Validation of an Algorithm for Atrial Fibrillation and Flutter Diagnosis

BF Giraldo[1], C Ferré[1], P Laguna[2], R Jané[1], M Martín[3], P Caminal[1]

[1] Dept. ESAII, Centre de Recerca en Enginyeria Biomèdica, UPC. Barcelona, Spain
[2] Dept. Ingeniería Electrónica y Comunicaciones, Centro Politécnico Superior. Univ. Zaragoza, Spain.
[3] Unitat d'Epidemiologia Clínica. Consorci Hospital Creu Roja. L'Hospitalet de Llobregat, Spain

Abstract

In this work, we developed a methodology to evaluate an algorithm (AAFF) for atrial fibrillation (AF) and flutter (AFL) detection and discrimination. Atrial electrical activity in pathologies such as AF and AFL are difficult to characterize quantitatively. The diagnoses provided by AAFF were compared to those established by 8 clinicians at different expertise levels in cardiology, and the MIT-BIH database annotations. The concordance between diagnoses supplied by a pair of experts was studied with the measure of distances between their diagnoses. The methods used to compute distances were: Euclidean distance, Mahalanobis distance and City-block distance. Using the resulting matrices of distances between experts, cluster analyses were carried out to classify AAFF among human experts. The results show that differences between diagnoses made by AAFF and those made by some specialists were smaller than differences between some specialists themselves.

1. Introduction

In the diagnostic of atrial fibrillation (AF) and flutter (AFL), the F wave determines the patterns of fibrillation/flutter [1,2]. In a previous work the authors developed an automatic detection algorithm (AAFF) [3], based on the measures of amplitudes and durations of F waves (figure 1, 2). The method uses the differentiated and low-pass filtered ECG signal for the detection of wave boundaries [4]. For the diagnosis of the AF or AFL, the algorithm analyses the regularity of F wave patterns with the variation coefficient of their amplitudes and durations.

In this work we present the validation of AAFF. The validation of any automatic diagnostic system is a fundamental step in its development. The special characteristic of this validation lies in the specific methodology used to deal with the problem of not having a gold standard available as a reference. This methodology is based in the analysis of concordance between diagnoses analyzed by means of the distances between experts, and the cluster analysis [5,6]. The validation of the algorithm for the diagnostic of AF and AFL was performed using 60 segments of ECG records, each one with a duration of 4 seconds, from MIT-BIH database [7]. These segments contain episodes annotated as AF and AFL. A sample of 60 ECG records was therefore available for analysis. The diagnoses provided by AAFF were also compared to those established by a group of clinicians at different expertise levels in cardiology.

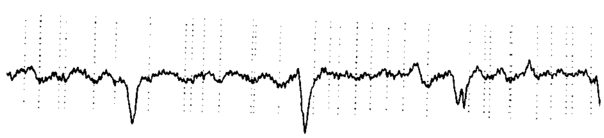

Figure 1. Example of F waves detection in record 210_18 :19, channel 1 (V1), MIT-BIH. Annotated Atrial Fibrillation.

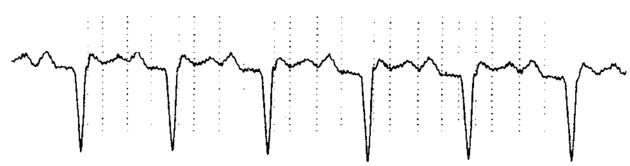

Figure 2. Example of F waves detection in record 202_25 :43, channel 1 (V1), MIT-BIH. Annotated Atrial Flutter.

2. Materials and Methods

The AAFF gave a diagnosis for each ECG record. Each diagnosis took the form of an array with 3 fields: atrial fibrillation, atrial flutter and other. These fields were qualified with 1 or 0, depending if the diagnostic is positive or not. Each one of the 8 clinicians involved in the process of validation received the same 60 ECG records plotted on standard paper. The group of clinicians include 6 specialists in cardiology (E1 to E6) and 2 non specialists in cardiology (E7 and E8). They were required to formulate a diagnosis for each case. Diagnosis results for all the clinicians and automatic algorithm were collected in order to obtain a cubic matrix, the three dimensions of which were the 10 experts (8 clinicians (E1 to E8), the MIT-BIH annotation (MB), plus the automatic diagnosis AAFF (AD)), the 60 ECG records selected, and the 3 possibilities associated with the pathologies.

The methods considered to compute distances between experts in each ECG record are:

- Euclidean distance, that is defined as

$$d(i,j)=\sqrt{\frac{1}{N}\sum_{m=1}^{N}(x_{im}-x_{jm})^2}$$

where d is the distance, i and j are a pair of experts, $N=3$ is the number of pathologies, m is the considered pathology and x is the numerically expressed diagnosis (1 or 0).

- Mahalanobis distance, that is a generalization of Euclidean distance useful in the case when the possibilities of the different diseases are not independent.

$$d(i,j)=\sqrt{\frac{1}{N}(x_i-x_j)' * w^{-1} * (x_i-x_j)}$$

where $(x_i - x_j)$ is the 3-dimensional column vector of differences between the diagnosis arrays i and j, is the $(x_i - x_j)'$, is the corresponding transposed vector, and the w^{-1} inverse of the variance-covariance matrix of the pathology possibilities.

- City-block distance, that penalizes the existence of many small differences.

$$d(i,j)=\frac{1}{N}\sum_{m=1}^{N}|x_{im}-x_{jm}|$$

Distances between each pair of experts obtained for the 60 cases were averaged to establish a 10 by 10 matrix of dissimilarities (Table 1). Based on this symmetrical matrix, a hierarchical cluster analysis [8] was carried out to get a progressive agglomeration of the experts in function of their similarity. When two experts are clustered, another matrix of dimension $(n-1)*(n-1)$ must be built for the process to continue. Every pair of similarity or dissimilarity coefficients of the two clustered experts to the rest of them is then substituted by a single new value. There are several methods for obtaining this new value. In the methods presented in this work, we use the average linkage criteria, which define the new values as an arithmetic mean of the previous ones. Both an advantage and disadvantage of the hierarchical cluster analysis is that the method does not provide a fixed number of cluster but rather a progressive agglomeration of the experts. The result of a hierarchical classification was graphically represented by a dendrogram depicting the nested structure of the clusters. This type of figure presents the experts or groups of experts as connected by links, and each node of the dendrogram represents a cluster [9]. The similarity of experts or groups of experts can easily be seen in the dendrogram. Cluster analysis is a descriptive analysis and does not supply information about the statistical significance of the observed differences between experts, but it is useful enough for the present study because it provides a relative classification of AAFF among human experts and the MIT-BIH annotation.

Table 1. Matrix of dissimilarities calculated using the City-block distance

	MB	AD	E1	E2	E3	E4	E5	E6	E7	E8
MB	0.000									
AD	0.255	0.000								
E1	0.255	0.122	0.000							
E2	0.305	0.283	0.238	0.000						
E3	0.244	0.277	0.166	0.272	0.000					
E4	0.233	0.266	0.200	0.227	0.144	0.000				
E5	0.288	0.300	0.266	0.294	0.200	0.122	0.000			
E6	0.222	0.300	0.211	0.205	0.211	0.144	0.233	0.000		
E7	0.233	0.311	0.322	0.327	0.300	0.355	0.388	0.311	0.000	
E8	0.238	0.305	0.316	0.300	0.261	0.327	0.361	0.283	0.094	0.000

3. Results

The results are presented from different points of view: frequency of diagnoses, distance between experts, singular diagnoses and omitted diagnoses.

3.1. Frequency of diagnoses

A descriptive statistical analysis for each expert was carried out by determining the number of the pathologies diagnosed (Table 2). When some specialists provided more than one pathology for each ECG record, the total number of diagnosed pathologies was greater than the number of records considered. This is the case of clinicians E1, E2, E6, E7 and E8. The last three clinicians also diagnosed *other* pathologies.

Table 2. Frequency of the pathologies diagnosed by each expert.

	MB	AD	E1	E2	E3	E4	E5	E6	E7	E8
Atrial Fibrillation	31	33	29	37	27	35	46	33	22	22
Atrial Flutter	29	27	34	26	33	25	14	40	39	40
Other	0	0	0	0	0	0	0	2	3	1
Total	60	60	61	63	60	60	60	75	64	63

3.2. Distance between experts

Euclidean distance.- In Figure 3 we represent graphically as a dendrogram the results of a cluster analysis resulting from the matrix of dissimilarities between experts calculated using the Euclidean distance. As it can be observed in this dendrogram, the distance between specialists E7 and E8 was the smallest. The distances between E4 and E5, and also between E1 and AD were nearly of the same value. When considering the Euclidean distance two big clusters may be considered. One includes the automatic diagnosis AAFF (AD) and the clinicians E1 to E6. The other cluster includes the MIT-BIH annotation (MB) and the clinicians E7 and E8.

Mahalanobis distance.- Figure 4 shows the dendrogram of the cluster analysis resulting from matrix of Mahalanobis distances. In this dendrogram the distance between experts E4 and E5 was the smallest, followed by the distance between experts E1 and AD. Considering the Mahalanobis distance, the diagnoses of AAFF (AD) are closer to the MIT-BIH annotation (MB) that when using other distances. Small distances are obtained in the cluster formed by AD, MB, E1, E3, E4 and E5.

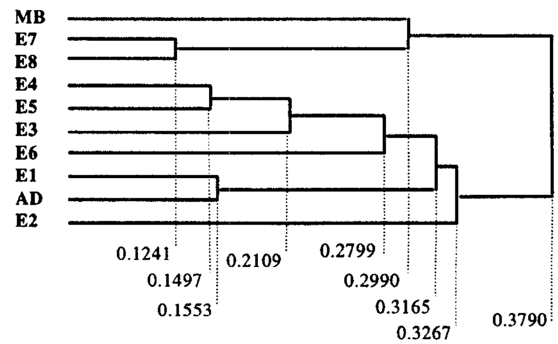

Figure 3. Dendrogram of cluster analysis from Euclidean distances.

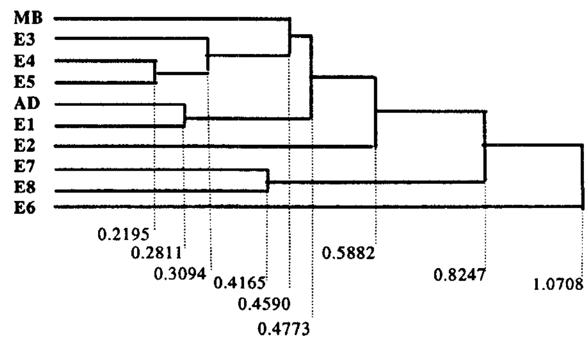

Figure 4. Dendrogram of cluster analysis from Mahalanobis distances.

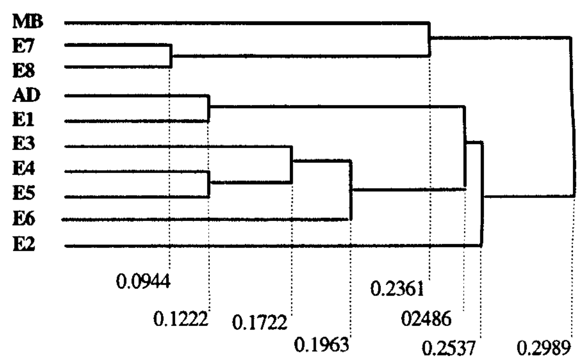

Figure 5. Dendrogram of cluster analysis from City-block distances.

City-block distance.- Table 1 shows the 10 by 10 dissimilarity matrix obtained using the City-block distance once all cases were averaged. In Figure 5 we represent graphically as a dendrogram the results of cluster analysis using this distance. This dendrogram is similar to the one obtained when using the Euclidean distance. The diagnosis of AD and E1 are very near, and belong to the same group than E2 to E6.

3.3 Singular diagnoses

We considered singular diagnosis those mentioned only by one expert. One way of determining the concordance of a single expert with the consensus of the group is to note the occurrence of singular diagnoses. Table 3 shows the percentage of singular diagnoses recorded by each expert. The automatic diagnosis AAFF did not record singular diagnosis, while clinicians E2, E5, E6, E7 and E8 recorded different cases.

Table3. Percentage of singular diagnoses recorded by each expert over the 60 ECG records.

Expert	MB	AD	E1	E2	E3	E4	E5	E6	E7	E8
%	0	0	0	10.0	0	0	3.3	3.3	6.6	1.6

3.4 Omitted diagnoses

An omitted diagnosis was considered as a diagnosis mentioned by every expert but one. The number of omitted diagnoses is a measure of disagreement with the consensus. Table 4 shows the percentage of diagnoses omitted by each expert. The automatic diagnosis AAFF did not omit any diagnosis, while clinicians E2, E5, E7 and E8 made different omissions.

Table 4.. Percentage of diagnoses omitted by each expert over the 60 ECG records.

Expert	MB	AD	E1	E2	E3	E4	E5	E6	E7	E8
%	0	0	0	8.3	0	0	3.3	0	1.6	1.6

4. Discussion and Conclusions

We have developed a methodology to evaluate an algorithm (AAFF) for atrial fibrillation and flutter detection and discrimination. The special characteristic of the validation methodology lies in the fact that it evaluates an automatic diagnostic system in a field where no absolute gold standard exists as a reference.

The results obtained applying this approach to AAFF validation show that differences between diagnoses made by AAFF and those made by some specialists were smaller than differences between some specialists themselves.

Specialists with highest proficiency scores gave the closed diagnoses. Results with three different distances showed that AAFF was closer to the clinicians with high expertise level in cardiology (E1 to E6) than to the other clinicians (E7 and E8). The best specialists and AAFF gave the least singular diagnoses and the least omitted diagnoses.

The results obtained with the Mahalanobis distance could be more reliable than the obtained with the other distances, because it takes into account that the different pathologies are not independent. In this case the MIT-BIH annotation belongs to same big cluster than the diagnoses of clinicians with high expertise level.

Acknowledgments

We acknowledge Drs. R. Baranowski, J.A. Casanovas, C. Castaño, M.C. Gallardo, E. Homs, I. Lacambra, A. Sánchez, P.J. Serrano, for their contribution to the validation of the algorithm with their annotations in the ECG records.

This work was supported in part by grant TIC94-0608-C02, from CICYT, Spain.

References

[1] H.Aberg, (1970). "Failure of Conversion of Atrial Fibrillation. Relation to Fibrillatory Wave Size". *Acta. Med. Scand.* 188, pp 197-199.

[2] J.Morganroth, L.N.Horowitz, M.E.Josephson, J.A.Kastor, (1979). "Relationship of atrial fibrillatory wave amplitude to left atrial size and etiology of heart disease". *Am. Heart J.* 97, pp. 247-251.

[3] B.F.Giraldo, P.Laguna, R.Jané, and P.Caminal, (1995). "Automatic Detection of Atrial Fibrillation and Flutter using the Differentiated ECG Signal". *Computers in Cardiology*, IEEE Computers Society Press, pp. 369-372.

[4] P.Laguna, R.Jané, P.Caminal, (1994). "Automatic detection of wave boundaries in multilead ECG signals: Validation with the CSE database". *Computers and Biomedical Research* 27, pp. 45-60.

[5] A.Verdaguer, A.Patak, J.J.Sancho, C.Sierra, F.Sanz, (1992). "Validation of the Medical Expert System PNEUMON-IA". *Computers and Biomedical Research* 25, pp. 511-526.

[6] C.Hernández, J.J.Sancho, M.A.Belmonte, C.Sierra, F.Sanz, (1994). "Validation of the Medical Expert System RENOIR". *Computers and Biomedical Research* 27, pp. 456-471.

[7] Harvard University, (1988). "MIT-BIH Arrhythmia Database Directory". *Massachusetts Institute of Technology. Division of Health Sciencies and Technology.*

[8] M.J.Norusis, (1985). "SPSSx Advanced Statistics Guide". *SPSS, Chicago.*

[9] G.J.McLachlan, (1992). "Cluster analysis and related techniques in medical research". *Stat. Methods Med. Research 1*, pp. 27-48.

Address for correspondence:
Beatriz F. Giraldo.
Dept.ESAII,Centre de Recerca en Enginyeria Biomèdica, UPC. Pau Gargallo 5, 08028, Barcelona, Spain.
e-mail: *giraldo@ic.upc.es*