectures the max rule is followed in the output layer.

Given the high dimensions of the designed systems, a pruning technique was applied, in order to reduce the system size and to improve its generalization capability. Two pruning techniques have been evaluated.

The first one, applied to the 7-NRBF structure, added a penalty term to the cost function [11], in order to lead the network to configurations where weights assume low values.

The second one, applied to the p-NRBF, during the learning process defined a sensitivity parameter for each weight and node of the system, in such a way as to describe their contribution to the learning procedure. At the end of the learning process, weights and nodes with lowest sensivities are pruned out of the network. Such technique was more effective, though a bigger computational effort was required.

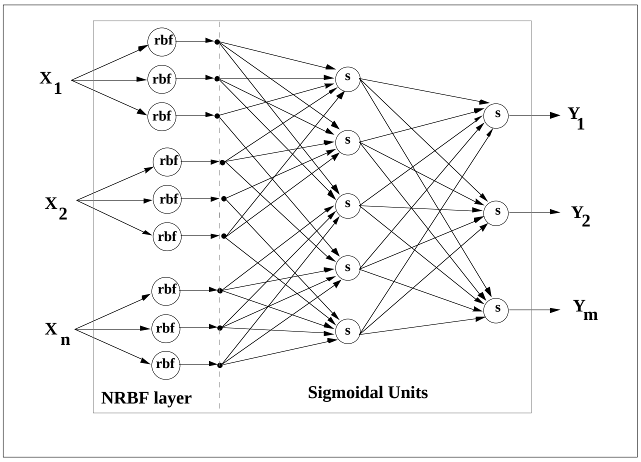

Figure 2. NRBF architecture

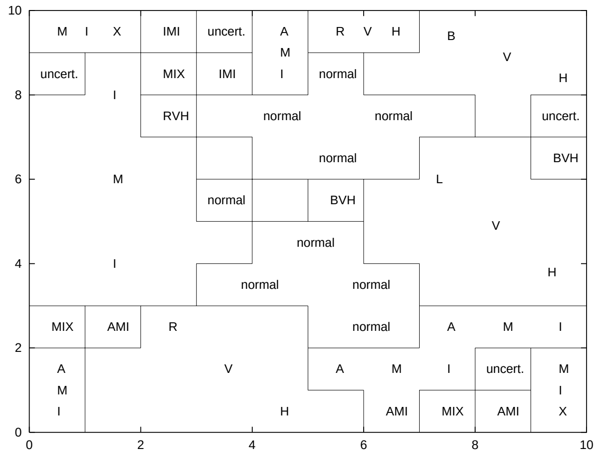

Figure 3. 10x10 Kohonen map.

3.3. Self Organizing Maps

Kohonen maps define a mapping from the input data into a regular two-dimensional array of nodes, with an unsupervised learning procedure.

In order to investigate how information is coded in different arrays, three squared self-organizing maps have been here implemented and evaluated with different sizes: 7x7, 10x10, and 28x28. Those dimensions are chosen, after taking into account the number of output classes and the dimension of the input space. An hexagonal lattice is chosen, since it allows an easier visual inspection [13].

The learning process has been split in two phases, a shorter one with a learning rate of 0.05 and a radius set to 10.0 and a second longer one with a lower learning rate (0.03) and a smaller radius (3.0). The first phase allowed the map to quickly reach a matrix status close to the best clustering of the input space, i.e. with a small quantization error [13], while the second phase performed a finer tuning of the found groups.

Since Kohonen maps are sensitive to the initial conditions, 10 maps have been trained for each dimension and the map with the lowest quantization error has been evaluated.

The most used classification strategy classifies each input pattern according to the winning unit label. To avoid a too rigid classification, we assigned the input pattern to all the output classes corresponding to units which output is lower than the minimum + 5%. More precisely:

IF: $min = \min_j {o_j}^M$
THEN: pattern M $\in \left\{class(h) : {o_h}^M \leq min + 5\%\right\}$

where ${o_h}^M$ represents the output of the unit h for the input pattern M, and the $class(h)$ the associated diagnostic class.

Every output class is represented by some topological regions in the map. If the dimension of the map is not too high, those regions are also quite compact and well defined. *"Wandering" units* in the maps, placed far away from the main region of the corresponding output class, represent clusters with ambiguous features.

As bigger the map size is as bigger the fragmentation of represented clusters and the number of uncertain and never winning units is.

Under-dimensioned maps, as the 7x7 map, force the creation of a very compact clustering, leaving small space for describing ambiguous data. Over-dimensioned maps, as the 28x28 map, produce units representing such particular cases of input data that the resulting map is not compact at all and the number of uncertain or never winning units becomes high. In fig. 3 the final mapping of the 10x10 map is reported.

Neverthless, the spatial organization of the output classes shows common features in all the evaluated maps. The main regions correspond to the two lateral hypertrophies, left and right, and to the two simple infarctions, anterior and inferior. The two hypertrophies are usually placed in two opposite sides and divided by the normal status and uncertain clusters, setting the borderline. The same happens with the two simple myocardial infarctions.

The normal status is then coded as a transition condition from where it is possible to move to different kinds of infarction or hypertrophy, according to the changes in the input features.

The two combined diagnostic classes, MIX and BVH, share the input features of the simple cardiac diseases from which originate. The maps represent such information by placing the combined classes in between or just close to the corresponding simple diseases. This way the BVH is represented by units placed close to LVH and RVH, and the MIX by units close to AMI and IMI.

4. COMPARISON AND DISCUSSION

In table 1 the performances of all the evaluated ECG classifiers are reported. The results show that the classifiers based on NN produce performances at least comparable with those from traditional classifiers.

In addition the combination of a fuzzy technique and of the connectionist approach allows a sensitive improvement. In particular the main one is obtained when the number p of RBF units moves from 1 to 2. Further improvements

Table 1. Performance comparison

method	aver. sensit.	aver. specif.
LDA	63%	-%
LOG	63%	94%
sNN	64%	94%
cNN	66%	95%
1-NRBF	59%	94%
7-NRBF	67%	94%
p-NRBF(p=2)	63%	94%
K-7x7	50%	92%
K-10x10	52%	92%
K-28x28	48%	91%

derived from a higher number of RBF units come together with an increased complexity of the architecture.

The compact representation of the input space into a two-dimensional array with reduced size (K-7x7) does not allow a fine clustering of the input data, leading to low percentages of average sensitivity and specificity. On the other hand the bad performance of the biggest map (K-28x28) shows that a too detailed description does not help for the final classification as well. Thus the optimal size of the Kohonen map is a parameter to be carefully investigated.

The lower performances produced by the Kohonen maps in comparison with those of the systems trained with a supervised technique (Tab. 1) do not necessarely mean a worse classification process. On one side the supervised learning prevents the network from "wrong" classifications, by ignoring ambiguous clusters or contraddictory patterns. On the other side, right because of their unsupervised learning, the self-organizing maps become able to catch similarities among input patterns (fig. 3). In that sense, they can produce a more accurate and detailed description of the input space, better than just using pre-defined diagnostic classes.

5. CONCLUSIONS

In this work the statistical and the neural approach have been implemented in performing the automatic analysis of the diagnostic ECG. Techniques with unsupervised learning, neuro-fuzzy architectures, and self-organizing maps have been considered and compared with traditional ECG analyzers on the basis of the same ECG database [1, 8].

A large validated database of ECG signals has been used for this study. The availability of a significative, large, and clinically validated ECG database guarrantees general and meaningful evaluation and comparison processes.

The results obtained by the connectionist approach were always at least comparable with those obtained by more traditional classification methods (Tab. 1), though the design of a neural classifier did not involve the definition of statistical properties and did not require any a priori hypothesis on the input population.

Architectures trained with unsupervised techniques produced worse classification performance, from the point of view of some pre-defined diagnostic classes, but a finer clustering of the input patterns was produced, based on their similarities.

Finally, the best performances were obtained by the combination of the connectionist approach with the fuzzy technique, since that allowed to yield advantages from both procedures.

REFERENCES

[1] J.L. Willems, E. Lesaffre, "Comparison of multigroup logisitic and linear discriminant ECG and VCG classification.", *J. of Electrocardiology* , Vol. 20, pp. 83-92, 1987.

[2] J.L. Talmon, "Pattern recognition of the ECG", Akademisch proefscrift, 1983.

[3] A. Gallino, *et al.*, "A Computer System for Analysis of ST Segment Changes on 24 Hours Holter Monitortapes: Comparison with other available Systems", *Journ. of American College of Cardiology*, Vol.4, pp. 245-252, 1984.

[4] A.D. Coast, R.M. Stern, G.G. Cano,S.A. Briller, "An Approach to Cardiac Arrhythmia Analysis using Hidden Markov Models", *IEEE Trans. on Biomedical Engineering*, Vol. 37, n. 9, pp. 826-835, 1990.

[5] G. Bortolan and J.L. Willems, "Diagnostic ECG Classification Based on Neural Networks", Journal of Electrocardiology, Vol. 26, pp. 75-79, 1994.

[6] R. Silipo, M. Gori, A. Taddei, M. Varanini, C. Marchesi, "Classification of Arrhythmic Events in Ambulatory ECG, using Artificial Neural Networks"", *Computers and Biomedical Research*, Vol. 28, pp. 305-318, 1995.

[7] J.H. Frenster, "Neural Networks for Pattern Recognition in Medical Diagnosis." In: Biomedical Engineering Perspectives: health care technologies for 1990's and beyond, Proc. 12th IEEE EMBS, P.C. Pedersen and B. Onaral (Eds), pp. 1423-1424, 1990.

[8] J.L. Willems, E. Lesaffre, and J. Pradens,"Comparison of Classification Ability of the Electrocardiogram and Vectorcardiogram.", *Am. Journal of Cardiology* , Vol. 59, pp. 119-124, 1987.

[9] G. Bortolan, R. Degani, J. Willems, "ECG classification with neural networks and cluster analysis", In: Computers in Cardiology 1991, A. Murray, R. Arzbaecher (Eds.), IEEE Computer Society, Los Alamitos, CA, pp. 177-180, 1991.

[10] G. Bortolan, R. Silipo, C. Marchesi, "Fuzzy pattern classification and the connectionist approach.", *Pattern recognition letter*, Vol. 17, pp. 661-670, 1996.

[11] G. Bortolan, S. Fusaro, "Some aspects of fuzzy preprocessing in the connectionist approach.", In: Fourth European Congress on Intelligent techniques and soft computing EUFIT-96, Verlag Mainz, Aachen, pp. 1425-1429 (1996).

[12] T. Kohonen, "Self organization and Associative memory." Springer-Verlag, 1989.