

Evolutionary Algorithms for Multiobjective and Multimodal Optimization of Diagnostic Schemes

Francisco de Toro*, Eduardo Ros, Sonia Mota, and Julio Ortega

Abstract—This paper addresses the optimization of noninvasive diagnostic schemes using evolutionary algorithms in medical applications based on the interpretation of biosignals. A general diagnostic methodology using a set of definable characteristics extracted from the biosignal source followed by the specific diagnostic scheme is presented. In this framework, multiobjective evolutionary algorithms are used to meet not only classification accuracy but also other objectives of medical interest, which can be conflicting. Furthermore, the use of both multimodal and multiobjective evolutionary optimization algorithms provides the medical specialist with different alternatives for configuring the diagnostic scheme. Some application examples of this methodology are described in the diagnosis of a specific cardiac disorder—paroxysmal atrial fibrillation.

Index Terms—Evolutionary algorithms, multiobjective and multimodal optimization, noninvasive medical diagnosis, paroxysmal atrial fibrillation.

I. INTRODUCTION

MANY real-world optimization problems deal with a multimodal function with local and global *optima*. In this paper, multimodal optimization problems refer to those ones dealing not just with the retrieval of the global *optima*, but also the local *optima* of sufficient quality. There are two good, practical reasons which may prompt the location of multiple *optima* in such types of problems. First, by encouraging the location of multiple *optima*, the chances of situating the global *optimum* are increased. Second, in a design context, identifying a diverse set of high-quality solutions will provide an insight into the nature of the input space and suggest alternative solutions [1]. Herein, we focus on this second objective.

Sometimes optimization problems exist with several (and normally conflicting) objectives, which need to be accomplished at the same time. A multiobjective optimization problem (MOP) can be defined [2] as that of finding a *vector of decision variables* $x \in \Phi \subseteq \mathbb{R}^n$, which meets a series of constraints and optimizes a *vector of objective functions*

$$f(x) = [f_1(x), f_2(x), \dots, f_k(x)]^T \quad (1)$$

where the k elements represent the objectives. These functions, usually in conflict with each other, are a mathematical descrip-

Manuscript received March 16, 2004; revised April 17, 2005. This work was supported in part by the Spanish national grants DPI-2004-07032 and TIN2004-01419. *Asterisk indicates corresponding author.*

*F. de Toro is with the Department of Signal Theory, Telematics and Communications, E.T.S. Informática, c/ Periodista Daniel Saucedo s/n, 18071 Granada, Spain (e-mail: ftoro@ugr.es)

E. Ros, S. Mota, and J. Ortega are with the Department of Computer Architecture and Technology at University of Granada, 18071 Granada, Spain (e-mail: eduardo@atc.ugr.es; smota@atc.ugr.es; julio@atc.ugr.es).

Digital Object Identifier 10.1109/TBME.2005.862539

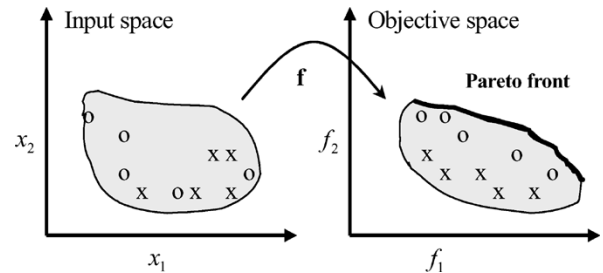


Fig. 1. Maximization of f_1 and f_2 : Some Pareto optimal solutions (o) and dominated solutions (x). After the optimization procedure, the Pareto optimal solutions (o) encountered will approach the Pareto front.

tion of performance criteria. The meaning of optimum is not well defined in this context, so it is difficult to have a vector of decision variables that optimizes all the objectives simultaneously. Therefore, the concept of *Pareto optimality* is used. Consider that all the components of (1) are to be maximized. Then, a solution $x^* \in \Phi \subseteq \mathbb{R}^n$ is defined as Pareto optimal if the condition expressed in (2) is satisfied

$$\forall x \in \Phi, \exists i \in \{1, \dots, k\} / f_i(x^*) > f_i(x) \wedge \forall j \neq i \in \{1, \dots, k\} f_j(x^*) \geq f_j(x). \quad (2)$$

This means that x^* is Pareto optimal, if no feasible vector x exists that would increase one criterion without causing a simultaneous decrease in, at least, one of the others. The notion of Pareto optimum almost always gives not just a single solution, but rather a set of solutions known as the Pareto front. Given that, in general, it is not easy to find an analytical expression for the Pareto front, the usual procedure is to determine a set of Pareto optimal solutions (see Fig. 1) that provides a good approximate description of the Pareto front. If $x \in \Phi$ is not a Pareto optimal point, then it is referred to as *dominated* solution (see Fig. 1).

Evolutionary algorithms (EAs) [3] are stochastic optimization procedures which apply a transformation process (crossover and mutation operators), inspired by the species natural selection process, to a set (population) of coded candidate solutions (individuals) to the problem. These procedures are especially suitable for solving multiobjective and multimodal optimization problems because they are able to capture multiple solutions in a single run. Typically, in order to avoid premature convergence, a diversity preservation technique has to be introduced in the design of the evolutionary algorithm. This becomes compulsory when EAs are used to locate different optima in multimodal or multiobjective optimization problems so that the chances of the algorithm converging into a single solution or a small group of solutions are minimized.

Multiobjective optimization evolutionary algorithms (MOEAs) is a very active area of research [4], [5]. A good summary of real-world optimization problems addressed with MOEAs can be found in [4]. Some researchers have suggested that multiobjective optimization might be an area where EAs perform better than other search strategies. The considerable amount of research which has been done on MOEAs contrasts, greatly, with the limited use made so far of EAs in locating the multiple *optima* of multimodal functions [1].

This paper addresses the use of evolutionary multimodal and multiobjective optimization techniques for the adjustment of medical diagnostic schemes dealing with biosignal analysis and processing. Both approaches give a set of different solutions for the diagnostic scheme providing medical specialists with a certain amount of flexibility to customize it. In particular, evolutionary multimodal optimization techniques are used in this work to obtain a set of different characteristics related to the physical properties of the biomedical signal on which the diagnosis can be based. This enables specialists to make a diagnosis based on a subset of characteristics. This is of great interest, since some characteristics may be unreliable due to the interference of other medical disorders that the patient may have. It also allows specialists from different centers to adopt a concrete subset of characteristic criteria depending on the equipment available at the particular center. Sometimes, the specialist may not have access to the instrument required to determine a characteristic so some of the solutions cannot be used. After the multimodal optimization process, the diagnostic performance based on different subsets of characteristics is known. On the other hand, multiobjective optimization techniques are useful in the framework of diagnostic schemes where several concurrent objectives are to be taken into account (e.g., classification rate, sensitivity, invasive measurement, etc). Furthermore, more than one Pareto optimal solution is obtained when dealing with conflicting objectives, so the medical specialist is also able to benefit from this set of solutions.

In the following section, a general modular methodology for noninvasive medical diagnostic applications is presented, and it addresses the use of evolutionary multiobjective and multimodal optimization techniques in the diagnostic methodology. In the Section II, some of the results from the application of this improved methodology in the detection of paroxysmal atrial fibrillation (PAF) [6] are given. We conclude this work with a summary of the conclusions drawn from this investigation.

II. MATERIALS AND METHODS

In many situations, the medical diagnosis of health disorders (e.g., heart disease) can be addressed through the extraction of characteristic parameters from biosignals such as electrocardiograms (ECGs). In many cases, this kind of noninvasive diagnosis is more appropriate than other solutions involving surgery. In Fig. 2, a general noninvasive diagnostic methodology based on biosignals is illustrated.

This methodology involves the first stage of the general parameter definition, followed by the extraction of a set of characteristics (*characteristic vector*), which are used by the *diagnostic scheme* to obtain the decision label (positive/negative di-

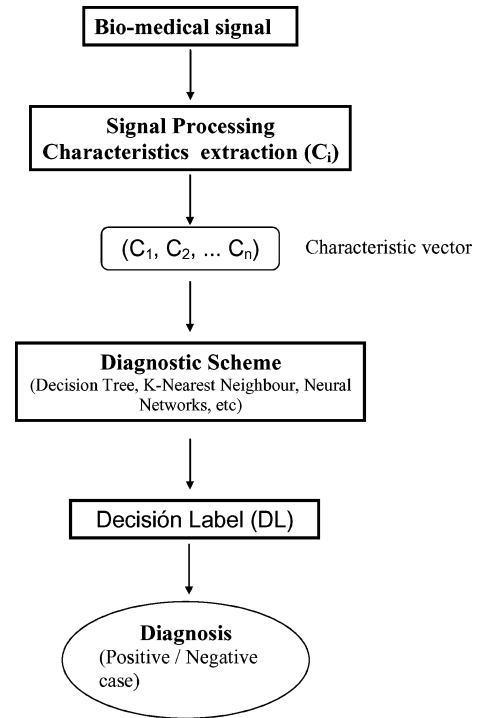


Fig. 2. Modular diagnostic methodology.

agnosis) of an unknown case. The characteristics are related to the physical properties of the biomedical signal. Thus, the diagnostic decision is a label obtained as a function of certain input items or characteristics (C_i) as indicated in (3). A diagnostic scheme (see Fig. 2) is applied to the input items, after which the diagnosis is obtained for each subject. The diagnosis methodology in Fig. 2 is modular so that different researchers are able to easily add (or remove) new characteristics to the vector

$$DL = f(C_1, C_2, \dots, C_n). \quad (3)$$

In what follows, we present different possible approaches, using multimodal (Section II-A) and multiobjective (Section II-B) evolutionary algorithms, to optimize the diagnosis used in the methodology of Fig. 2. The optimization stage (see dashed frame enclosed in Figs. 3 and 4) reduces the dimension of the characteristic vector (see Fig. 3) or determines the best values for the adjustable parameters of the diagnostic scheme (see Fig. 4) while taking into account one or several performance indicators. The optimization stage (see Figs. 3 and 4) requires the use of a training database (TDB) with known cases: positive or negative diagnosis. The values of the characteristics defined in the characteristics vector for each register in the TDB need to be extracted. A solution is a subset of characteristics (see Fig. 3) or certain values for the diagnostic scheme parameters (see Fig. 4). To evaluate each candidate solution, the diagnostic scheme computes a decision label (positive or negative) for each one of the TDB registers. In order to calculate the corresponding performance indicators of the diagnostic scheme, these computed labels are compared with ones already known. The EA maintains a population of candidate solutions of size: *popsiz*e (see Figs. 3 and 4). The evaluation procedure for each solution will depend on the specific diagnostic scheme

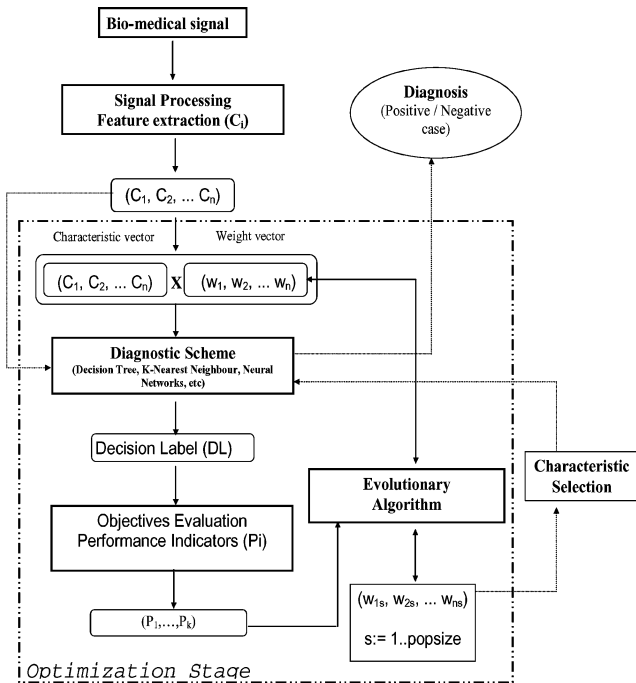


Fig. 3. Characteristic selection using an evolutionary algorithm. If more than one Objective functions are taken into account, a MOEA must be used; otherwise a multimodal EA is applied. Applying *diversity-maintaining-techniques* in the population of weight vectors enable to obtain different subsets of characteristics to be used in the diagnosis.

(see for example Appendix C). The crossover and mutation operators transform the candidate solutions and only the most suitable ones, those that lead to high performance indicators values, are maintained in the next iteration. After the optimization stage, in Figs. 3 and 4, dashed arrows are used to indicate the diagnostic methodology for new cases.

A. Using EAs for Multimodal Optimization

The *classification accuracy* (see Appendix A) reached after the diagnosis of a group of subjects is the most widely used performance indicator. The selection of the best components of the characteristic vector, in terms of classification accuracy maximization, can be achieved by means of a statistical study on the influence of each parameter in the diagnosis. Ideally, if the number of characteristic vector components could be reduced without affecting the diagnostic performance (measured in terms of classification accuracy), then the computational cost of the modular diagnostic methodology could be lowered. On other hand, the problem of finding a characteristic vector (see Fig. 2), which maximizes the classification accuracy, turns out to be a multimodal optimization problem, since several subsets of characteristic parameters that give reasonably high classification accuracy (global and local optima of the problem) can be found (see Section III). The existence of several solutions is an advantage since it gives the medical specialist some flexibility, i.e., the possibility of using different subsets of characteristics in the diagnostic methodology, as shown in Fig. 2. This multiplicity can be useful in those cases where some characteristics are unreliable due to the interfering presence of diseases different from the one being diagnosed, where there is instrumental error or even in the absence of specific equipment needed for

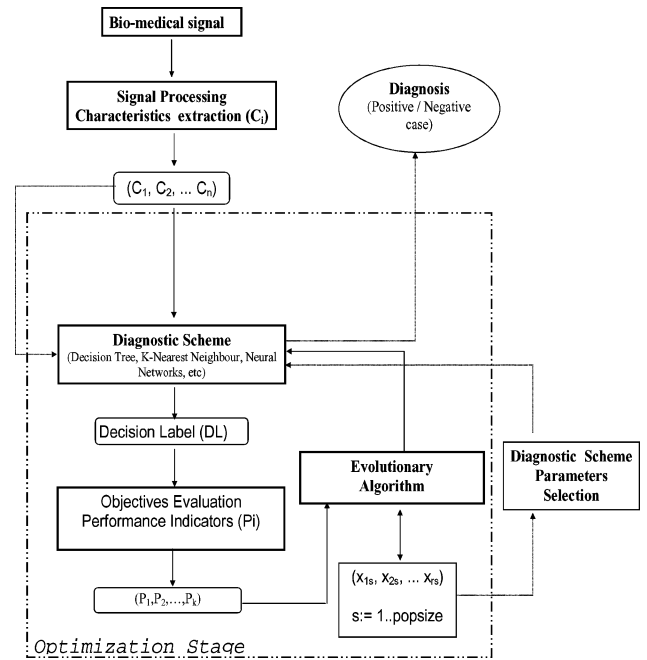


Fig. 4. Determining r parameters of the diagnostic scheme using an evolutionary algorithm.

the determination of a particular characteristic. Following the multimodal optimization process the medical specialist is provided with a diagnostic scheme that can be customized, i.e., he or she can choose which characteristics they prefer to be used in the diagnostic scheme. After this is done, only a single solution with these characteristics, from the ones provided by the multimodal optimization process, is selected. In this way, the medical specialist can bias the diagnostic scheme so that it is based on concrete features.

In order to both, reduce the number of characteristic vector components and to provide the medical specialist with the flexibility previously discussed, we propose using the optimization stage shown in Fig. 3.

The basic *characteristic vector* containing all defined parameters is multiplied by a *weight vector*. A value ranging between 0 and 1 is assigned to each component of the weight vector, according to the degree of influence that each component of the parameter vector has on the diagnostic scheme. To optimize the performance objectives (classification accuracy) the values of the weights are iteratively adjusted by an evolutionary algorithm in the way explained at the beginning of this section. Moreover, we are able to decide whether the components of the weight vector will take real values (ranging from 0 to 1) or binary values (0 or 1). In the first case, each solution (weight vector) shows the importance of each component of the characteristic vector. In the second case, each solution indicates which components of the characteristic vector should be considered in the diagnosis. The binary choice highly reduces the searching space, accelerating the convergence of the optimization process. Due to the fact that EAs work with a population of candidate solutions, different choices of weight vectors can be explored in a single iteration of the algorithm. If the necessary diversity mechanism [7]–[9] is incorporated into the EAs, at the end of the convergence, the set of solutions obtained (each solution is a different

subset of weights) is *diverse* and provides the specialist with the necessary flexibility.

B. Using EAs for Multiobjective Optimization

In what follows, we discuss two situations, where the optimization of more than one objective could be of interest. In both cases, a multiobjective evolutionary algorithm addresses the optimization of the two objectives.

1) *Maximization of the Classification Accuracy and Sensitivity*: Let us assume that a specific disease is to be detected. Label *P* (positive) means that the subject is suffering from this medical condition and label *N* (negative) represents a subject not suffering from the condition.

In addition to classification accuracy, there are two well-known basic functions to describe the diagnostic application performance: sensitivity and specificity (see Appendix A). The optimization of the sensitivity has a bias effect on the diagnostic scheme in favor of the accurate detection of subjects with a particular illness, regardless of the number of false positives produced. Whereas, if the specificity is optimized, the result is a system which accurately detects subjects who are not affected by the illness (healthy subjects), regardless of the number of false negatives produced.

The automatic individual optimization of one of these functions would generate inadequate systems, since a diagnostic scheme producing only *N* labels for all subjects would have a specificity of one and a sensitivity of zero. In contrast, a diagnostic scheme producing only *P* labels would obtain a sensitivity of one and a specificity of zero. Clearly, both systems are ineffective as diagnostic aids.

However, we are able to simultaneously optimize classification accuracy and sensitivity (or specificity) to favor the performance toward the accurate detection of patients (or subjects not suffering from the illness under consideration). If we were to use a noninvasive technique, such as ECG processing, for the parameter extraction, we might want to run an initial scan to look for potential patients in a wide population. In this case, high sensitivity is preferred over high specificity for the same classification accuracy. This means, that such an automatic diagnostic application could be used in routine explorations. Positive diagnosis would prompt more complex exploration processes. For this reason, we propose to optimize both the classification accuracy and the sensitivity of the algorithm (see Section III-B).

There are many other examples in which the optimization of two objective functions is a key issue. For instance, the accurate diagnosis of a particular disease among a specific phenotype of subjects may be of special interest at a given point. In that case, disease and specific phenotype detection define two different goals which would need to be dealt with simultaneously.

2) *Maximization of the Classification Accuracy and the Coverage Level*: Certain diagnostic schemes produce an output diagnostic indicator in such a way that a low value for this indicator is interpreted as a negative diagnosis and a high value as a positive diagnosis (see Fig. 5).

A confidence threshold is normally used in such classification schemes. Here, we would distinguish three different situations depending on the diagnostic indicator value: negative case, undiagnosed case and positive case. Thus, a confidence interval

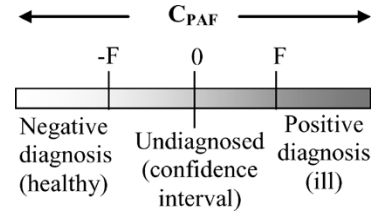


Fig. 5. Diagnostic indicator showing the confidence interval $[-F, F]$.

is provided for the diagnosis. Ideally, we would like to achieve both, a high level of classification accuracy and a high coverage level (4) with certain parameter vectors. In this scenario, the algorithm in Fig. 3 (characteristic selection) or Fig. 4 (determining the parameters of the diagnostic scheme) should maximize both the classification accuracy and the coverage level

$$CL = \frac{\text{Number of Diagnosed Cases}}{\text{Total Number of Cases}}. \quad (4)$$

III. RESULTS OF DIFFERENT CLASSIFICATION SYSTEMS FOR PAF DIAGNOSIS

We have applied the aforementioned methodology to the diagnosis of PAF. Atrial fibrillation is a heart arrhythmia and a common cause of systemic emboli, 75% of which generate cerebrovascular incidents [6]. PAF often leads to chronic or more serious types of arrhythmia, so an early diagnosis based on ECG traces would be very useful in medical practice. PAF diagnosis based on ECG traces that do not contain explicit fibrillation episodes is a topic still open to research [10]. Different approaches are presented in [10] which give promising results.

Diverse strategies have been applied for PAF detection on nonfibrillating recordings [10]. Diagnostic schemes based on incidence of isolated Atrial premature complexes [11]–[13] and on P-wave morphology variability [12] achieved the best performance. Other methods based on time-domain and frequency-domain analysis of heart rate variability were less successful [14]–[17]. Overall, the best performance so far has been obtained by combining different strategies, such as analysis of premature contractions, P-wave morphology and heart rate variability [12], [13]. This encourages the development of diagnostic schemes based on diverse characteristics. In this context the presented optimization scheme becomes interesting. To illustrate this we have extracted diverse characteristics mainly based on the morphological features of the P-wave [18]–[20] and the heart rate. We have then applied the presented optimization methodology and evaluated the performance of different approaches.

In order to evaluate the performance of this approach, we have optimized the diagnostic scheme using the characteristics described in Appendix B. However, if one of these characteristics is also influenced by another pathology that the subject may have, the accuracy of the diagnosis would be affected if that characteristic is used in the optimized classification approach. For example, if we were to use the RR distance in the diagnosis of a subject with a known cardiac pathology, the diagnostic accuracy would be affected if that particular characteristic were utilized in the optimized classification approach. On the other hand, if the medical specialist were able to run the diagnostic application invalidating the unreliable features, the resultant diag-

nosis would not be affected by the known cardiac pathology. In order to proceed in this way, the most suitable approach would be to address a multimodal optimization strategy that generates multiple solutions (i.e., diagnostic methods based on different features). Subsequent to the multimodal optimization process, the medical specialist is provided with a number of diagnostic approaches based on different subsets of characteristics. This may prove to be useful in the context of co-existing pathologies or when, for some reason, certain features cannot be extracted.

A public database for PAF prediction evaluation [10] provided by Physiobank [21] has been used in this diagnostic problem. It comprises the ECG records of 25 healthy individuals (n files) and 25 patients diagnosed with PAF (p files). The records are labeled (healthy or ill). It is important to note that none of these cases contains any explicit PAF episode and, therefore, the diagnostic algorithms proposed for this database will focus on ECG characteristics present in sinus rhythm (normal heart state).

Although several authors have used the same database to evaluate PAF screening systems, the comparison between the different approaches is not straightforward. Since some of them reported statistical differences between the training and test database, we have opted to focus on the results obtained using the training set (see Table I).

Both our approach and that employed in [22] used cross-validation to estimate the performance of the classifier on the training set. In other approaches [11], [12], [14]–[17], [23], simple decision rules with optimized thresholds for the training set have been implemented. This would possibly give better results than the ones expected using a valid test set. In [16] the authors used a K-nearest neighbor classifier and reported no results on the training set. Different kinds of features are used in each approach: Heart rate variability (HRV) features, power spectral density of the P wave and of HRV, P wave morphology, atrial premature complexes (APCs) number and timing.

The best results are obtained in [22] using HRV statistic features. However, as the authors point out, the results reported were obtained with the best configurations using parameter subsets. Therefore, they must be considered biased as these particular parameter subsets may be optimal for this particular training set, in spite of the use of cross validation. They can be compared to our best classification results, which are in the range of 84%. Therefore, in our approach we obtain performance rates similar to the best results reported in other works, in which power spectral density of HRV and HRV statistical features were used [22], or which were based on the P wave morphology and APCs [12]. Nevertheless, the methodologies found in [11], [12], [14]–[17], [22], [23] are based on a very restricted set of features. The efficient management of a large set is not straightforward. The main advantage of our classification scheme is its modularity, whereby it is easy to incorporate alternative features, such as the ones used by other researchers. The genetic algorithm will optimize the classifier, selecting the most representative sets of features. Furthermore, as explained in Sections III-B and III-C, several goals can be achieved simultaneously with this feature selection mechanism. Therefore, the results obtained confirm that the features used in this approach contain useful information for PAF screening. Furthermore, the proposed optimization scheme enables easy

TABLE I
COMPARATIVE RESULTS OF PAF SCREENING SYSTEMS BASED ON DIVERSE FEATURES AND CLASSIFICATION METHODS

Author	Features	Classification Method	Accuracy (%) (Training Set)
Chaz and Heneghan [22]	Power spectral density of HRV	Linear discriminant classifier	83.2*
Chaz and Heneghan [22]	HRV statistical features	Linear discriminant classifier	87.2*
Chaz and Heneghan [22]	P wave shape	Linear discriminant classifier	80.0*
Chaz and Heneghan [22]	Power spectral density of P wave	Linear discriminant classifier	71.9*
Langley et al. [23]	APCs	Decision rule	71
Zong et al. [11]	APCs	Decision rules	76
Maier et al. [14]	HRV statistical features, APCs and VPCs	Decision rules	80
Schreier et al. [12]	P wave morphology and APCs	Decision rule	82
Yang and Yin [15]	HRV features	Decision rule	74
Lynn and Chiang [16]	HRV features	K-nearest neighbour Inductive	0.68 (on the test set)
Krstacic et al. [17]	HRV features	Machine Learning by Logic Minimization (ILLM)	70
Toro et al. (our approach)	P wave morphology features and HRV features	K-nearest neighbour optimized with Genetic algorithms	82-84

PSD: Power Spectral Density

HRV: Heart Rate Variability

APC: Atrial Premature Complex

VPC: Ventricular Premature Complex

HRV statistical features: such as NN50, pNN50, RMSSD and SDDSD

P wave shape: sampled P wave interval

P wave morphology: features extracted from the P wave interval as presented in this contribution or by template comparison as in [12].

* Best results out of different classifiers configuration.

multimodal and multiobjective optimization of any modular classifier as described in previous sections.

In the presented contribution a low-level processing algorithm was used to extract 48 definable characteristics from the ECG records. The extracted characteristics were those considered important in mapping an ECG trace for PAF diagnosis from the study carried out in [18]–[20]. We have studied combinatorial searching schemes for the optimization of the classification system [19], [20]. However, in this paper we use evolutionary algorithms for the first time, which introduce interesting new features to the optimization process. The evolutionary algorithms perform multipath searching and widely explore the input space. In addition, in the previous approach [19], [20], local minima could not be avoided. Furthermore, this paper illustrates how the use of evolutionary algorithms allows multiobjective and multimodal optimization, which is interesting in the context of diagnostic systems. The characteristics extracted for PAF diagnosis are briefly described in Appendix B.

```

01 Create randomly Population  $P$  of  $Popsiz$ e individuals
02 While (stop_condition) FALSE
03  $P^* = \emptyset$ 
04 While (Sizeof( $P^*$ )  $\neq Popsiz$ e)
05   Select  $p_1$  and  $p_2$  from  $P$ 
06   Crossover  $p_1$  and  $p_2$  to obtain  $h_1$  and  $h_2$ 
07   Mutate  $h_1$  and  $h_2$  to obtain  $c_1$  and  $c_2$ 
08   If [ $Dis(p_1, c_1) + Dis(p_2, c_2) \leq [Dis(p_1, c_2) + Dis(p_2, c_1)]$ ]
09     If  $c_1 < p_1$  then  $P^* = P^* \cup \{c_1\}$  else  $P^* = P^* \cup \{p_1\}$ 
10     If  $c_2 < p_2$  then  $P^* = P^* \cup \{c_2\}$  else  $P^* = P^* \cup \{p_2\}$ 
11   Else
12     If  $c_1 < p_1$  then  $P^* = P^* \cup \{c_1\}$  else  $P^* = P^* \cup \{p_2\}$ 
13     If  $c_2 < p_2$  then  $P^* = P^* \cup \{c_2\}$  else  $P^* = P^* \cup \{p_1\}$ 
14   EndWhile
15  $P = P^*$ 
16 Evaluate individuals in  $P$  using the Diagnostic Scheme
17 EndWhile

```

Fig. 6. Pseudocode of the genetic algorithm based on deterministic crowding used for PAF diagnosis scheme optimization. Selection of individuals (step 05) is carried out without replacement. Mutation operator (step 07) is applied with a given mutation rate probability ($pmut$). The symbol $<$ means “better than.”

Hereafter, we describe three different optimization schemes corresponding to each of the examples introduced in the previous section.

A. Multimodal Optimization of the Classification Accuracy

For this application, we have chosen a K-nearest neighbor (KNN) classifier as the diagnostic scheme using the method *leaving one out* [24] (see Appendix C). Since the KNN algorithm is based on distance computing, the characteristics’ values were normalized so that they all lie between 0 and 1. In order to obtain sufficiently diversified solutions, population diversity maintaining techniques (i.e., *nicing methods* [7]) were embodied in the design of the evolutionary algorithm. An evolutionary algorithm with *deterministic crowding* [8] has been implemented for this purpose (see Fig. 6). Deterministic crowding has been chosen because of its good behavior [9] and user parameter absence. Here, we have used binary coding for the weight vector, so each solution is composed of 48 binary digits. A weight component equal to zero means that the characteristic is not used for the diagnosis and a weight component equal to one means that it is used. Crossover and mutation operators are illustrated in Fig. 7. The mutation rate (see Figs. 6 and 7) or $pmut$ was set to 0.1 after obtaining similar results with values of between 0.05 and 0.3. We maintained a population of 10 000 candidate solutions to obtain as many global and local optima as possible, so as to produce a good surface map of the optimization problem.

The algorithm has been executed for 1 000 000 evaluations of the objective function (classification accuracy) because the performance does not significantly improve afterwards. If a systematic search had been carried out, then $2^{48} \approx 3 \times 10^{14}$ different weight vectors would have been evaluated in the diagnostic scheme. By using a genetic algorithm we lowered that amount to 10^6 evaluations. The KNN classifier has been used

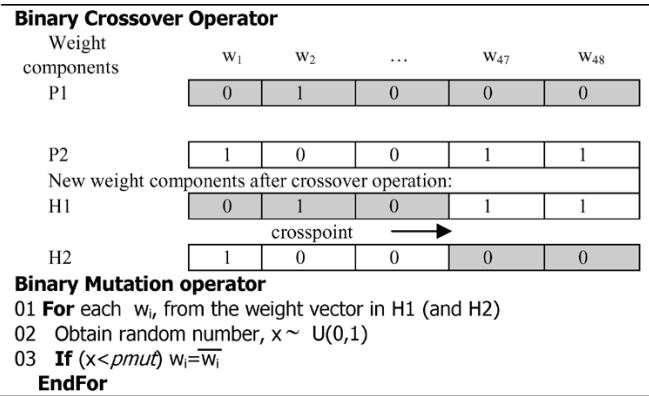


Fig. 7. Binary 1-point crossover and mutation operators. The “crosspoint” is a random number from a uniform distribution $U(1, 47)$.

TABLE II
BEST SOLUTIONS FOR THE PAF DIAGNOSIS

<i>Solution</i>	#1	#2	#3	#4
<i>QR</i>	0000	0000	0110	1010
<i>RR</i>	1011	1010	0101	1101
<i>PR</i>	0001	1001	1010	1110
<i>P_amp_N</i>	0001	1010	0101	1111
<i>P_wide</i>	0110	1011	0100	0010
<i>P_ini</i>	0001	0101	0001	1100
<i>P_int_N</i>	1010	1001	1001	1001
<i>PR2</i>	0101	0101	0001	0000
<i>P_amp2_N</i>	1000	1000	1111	1010
<i>P_wide2</i>	0000	0001	0101	0100
<i>P_ini2</i>	0111	1011	0110	0011
<i>P_int2_N</i>	1111	0011	1010	0111
<i>CA (%)</i>	84	82	82	80
<i>SENSI (%)</i>	92	88	88	84
<i>SPECI (%)</i>	76	76	76	76
<i>Number of Active Characteristics</i>	20	22	23	25

Each of the four solutions (weight vectors) shows the active characteristics (1) and the non-active characteristics (0) and the Classification Accuracy (CA), Sensitivity (SENSI) and Specificity (SPECI) reached. Sensitivity and Specificity have been computed a posteriori, i.e. they have not been optimised on the loop. The characteristics labels are explained in detail in Appendix B. The four components under each label correspond to the Average, Modified Maximum, Modified Minimum and Standard Deviation of each characteristic along the register (also explained in Appendix B).

with 5 neighbors. After 10 executions of the algorithm with different initial populations, 1.182 solutions with a classification accuracy above 70% were obtained.

The large number of quality solutions that have been found makes it difficult to extract conclusions on how the diagnostic performance is dependant on the various characteristics. In this sense, a final analytical stage following that of optimization would be desirable to automatically extract information from the set of solutions. In this way, conclusions on how sensitive the diagnostic scheme is with respect to the different characteristics may be reached. Here, we only focus on the four best solutions (weight vectors); those with a classification performance above 80%. Their characteristics are summarized in Table II, which shows how the four solutions are based on different ECG features (see Appendix B for features labels). The first solution centers on RR intervals and features extracted when several waves are detected in the P wave region (P_ini2 and P_int2_N).

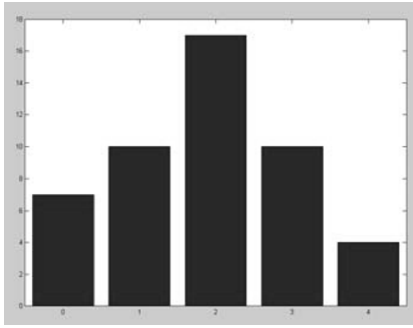


Fig. 8. The histogram illustrates the frequency of shared characteristics from the four best solutions found by an AE with deterministic crowding. Only four characteristics are shared by the four solutions, seven characteristics are not used by any solution and 17 characteristics are shared by two solutions.

The second solution focuses mainly on P_wide and P_ini2 but also covers RR , PR , P_amp_N , P_ini , and P_int_N . The third solution concentrates mainly on P_amp2_N but also considers a wide variety of different features (QR , RR , PR , P_amp_N , P_int_N , P_wide2 , P_ini2 , and P_int2_N). Finally, the third solution focuses strongly on P_amp_N and also interestingly, on RR , PR , and P_int2 . The solutions are significantly diverse. Nevertheless, there are four features used by these four best configurations: mean (P_int_N), mean (P_amp2_N), modified min (P_ini2) and modified min (P_int2_N). Therefore, it can be concluded that features related to the P wave morphology contain relevant information for PAF detection. Furthermore, the use of minimum measurements of features extracted, when more than a single wave is detected in the P wave region, may be relevant because records exist in which no single event of this kind is detected and, therefore, this causes the system to produce a N (negative) classification. That means that the specific detection of more than one wave in the P-wave region is of interest in PAF diagnosis.

To illustrate that the four solutions are functionally different, we calculate the *distance* between them as the number of *not common characteristics*: $d_{12} = 16$, $d_{13} = 21$, $d_{14} = 23$, $d_{23} = 27$, $d_{24} = 21$, and $d_{34} = 20$. The *distance* represents the difference between the solutions. Thus, for example, solutions 1 and 3 together are based on 43 characteristics and only 22 of them are shared since the distance is 21. This fact can also be corroborated by counting the number of *active-shared-characteristics*, as shown in the histogram in Fig. 8.

The classification accuracy average for the best four solutions is in the region of 82% and the sensitivity is around 88%. The diversity of the four best solutions represents a good example of the outcome of a multimodal optimization scheme. Therefore, the four solutions show similar performance but are based on different features (i.e., only 11 common components) due to the applied multimodal optimization strategy that assures diversity in the final solutions.

In this specific diagnostic task, the four best diverse solutions exhibit similar performance. However, in the framework of other diagnostic problems, if a feature that is very significant in the diagnosis of the target pathology is to be avoided, then a solution that excludes this characteristic will exhibit lower performance. Even so, the medical specialist may consider it in-

```

01 Create randomly Population P of PopsiZe individuals
02 Create external archive E
03 While (stop_condition) FALSE
04   Copy non dominated individuals from P to E
05   Delete dominated individuals from E
06   If Size (E) > maxESiZe then reduce Size of E with clustering
07   Assign fitness to each individual in P and E.
08   Binary Tournament Selection of PopsiZe individuals from P+E
09   Crossover and Mutation of individuals in P
10   Evaluate individuals in P using the Diagnostic Scheme.
EndWhile

```

Fig. 9. Pseudocode of SPEA used for PAF diagnostic scheme optimization.

teresting to carry out the diagnosis ignoring this feature for the reasons discussed in Section III.

B. Maximization of the Classification Accuracy and Sensitivity

For this application, we also used the KNN classifier as the diagnostic scheme. The strength pareto evolutionary algorithm (SPEA) [25] was chosen as the multiobjective evolutionary optimizer to maximize both classification accuracy and sensitivity concurrently.

In SPEA (see Fig. 9), an external archive is updated with the nondominated solutions found in each iteration of the algorithm from the creation of the initial population. Crossover and mutation operators are applied to the union of this external archive and to the current population. A nondominated individual is assigned a fitness value proportional to the number of solutions that it dominates. This fitness assignment ensures that the search process is biased toward nondominated solutions. Finally, a clustering technique (step 06 in Fig. 9) called the *average linkage method* [26] is used to introduce diversity to the external archive and to control its size.

In this application, the number of SPEA function evaluations during the experiments was 100 000 because after this, the results of the algorithm do not improve significantly. Crossover and mutation operators, shown in Fig. 7, were used with probabilities of 0.8 and 0.01 respectively and the external archive of Pareto optimal solutions has a maximum size of 30 candidate solutions (*maxESiZe* in Fig. 9). We used a population of 100 individuals. As a result, between 18–25 Pareto optimal solutions were obtained. The best Performance was obtained, again, with five neighbors in the KNN algorithm. The average of the four best Classification accuracies was 82% and 94.7% was the average sensitivity. For these solutions, the *a posteriori* calculated specificity average was 69.3%. In this case, it was observed that the sensitivity increases regardless of the specificity degradation because sensitivity is explicitly maximized in the procedure. As mentioned in the previous section, this optimization scheme could be of interest for the detection of possible PAF patients among the general population in preventive examinations. Therefore, it could be used as the scanning procedure for a general population provided that it is based on a noninvasive exploration, such as ECGs.

C. Maximization of the Classification Accuracy and the Coverage Level

Within this application, we use a diagnostic scheme based on weighted threshold dependent decision rules. A multiobjective

evolutionary algorithm optimizes the rules in order to maximize both the classification accuracy and the coverage level (see Section II-B). Again, SPEA [25] has been used as optimizer. For each of the 48 extracted characteristics, four different decision rules were applicable

$$\begin{aligned} \text{If } p_i < U_{i(\text{Low}_1)}, & \text{ then } C_{\text{PAF}} = C_{\text{PAF}} + W_{i1} \\ \text{If } p_i < U_{i(\text{Low}_2)}, & \text{ then } C_{\text{PAF}} = C_{\text{PAF}} - W_{i2} \\ \text{If } p_i > U_{i(\text{High}_1)}, & \text{ then } C_{\text{PAF}} = C_{\text{PAF}} + W_{i3} \\ \text{If } p_i > U_{i(\text{High}_2)}, & \text{ then } C_{\text{PAF}} = C_{\text{PAF}} - W_{i4}. \end{aligned}$$

where U represents different thresholds, C_{PAF} is the diagnostic indicator level which will determine the final diagnosis (see discussion in Section II-B) and the weights (W_{ij}) are constrained in the interval $[0,1]$. As described below, the first and third rules increase C_{PAF} , resulting in a positive PAF diagnosis, whereas the second and fourth rules decrease C_{PAF} , resulting in a negative PAF diagnosis.

In the diagnostic procedure, the C_{PAF} level is finally compared with a confidence interval $[-F, F]$. If C_{PAF} remains within this interval, we consider that there is not enough certainty about the diagnosis and this case is left undiagnosed. If $C_{\text{PAF}} > F$, the subject is diagnosed positive (as PAF patient); while if $C_{\text{PAF}} < F$, it is diagnosed negative.

High levels of F lead to an increase in the classification accuracy because the certainty of the diagnosis rises but it leads to a reduction in the coverage level [see (4)], since it leaves more cases undiagnosed. As can be seen, these two objectives clearly conflict with each other.

This approach could take into consideration the 48 extracted characteristics with four decision rules associated with each of them. In each decision rule, the variables to be fixed are the threshold (U) and the weight (W). Therefore, we are able to configure a vector of decision variables with 192 weights and 192 thresholds to be optimized (384 adjustable parameters).

In order to reduce the complexity of the optimization problem, and after a statistical study, an expert selected a subset of 32 rules and their associated 32 thresholds (which apparently maximize the discrimination power). In this way, if the expert thresholds are adopted, the vector of decision variables is reduced to 32 components (weights) or to 64 components if weights and thresholds are optimized.

Each EA has been executed during 200 000 function evaluations. The maximum level reached for C_{PAF} in the diagnosis was roughly $+2.5$ and the minimum -2.5 . Thus, the diagnostic range is 5 (Fig. 5). Parameter F is used with three different values, each representing percentages of the diagnostic range: $F = 0.5$ represents 20%; $F = 1$ is 40%; $F = 1.5$ makes up 60%. In Figs. 10–12, the solutions in the objective space are shown: classification accuracy (vertical axis) *versus* coverage level (horizontal axis).

The two optimization problems addressed in this application are as follows.

PAF Diagnosis: Optimization of Weights of Decision Rules Given by an Expert: In this case, the vector of decision variables has 32 components, and only the weights associated with the decision rules given by an expert are optimized [see Figs. 10(a)–12(a)]

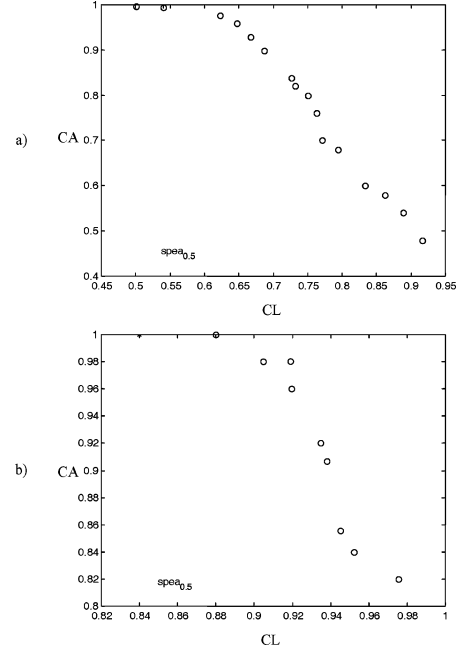


Fig. 10. Pareto optimal solutions using a confidence interval of 20%.

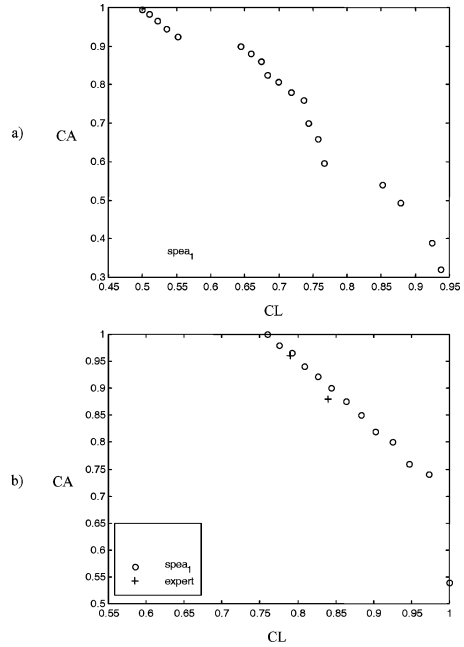


Fig. 11. Pareto optimal solutions using a confidence interval of 40%.

PAF Diagnosis: Optimization of Threshold and Weights of the Decision Rules Given by an Expert: In this case the vector of decision variables has 64 components. The parameters to be optimized are the 32 weights and 32 thresholds for the rules selected by an expert [see Figs. 10(b)–12(b)]. The solution found by an expert are available for $F = 1$ and included in Fig. 11(b).

D. Comparison Between the Different Approaches

The applications described in Sections III-A and III-B use KNN as their labeling scheme which is modular, i.e., new features can easily be included in the diagnostic scheme, although a posterior optimization process would be required.

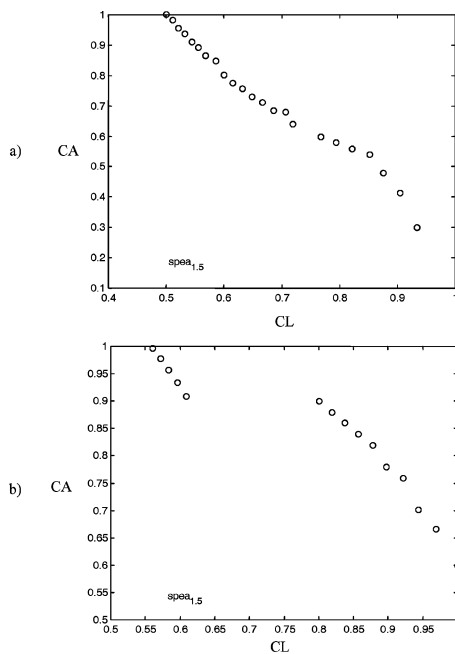


Fig. 12. Pareto optimal solutions using a confidence interval of 60%.

In Section III-A, the goal is to optimize a single objective (classification accuracy) by adopting a multimodal optimization strategy. In this case, the best solution achieves a classification accuracy of 84% and a sensitivity of 92%; and, more importantly, a population of diverse solutions is generated. Section III-B addresses the multiobjective optimization of the diagnostic scheme whereby the target objectives are classification accuracy and sensitivity. In this case, the population of solutions are not guaranteed to be based on different input characteristics, as the optimization strategy of this approach is not multimodal. In this instance, the best solutions achieve a classification accuracy of 82% and a sensitivity of 94.7% (the average of the four best solutions). The higher sensitivity is obtained because it is explicitly addressed in the multiobjective optimization process. Section III-C is based on a different classification scheme using decision rules. This labeling technique makes it easy to include medical specialist experience, by choosing key rules and thresholds. On the other hand, this approach is not modular and the inclusion of new characteristics is not as straightforward as in previous approaches, since new rules and thresholds need to be defined.

Two different approaches are considered in Section III-C. In the first one, the optimization process only adjusts the weights of the decision rules selected by the expert, using thresholds defined by him, whilst in the second approach, the optimization process changes both the weights and the thresholds of the decision rules. This second approach leads to better results, as illustrated in Figs. 10–12. This means that the thresholds selected by the expert, following visual inspection of the different characteristic topology, were not optimal

IV. CONCLUDING REMARKS

This paper addresses the use of evolutionary multimodal and multiobjective optimization techniques in noninvasive medical

diagnostic methodology based on biosignal processing. It has been put forward as a general diagnostic methodology whose adjustable parameters are optimized using an evolutionary algorithm addressing a multimodal or a multiobjective optimization problem. Three different examples of the application of this general methodology have been considered. All of these approaches have in common the production of a set of different solutions to be used in the diagnostic decision. In this scenario, the specialist has certain amount of freedom to select the parameters for the diagnostic decision process.

Evolutionary algorithms are used in a multimodal optimization scheme to provide different subsets of characteristics according to the physical properties of the biomedical signal. The specialist will then choose a specific subset of characteristics depending on the instrumentation available, the subject phenotype or other known diseases suffered by the patient treated.

Multiobjective optimization schemes make it possible to allow the simultaneous optimization of different goal functions (simultaneously). Multiobjective solutions represent good trade-offs between the different concurrent goals. This is particularly interesting in the detection of a specific illness among a certain phenotype of subjects, or when sensitivity is of special interest when scanning over a wide population for a specific disease.

Multimodal and multiobjective optimization schemes represent powerful tools, which enable modern medical centers to carry out thorough disease diagnostic studies. The various solutions can then be widely distributed so that other medical centers can benefit from complete studies without having to implement their own optimization procedures. It can even be useful when deciding which equipment to use in the diagnosis of a particular illness. It will also provide reference performance. The multiobjective solutions represent highly optimized trade-offs between different goal functions. Furthermore, the diversity of the solutions provides the medical specialist or the medical center with a degree of flexibility which enables the diagnosis based on the specific physical properties of the biosignals (characteristics).

Different optimization schemes for a specific diagnostic application (PAF pathology) are adopted in Section III to illustrate how multiobjective and multimodal optimization strategies can be used in this field. Section III-A describes the multimodal optimization of a diagnostic scheme with a single target function – classification accuracy. For PAF diagnosis, we obtain a set of diverse solutions thanks to this multimodal strategy. The four best solutions achieve an average classification accuracy of 82% and a sensitivity of 88%, which is not optimized in the loop. Section III-B addresses multiobjective optimization to include both classification accuracy and sensitivity. In this case, solution diversity cannot be guaranteed. If the four best solutions are taken into account, this approach achieves an average classification accuracy of 82% and a sensitivity of 94.7%. Note that the sensitivity rate is higher because it is explicitly optimized. Finally, Section III-C adopts a different classification method based on decision rules. This scheme facilitates the input of medical expertise, but it is not modular and, therefore, the inclusion of new characteristics is not straightforward. The set up described in Section III-C achieves higher results, i.e., a classification accuracy of above 98% using optimized weights and thresholds.

However, in this approach some cases are left undiagnosed due to the confident parameter F . The application described in Section III-C makes the most of the medical specialist's experience, since he or she can choose the rules which make up the classification system (whereas the optimization process only adjusts the weights and thresholds of these rules).

More detailed results from characteristic scanning procedures can be found in [19] and [20], which shows single path and single objective searching techniques. The optimization strategies presented in this paper represent a development on results previously achieved.

All the applications for PAF prediction considered here are based mainly on characteristics extracted from the morphological analysis of the P-wave (see Appendix B). But the presented diagnostic schemes and optimization methodology are completely modular. This allows the easy inclusion of new characteristics in the study. In future work related with PAF pathology prediction we plan to include other time-domain and frequency-domain characteristics as well as premature atrial contraction analysis. We also plan to study *incremental optimization schemes* that facilitate the inclusion of new characteristics in the diagnostic scheme without requiring a complete re-optimization process. That means that the optimization procedure should start from the already preoptimized set of parameters and include the new ones. Finally, our current work is focused on knowledge extraction techniques from the multimodal optimized schemes (solutions surface). This is important for the interpretation of the diagnosis by a specialist. It will also facilitate hybrid diagnostic schemes which combine automatic tools and specialist knowledge. For this purpose, we are studying function approximation techniques in order to translate the solutions surface to a set of multiple-characteristic-based diagnostic rules.

APPENDIX I

CLASSIFICATION ACCURACY, SENSITIVITY AND SPECIFICITY

For biomedical diagnostic applications, the final diagnosis is that either a patient is *ill* (suffering a certain pathology) or *healthy* (free from this particular pathology). This means that the classification result can be one of the following cases:

True Positive (TP)	The algorithm classifies the subject as ill and the subject is in fact ill.
True Negative (TN)	The algorithm classifies the subject as healthy and the subject is in fact healthy.
False Positive (FP)	The algorithm classifies the subject as ill but the subject is healthy.
False Negative (FN)	The algorithm classifies the subject as healthy but the subject is ill.

Within these cases, different functions of interest can be defined:

Classification Accuracy (CA):

$$CA = \frac{TP + TN}{TP + TN + FP + FN}. \quad (5)$$

Sensitivity: represents the ratio between the detected ill patients and the total ill patients.

$$SENSI = \frac{TP}{TP + FN}. \quad (6)$$

Specificity: represents the ratio between the detected healthy subjects and the total healthy subjects.

$$SPECI = \frac{TN}{TN + FP}. \quad (7)$$

APPENDIX II

CHARACTERISTICS EXTRACTED FROM ECG TRACES FOR PAF DIAGNOSIS

The ECG of the PAF database [10] was provided with automatically generated QRS annotations. Using these marks, we refined the R wave allocation and marked candidate points for Q, S and T waves. The interval between T and the next Q wave was scanned in detail to try to detect the P wave. Most of the proposed parameters related to the P wave morphology, as the main auricular activity wave. From the interval between T and Q (the *P wave region*), we extracted a number of morphology parameters: maximum voltage amplitude or time interval of the widest wave, among others, in order to decide whether a clear single P wave was present or not.

For every heartbeat, extremely heterogeneous characteristics were defined, which might be of interest in PAF diagnosis. We extracted the following characteristics for each cardiac pulsation.

- a.1) **QR**: QR voltage amplitude.
- a.2) **RR**: Time distance between two consecutive R waves.

A time window was defined in which the P wave should be localized, known as the P wave region. Then, it was necessary to determine whether there was a single P wave in this time window. If so, the following characteristics were extracted.

- b.1) **PR**: Time distance between the maximum of the single detected P wave and that of the next R wave.
- b.2) **P_amp_N**: The voltage amplitude of the P wave (P_amp_N) from the ECG. This was normalized to reduce the effect of different amplitude registers, by calculating the ratio $P_amp_N = P_amp/QR$.
- b.3) **P_wide**: Time width of the P wave measured at the isoelectric level of a particular heartbeat.
- b.4) **P_ini**: Time distance from the beginning of the P wave to its maximum.
- b.5) **P_int_N**: P_int was taken as the integral of the P wave (above the isoelectric level of this particular heartbeat). The voltage was normalized in the case of the final characteristic (P_int_N) to avoid the effects of different amplitude registers by calculating the ratio: $P_int_N = P_int/QR$.

On the other hand, if several waves were detected in the time window (*P wave region*), the following characteristics were extracted:

- c.1) **PR2**: Time distance between the maximum of the last wave of this time window and the maximum of the next R wave.
- c.2) **P_amp2_N**: The average voltage amplitude of the waves detected in this window (P_{amp2}) was calculated, after which it is normalized by calculating $P_{amp2_N} = P_{amp2}/QR$.
- c.3) **P_wide2**: The average time width of the waves detected in a window.
- c.4) **P_ini2**: The average of time distances in a window from the beginning of each wave to its maximum.
- c.5) **P_int2_N**: P_{int2} was calculated as the average of the wave integrals detected in the window. Then, $P_{int2_N} = P_{int2}/QR$ was normalized.

All these characteristics were calculated for heartbeats in which one or more waves of auricular activity were present: P or Fibrillation (F) waves. This fact limited the number of characteristic estimations extracted from each ECG file. ECG files might exist with no estimation for the third set of parameters (c), because no clear F wave was detected.

Once individual estimations of the parameters were obtained for each cardiac pulse along the entire ECG register, four global register descriptors from each characteristic sequence were calculated:

- 1) *Average (Mean)*: The average value of a characteristic in a register.
- 2) *Standard Deviation (Std)*: The stability of the characteristic along the register.
- 3) *Modified Maximum (Max)*: The maximum value of a characteristic.
- 4) *Modified Minimum (Min)*: The minimum value of the characteristic.

In order to reduce the effects of possible artefacts in the register, the *modified* maximum (minimum) was calculated as the average of the five greatest (lowest) values along the register.

In this way, each ECG register is characterized by a 48-component vector. Many of these components are related to the morphology of the P-wave. These characteristics are not orthogonal; in fact, some of them are highly redundant. However, these redundancies are compensated for during the optimization process by the evolutionary algorithms used in the searching space. Other characteristics, such as frequency components, have also been studied by other authors in the context of PAF diagnosis [10]. There were great similarities in accuracy between diagnostic schemes using these characteristics and those presented in this work.

APPENDIX III

THE K-NEAREST NEIGHBOR CLASSIFIER

A modular classification algorithm based on the nearest K-neighbors was used for this application. The labeled 48-component vectors from the database [10] act as references for the classification system. The KNN classifier works as follows: the Euclidean distances to the labeled vectors used as classification reference are calculated for each new unlabeled vector, the labels of the nearest K-neighbors consulted, and the final label is calculated through a voting scheme as the label of

the majority of the K-neighbors. In this way, the classification algorithm is modular and new parameters can be easily added, as only the dimension considered in the Euclidean distance calculation step needs to be modified. The modularity of the classification algorithm facilitates the parameter optimization scheme illustrated in.

Due to the small size of the test database (25 PAF patients and 25 non-PAF subjects), the evaluation of the classification accuracy (and sensitivity) is calculated in 50 cycles by the *leaving one out* method [24], i.e., in each cycle, one vector is selected from the database as the test element. This vector is classified according to the scheme described above, with the other 49 labeled vectors serving as classification references. In each cycle the classification results are updated in four counters (see Appendix A): True_Positive (TP), True_Negative (TN), False_Positive, (FP) and False_Negative (FN). Finally, the classification accuracy (and sensitivity) is calculated following the expressions (5) and (6).

REFERENCES

- [1] J. P. Li, M. E. Balazs, G. T. Parks, and P. J. Clarkson, "A species conserving genetic algorithm for multimodal function optimization," *Evol. Comput.*, vol. 10, no. 3, pp. 207–234, 2002.
- [2] J. P. Cohoon and D. H. Marks, "A review and evaluation of multiobjective programming techniques," *Water Resources Res.*, vol. 11, no. 2, pp. 208–220, 1975.
- [3] A. E. Eiben and J. E. Smith, *Introduction to Evolutionary Computing*. New York: Springer, 2003, Natural Computing Series.
- [4] C. A. Coello, D. A. Van Veldhuizen, and G. B. Lamont, *Evolutionary Algorithms for Solving Multi-Objective Problems*. Norwell, MA: Kluwer Academic, 2002.
- [5] K. Deb, *Multiobjective Optimization Using Evolutionary Algorithms*. New York: Wiley, 2002.
- [6] P. Petersen and J. Godtfredsen, "Embolic complications in paroxysmal atrial fibrillation," *Stroke*, vol. 17, pp. 622–626, 1986.
- [7] B. Sareni and L. Krähenbühl, "Fitness sharing and niching methods revisited," *IEEE Trans. Evol. Comput.*, vol. 2, no. 3, pp. 97–106, Sep. 1998.
- [8] S. W. Mahfoud, "Crowding and preselection revised," in *Parallel Problem Solving from Nature*, R. Manner and B. Manderick, Eds. Amsterdam, The Netherlands: Elsevier, 2002, vol. 2, pp. 27–36.
- [9] —, "A comparison of parallel and sequential niching methods," presented at the 6th Int. Conf. Genetic Algorithms, San Mateo, CA, 1995.
- [10] (2001) Computers in Cardiology. IEEE press. [Online]. Available: <http://physionet.cps.unizar.es/physiobank/database/afpdb/>
- [11] W. Zong, R. Mukkamala, and R. G. Mark, "A methodology for predicting paroxysmal atrial fibrillation based on ECG arrhythmia feature analysis," in *Proc. Computers in Cardiology*, Rotterdam, The Netherlands, 2001, pp. 125–128.
- [12] G. Schreier, P. Kastner, and W. Marko, "An automatic ECG processing algorithm to identify patients prone to paroxysmal atrial fibrillation," in *Proc. Computers in Cardiology*, Rotterdam, The Netherlands, 2001, pp. 133–136.
- [13] T. Thong, J. McNames, M. Aboy, and B. Goldstein, "Prediction of paroxysmal atrial fibrillation by analysis of atrial premature complexes," *IEEE Trans. Biomed. Eng.*, vol. 51, pp. 561–569, Apr. 2004.
- [14] C. Maier, M. Bauch, and H. Dickhaus, "Screening and prediction of atrial fibrillation by analysis of heart rate variability parameters," in *Proc. Computers in Cardiology*, Rotterdam, The Netherlands, 2001, pp. 129–132.
- [15] A. C. Yang and H. W. Yin, "Prediction of paroxysmal atrial fibrillation by footprint analysis," in *Proc. Computers in Cardiology*, Rotterdam, The Netherlands, 2001, pp. 401–404.
- [16] K. S. Lynn and H. D. Chiang, "A two-stage solution algorithm for paroxysmal atrial fibrillation," in *Proc. Computers in Cardiology*, Rotterdam, The Netherlands, pp. 405–408.
- [17] G. Krstacic, D. Gamberger, T. Smuc, and A. Krstacic, "Some important R-R interval based paroxysmal atrial fibrillation predictors," in *Proc. Computers in Cardiology*, Rotterdam, The Netherlands, 2001, pp. 409–412.

- [18] S. Mota, E. Ros, F. J. Fernandez, A. F. Diaz, and A. Prieto, "ECG parameter characterization of paroxysmal atrial fibrillation," in *Proc. 4th International Workshop on Biosignal Interpretation*, 2002, pp. 247–250.
- [19] E. Ros, S. Mota, F. J. Toro, A. F. Díaz, and F. J. Fernández, "Automatic paroxysmal atrial fibrillation based on not fibrillating ECGs," *Meth. Inf. Med.*, vol. 43, no. 1, pp. 94–98, 2004.
- [20] E. Ros, S. Mota, F. J. Fernández, F. J. Toro, and J. L. Bernier, "ECG characterization of paroxysmal atrial fibrillation: Parameter extraction and automatic diagnosis algorithm," *Comput. Biol. Med.*, vol. 34, no. 8, pp. 679–696, 2004.
- [21] A. Goldberger, L. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. 215–220, 2000.
- [22] P. de Chanzal and C. Heneghan, "Automated assessment of atrial fibrillation," in *Proc. Computers in Cardiology*, Rotterdam, The Netherlands, 2001, pp. 117–120.
- [23] P. Langley, D. di Bernardo, J. Allen, E. Bowers, F. E. Smith, S. Vechietti, and A. Murray, "Can paroxysmal atrial fibrillation be predicted?," in *Proc. Computers in Cardiology*, Rotterdam, The Netherlands, 2001, pp. 121–124.
- [24] D. Hand, *Discrimination and Classification*. New York: Wiley, 1981.
- [25] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: A comparative case study and the strength pareto approach," *IEEE Trans. Evol. Comput.*, vol. 3, pp. 257–271, Nov. 1999.
- [26] J. Morse, "Reducing the size of the nondominated set: Pruning by clustering," *Comput. Oper. Res.*, vol. 7, no. 2, pp. 55–66, 1980.



Francisco de Toro received the B.S. degree in physics, the M.S. degree in electronics engineering, and the Ph.D. degree from the University of Granada, Granada, Spain, in 1993, 1996, and 2003, respectively.

He is currently Assistant Professor at the Department of Signal Theory, Telematics and Communications at the same university. His research interests include evolutionary computation for multiobjective optimization, machine learning, and parallel processing.



Eduardo Ros received the B.S. degree in physics, the M.S. degree in electronics engineering, and the Ph.D. degree from the University of Granada, Granada, Spain, in 1993, 1996, and 1997, respectively.

He is currently Associate Professor at the Department of Computer Architecture and Technology at the same university. He is currently the responsible Researcher at the University of Granada of two European projects related with bio-inspired processing schemes. His research interests include biomedical signal processing, hardware implementation of digital circuits for real time processing in embedded systems and computer vision.



Sonia Mota received the BS degree in physics in 1998 and the MS degree in 2002 from the University of Granada, Granada, Spain. She is currently working towards the Ph.D. degree as an Assistant Researcher at the Department of Computer Architecture and Technology.

Her research interests include biomedical signal processing, artificial vision and hardware implementation of bio-inspired processing schemes.



Julio Ortega received the B.S. degree in electronic physics, the M.S. degree in electronics, and the Ph.D. degree from the University of Granada, Granada, Spain, in 1985, 1986, and 1990, respectively.

He was at the Open University, U.K., and with the Department of Electronics, University of Dortmund, Dortmund, Germany, as an Invited Researcher. Currently he is a Full Professor in the Department of Computer Architecture and Technology of the University of Granada. His research interest are in the fields of parallel processing and parallel computer architectures, artificial neural networks, and evolutionary computation. He has led research projects supported by the Spanish Commission of Science and Technology (CICYT) in the area of parallel algorithms and architectures for optimization problems.

Dr. Ortega's Ph.D. degree dissertation received the Award of Ph.D. dissertations of the University of Granada.