A Framework for Fuzzy Expert System Creation—Application to Cardiovascular Diseases

Markos G. Tsipouras, *Student Member, IEEE*, Costas Voglis, and Dimitrios I. Fotiadis*, *Senior Member, IEEE*

***Abstract*—A methodology for the automated development of fuzzy expert systems is presented. The idea is to start with a crisp model described by crisp rules and then transform them into a set of fuzzy rules, thus creating a fuzzy model. The adjustment of the model's parameters is performed via a stochastic global optimization procedure. The proposed methodology is tested by applying it to problems related to cardiovascular diseases, such as automated arrhythmic beat classification and automated ischemic beat classification, which, besides being well-known benchmarks, are of particular interest due to their obvious medical diagnostic importance. For both problems, the initial set of rules was determined by expert cardiologists, and the MIT-BIH arrhythmia database and the European ST-T database are used for optimizing the fuzzy model's parameters and evaluating the fuzzy expert system. In both cases, the results indicate an escalation of the performance from the simple initial crisp model to the more sophisticated fuzzy models, proving the scientific added value of the proposed framework. Also, the ability to interpret the decisions of the created fuzzy expert systems is a major advantage compared to "black box" approaches, such as neural networks and other techniques.**

***Index Terms*—Arrhythmic beat classification, expert systems, fuzzy modeling, ischemic beat classification.**

I. Introduction

MEDICAL expert systems are a challenging field, requiring the synergy of different scientific areas. The representation of medical knowledge and expertise, the decision making in the presence of uncertainty and imprecision, and the choice and adaptation of a suitable model are some issues that a medical expert system should take under consideration. Uncertainty is traditionally treated in a probabilistic manner; recently, however, methods based on fuzzy logic have gained ground [1], [2]. The model's parameter adaptation (training) amounts to optimizing a properly constructed "error" function. There is a variety of methods with diverse features that may be proper. Understanding the subtleties of the optimization procedures is a key to choosing an effective training approach.

Expert systems are a branch of artificial intelligence, which make extensive use of specialized knowledge to solve problems at the level of a human expert. This knowledge is represented in by a set of rules [3]. An expert system's review of applications is presented in [4]. An expert system is created by defining a crisp or fuzzy model (set of rules) and then optimizing its parameters to fit a given dataset. Several approaches have been proposed in the literature for the development of fuzzy or crisp models. In most of them, the model is trained using a known optimization technique, i.e., fuzzy rules with genetic algorithms [5], fuzzy rules with simulated annealing [6], multicriteria decision analysis with genetic algorithms [7]. Neuro-fuzzy algorithms have also been proposed, where, the fuzzy rules are modeled by artificial neural networks (ANNs) and popular training techniques are applied [8].

In this paper, a framework for the automated generation of a fuzzy expert system (FES) is proposed. The framework is based on rules, which are initially represented using the crisp membership function, forming a crisp model. The rules are then transformed from crisp to fuzzy ones, using a fuzzy membership function and $T_{\rm norm}$ and $S_{\rm norm}$, which are fuzzy equivalence for the binary AND and OR operators, respectively [1]. Using different selections for the fuzzy membership function and different definitions for the $T_{\rm norm}$ and $S_{\rm norm}$, several fuzzy models can be created. Then, the fuzzy model is tuned so as to find optimal parameters of the fuzzy membership functions, and, if necessary, parameters for the $T_{\rm norm}$ and $S_{\rm norm}$; the fuzzy model combined with the optimal parameters comprises a FES. The proposed framework is applied to two well-known cardiovascular domain problems, the arrhythmic beat classification and the ischemic beat classification from electrocardiograms (ECGs).

In the following, initially some basics of the classification problem and fuzzy logic are briefly described and then the framework for the automated FES creation is presented in detail. Next, the two domains of application are described (arrhythmic beat classification, ischemic beat classification) along with the employed datasets for each one, the initial medical rules, the respective crisp models and the FESs, automatically generated from the proposed methodology. Also, results from the evaluation of the created FESs are presented. In the following, the scientific added value of the methodology along with its advantages and disadvantages are addressed. Also, the

Manuscript received August 9, 2006; revised January 9, 2007. This work was supported by the program "Heraklitos" of the Operational Program for Education and Initial Vocational Training of the Hellenic Ministry of Education under the 3rd Community Support Framework. *Asterisk indicates corresponding author.*

M. G. Tsipouras is with the Unit of Medical Technology and Intelligent Information Systems, Department of Computer Science, University of Ioannina, GR 45110 Ioannina, Greece (e-mail: markos@cs.uoi.gr).

C. Voglis is with the Department of Computer Science, University of Ioannina, GR 45110 Ioannina, Greece (e-mail: voglis@cs.uoi.gr).

*D. I. Fotiadis is with the Unit of Medical Technology and Intelligent Information Systems, Department of Computer Science, University of Ioannina, P.O. Box 1186, GR 45110 Ioannina, Greece (e-mail: fotiadis@cs.uoi.gr).

Digital Object Identifier 10.1109/TBME.2007.893500

generated FESs and their results are discussed. Finally, further improvements of the automated methodology are discussed.

II. Methodology

First, definitions and related terminology used in the classification problems and fuzzy logic, are introduced. Having the data $D = \{d^l, c^l\}$, $l = 1, \ldots, K$, where $d^l \in \Re^n$ is a single pattern with n features, $c^l \in \{1, 2, \ldots, m\}$ is its class (m is the number of classes), and K is the total number of patterns (the size of D), a classification problem is defined as the determination of a mapping model $M(\cdot)$, where $M(d^l) = c^l$ [9]. An alternative representation for the class is $c^l \in \{0,1\}^m$, where, if d^l belongs to class i, then $c^l = e_i\,(e_1 = (1,0,0,\ldots,0)\,, e_2 = (0,1,0,\ldots,0), \ldots, e_m = (0,0,0,\ldots,1))$. A common methodology to treat a classification problem is to define a mapping model and train it, using a subset of the data and a cost function, which is minimized. A tool that is used for the evaluation of a classification model is the normalized confusion matrix, having dimension $m \times m$, defined as

$$X_{i,j} = \frac{\#\text{ of patterns in class } j \text{ classified to class } i}{\text{total } \# \text{ patterns in class } i} \tag{1}$$

where $X_{i,j}$ is the (i,j) element of the confusion matrix. Crisp logic is the binary reasoning. Membership functions are fundamentals in set theory, measuring the certainty of an object x belonging to a set S. The membership function used in crisp logic, is a binary operator and its value is 1 or 0, representing that x does or does not belongs to S, respectively. Fuzzy logic is a generalization of the classical set theory [1], [2]. It has been used to represent and manage the vagueness, which arises in data or in expert's knowledge [10]. The fuzzy logic is based on fuzzy membership functions, which are continuous approaches that have values in the interval $[0,1]$, representing the relationship between the object x and the set S.

A. Crisp Model

A crisp model consists of m crisp rules $R_i(d^l, \theta_i)$ $(i = 1, \ldots, m)$, where θ_i is a vector containing all parameters (thresholds) used in the ith rule and m is the number of classes; thus, one rule is defined for each class. Each R_i consists of several simple rules $r_{i,j}(d^l, \theta_{i,j})(j = 1, \ldots, \tau)$, defined as $r_{i,j}(d^l, \theta_{i,j}) = g_c(f_{i,j}(d^l), \theta_{i,j})$ (the jth simple rule in the ith rule), where $f_{i,j}(\cdot)$ is a function of the data d^l, $\theta_{i,j}$ is a parameter (the jth parameter in the θ_i vector), τ is the number of simple rules $r_{i,j}$ used in R_i and $g_c(\cdot)$ is the crisp membership function (increasing or decreasing), defined as

$$g_c^{\text{inc}}(x,\theta) = \begin{cases} 0, & x \le \theta \\ 1, & x > \theta \end{cases} \text{ (increasing) or}$$

$$g_c^{\text{dec}}(x,\theta) = \begin{cases} 1, & x \le \theta \\ 0, & x > \theta \end{cases} \text{ (decreasing).} \tag{2}$$

Each rule $R_i(d^l, \theta_i)$ can be expressed as a combination of $r_{i,j}(d^l, \theta_{i,j})$ simple rules, as follows [see (3), shown at the bottom of the page] where $\theta_i = \{\theta_{i,j}\}$, $j = 1, \ldots, \tau$. A simple rule $r_{i,j}$ is a rule that contains only one inequality (e.g., $x > 0$). Having several instances of an object belonging to category i, each row of the $R_i(d^l, \theta_i)$ includes all simple rules $r_{i,j}(d^l, \theta_{i,j})$, which are related to a single object instance. Then, the $R_i(d^l, \theta_i)$ combines all instances related to the same class. The final decision (class) of the crisp model (M_c) is made using the results from all rules: $M_c(d^l, \Theta) = F_c(R_1(d^l, \theta_1), R_2(d^l, \theta_2), \ldots, R_m(d^l, \theta_m))$, where Θ is a vector containing all thresholds $(\Theta = \{\theta_1, \theta_2, \ldots, \theta_m\})$ and $F_c(\cdot)$ is a function that combines the outcomes of all $R_i(d^l, \theta_i)$ crisp rules and results to one of the classes. Depending on the representation selected for c^l, the final decision is $F_c \in \{1, 2, \ldots, m\}$, where m is the number of classes, or $F_c \in \{0,1\}^m$, where, if d^l is classified to class i, then $F_c = e_i$. A more general definition of the $F_c(\cdot)$ function could include an additional result $(m+1)$, which states that the classification process failed (i.e., for a single case two or more rules were true). In this case the final decision is $F_c \in \{1, 2, \ldots, m, m+1\}$ or $F_c \in \{0,1\}^m$ (but not necessary $F_c = e_i$).

Each row of the $R_i(d^l, \theta_i)$ rule (i.e., $r_{i,1}(d^l, \theta_{i,1})$ AND $r_{i,2}(d^l, \theta_{i,2})$ AND…AND $r_{i,a_1}(d^l, \theta_{i,a_1})$) is a conjunction (sequence of AND) of one or more simple rules and the $R_i(d^l, \theta_i)$ rule is a disjunction (sequence of OR) of its rows. This form is known as disjunctive normal form (DNF) and has been chosen because every logical expression (i.e., set of rules) can be written in DNF.

B. Fuzzy Model

The crisp model is transformed into a fuzzy model using a fuzzy membership function $g_f(x,\theta)$ instead of the crisp $g_c(x,\theta)$. In this case, θ is a vector containing all parameters used in the fuzzy membership function and its size depends on the selection of the fuzzy membership function. Table I presents some monotonic fuzzy membership functions along with the parameters (θ) needed for each one. Also, T_{norm} and S_{norm} are used; Table II presents some common definitions for the T_{norm}

$$R_i(d^l, \theta_i) = \begin{cases} (r_{i,1}(d^l, \theta_{i,1}) \text{ AND } r_{i,2}(d^l, \theta_{i,2}) \text{ AND } \ldots \text{ AND } r_{i,a_1}(d^l, \theta_{i,a_1})) & \text{OR} \\ (r_{i,a_1+1}(d^l, \theta_{i,a_1+1}) \text{ AND } r_{i,a_1+2}(d^l, \theta_{i,a_1+2}) \text{ AND } \ldots \text{ AND } r_{i,a_2}(d^l, \theta_{i,a_2})) & \text{OR} \\ \ldots & \\ (r_{i,a_l+1}(d^l, \theta_{i,a_l+1}) \text{ AND } r_{i,a_l+2}(d^l, \theta_{i,a_l+2}) \text{ AND } \ldots \text{ AND } r_{i,\tau}(d^l, \theta_{i,\tau})) & \end{cases} \tag{3}$$

TABLE I
MONOTONIC FUZZY MEMBERSHIP FUNCTIONS

function	increasing	decreasing	parameters
linear	$g^{inc}_{f_{lin}}(x,a,b)=\begin{cases}0 & x\le a\\ \frac{x-a}{b-a} & a<x<b\\ 1 & b\le x\end{cases}$	$g^{dec}_{f_{lin}}(x,a,b)=\begin{cases}1 & x\le a\\ \frac{x-b}{a-b} & a<x<b\\ 0 & b\le x\end{cases}$	$\theta=\{a,b\}$
sigmoid	$g^{inc}_{f_{sig}}(x,a,b)=\frac{1}{1+e^{a(b-x)}}$	$g^{dec}_{f_{sig}}(x,a,b)=\frac{1}{1+e^{a(x-b)}}$	$\theta=\{a,b\}$
sum of sigmoid and its gradient	$g^{inc}_{f_{sig}}(x,a,b)+\left(g^{inc}_{f_{sig}}(x,a,b)\right)'$	$g^{dec}_{f_{sig}}(x,a,b)+\left(g^{dec}_{f_{sig}}(x,a,b)\right)'$	$\theta=\{a,b\}$
nested sigmoid	$g^{inc}_{f_{sig}}\left(g^{inc}_{f_{sig}}(x,a_1,b_1),a_2,b_2\right)$	$g^{dec}_{f_{sig}}\left(g^{dec}_{f_{sig}}(x,a_1,b_1),a_2,b_2\right)$	$\theta=\begin{Bmatrix}a_1,b_1,\\ a_2,b_2\end{Bmatrix}$
sum of two sigmoids	$t\cdot g^{inc}_{f_{sig}}(x,a_1,b_1)+(1-t)\cdot g^{inc}_{f_{sig}}(x,a_2,b_2)$	$t\cdot g^{dec}_{f_{sig}}(x,a_1,b_1)+(1-t)\cdot g^{dec}_{f_{sig}}(x,a_2,b_2)$	$\theta=\begin{Bmatrix}a_1,b_1,\\ a_2,b_2,t\end{Bmatrix}$

and S_{norm}. Depending on the definition, the T_{norm} and S_{norm} might need parameters or not (also shown in Table II).

A fuzzy model consists of m fuzzy rules $R_i\left(d^l,\theta_i\right)$ $(i=1,\ldots,m)$, where θ_i is a vector containing all parameters used in the ith rule. Again, each R_i consists of several simple rules $r_{i,j}\left(d^l,\theta_{i,j}\right)$ $(j=1,\ldots,\tau)$, defined as (the jth simple rule in the ith rule): $r_{i,j}\left(d^l,\theta_{i,1}\right)=g_f\left(f_{i,j}(d),\theta_{i,j}\right)$, where $f_{i,j}(\cdot)$ is the same function of the data d^l as in the crisp model and $\theta_{i,j}$ is a vector of parameters. Each rule $R_i\left(d^l,\theta_i\right)$ is again formed as a combination of $r_{i,j}\left(d^l,\theta_{i,j}\right)$ simple rules $(j=1,\ldots,\tau)$, as follows [see (4), shown at the bottom of the page], where $\theta_i=\{\theta_{i,j},\kappa_{i,k},\nu_i\}$, $j=1,\ldots,\tau$, $k=1,\ldots,\upsilon$, with each $\theta_{i,j}$ being a vector with parameters used in the membership function of the jth simple rule of the ith rule, each $\kappa_{i,k}$ being a parameter entering the T_{norm} of the kth raw of the ith rule (with υ being the total number of rows) and ν_i being a parameter (one for each rule) entering the S_{norm}. If the T_{norm} and S_{norm} does not need parameters, then $\theta_i=\{\theta_{i,j}\}$, $j=1,\ldots,\tau$. The final decision (class) of the fuzzy model (M_f) is made using the result of all rules: $M_f\left(d^l,\Theta\right)=F_f\left(R_1\left(d^l,\theta_1\right),R_2\left(d^l,\theta_2\right),\ldots,R_i\left(d^l,\theta_m\right)\right)$, where Θ is a vector containing all parameters used in the rule $(\Theta=\{\theta_1,\theta_2,\ldots,\theta_m\})$ and $F_f(\cdot)$ is a function that combines the outcomes of all $R_i\left(d^l,\theta_i\right)$ fuzzy rules (defuzzifier). Again, the definition of the $F_f(\cdot)$ function can include the "unclassified" state. Depending on the representation selected for c^l, the final decision of the model could be $F_f\in\{1,2,\ldots,m\}$ or $F_f\in\{0,1\}^m$, where, if d^l is classified to class i, then $F_f=e_i$. Also, if unclassified state is included, then the final decision of the model could be $F_f\in\{1,2,\ldots,m,m+1\}$ or $F_f\in\{0,1\}^{m+1}$.

The transformation of the crisp set of rules to the respective fuzzy greatly depends on the selection of the fuzzy membership function, the T_{norm} and S_{norm} and the defuzzifier; if specific combinations among these are selected then known solutions from the literature can be used to express the explicit mathematical input-output of the fuzzy model [2], [11].

C. Optimization

The parameters Θ entering a fuzzy model can be optimally determined using an optimization procedure. Formulating the training process of a model as an optimization problem is a common practice in order to construct efficient expert systems. The efficiency of the system highly depends on the quality of the cost function and the choice of a training dataset. Also, a robust optimization method increases the speed of training process and enhances the quality of the final solution. The selection of the optimization method greatly depends on the equations describing the fuzzy model and the selection of the cost function; if these are differentiable then an optimization method making use of the first derivatives information can be employed, else methods that do not require first derivatives must be used (e.g., [47]).

The optimization problem can be formulated as: minimize function $f(x)$ subject to $x\in\Omega$, where $f(\cdot):\Re^n\rightarrow\Re$ and $\Omega\subseteq\Re^n$. It can be viewed as a decision problem which involves the computation of the "best" vector x of the decision variables over all possible vectors in Ω. This vector is called the minimizer of $f(\cdot)$ over Ω. Considering the optimization problem, two kinds of minimizers can be distinguished, local and global minimizers. A point $x^*\in\Omega$ is a local minimizer of $f(\cdot)$ over Ω if there exists $\varepsilon>0$ such as $f(x)\ge f(x^*)$ for all $x\in\Omega/\{x^*\}$ and $\|x-x^*\|<\varepsilon$. A point $x^*\in\Omega$ is a global minimizer of $f(\cdot)$ over Ω if $f(x)\ge f(x^*)$ for all $x\in\Omega/\{x^*\}$ [12]. Finding global minimizers is a challenging task and several techniques have been proposed: Branch and Bound techniques [13], simulated annealing [6], [14], genetic algorithms [5], [7], and stochastic methods.

$$
\begin{aligned}
R_i\left(d^l,\theta_i\right)=S_{\text{norm}}\Bigg(\Bigg(&T_{\text{norm}}\left(r_{i,1}\left(d^l,\theta_{i,1}\right),r_{i,2}\left(d^l,\theta_{i,2}\right),\ldots,r_{i,a_1}\left(d^l,\theta_{i,a_1}\right),\kappa_{i,1}\right)\\
&T_{\text{norm}}\left(r_{i,a_1+1}\left(d^l,\theta_{i,a_1+1}\right),r_{i,a_1+2}\left(d^l,\theta_{i,a_1+2}\right),\ldots,r_{i,a_2}\left(d^l,\theta_{i,a_2}\right),\kappa_{i,2}\right)\\
&\cdots\\
&T_{\text{norm}}\left(r_{i,a_l+1}\left(d^l,\theta_{i,a_l+1}\right),r_{i,a_l+2}\left(d^l,\theta_{i,a_l+2}\right),\ldots,r_{i,\tau}\left(d^l,\theta_{i,\tau}\right),\kappa_{i,\upsilon}\right)\Bigg),\nu_i\Bigg)
\end{aligned}
\tag{4}
$$

TABLE II
DEFINITIONS OF T_{norm} AND S_{norm}

	T_{norm}	S_{norm}	parameters
Minimum & maximum	$min(a,b)$	$max(a,b)$	
Algebraic product & probabilistic OR	ab	$a+b-ab$	
Einstein product and sum	$\frac{ab}{2-(a+b-ab)}$	$\frac{a+b}{1+ab}$	
Dombi class	$\frac{1}{1+\left(\left(\frac{1}{a}-1\right)^{\kappa}+\left(\frac{1}{b}-1\right)^{\kappa}\right)^{\frac{1}{\kappa}}}$	$\frac{1}{1+\left(\left(\frac{1}{a}-1\right)^{-\nu}+\left(\frac{1}{b}-1\right)^{-\nu}\right)^{-\frac{1}{\nu}}}$	$\kappa,\nu\in(0,+\infty)$
Dubois-Prade class	$\frac{ab}{max(a,b,\kappa)}$	$\frac{a+b-ab-min(a,b,1-\nu)}{max(1-a,1-b,\nu)}$	$\kappa,\nu\in[0,1]$
Yager class	$1-min\left(1,\left((1-a)^{\kappa}+(1-b)^{\kappa}\right)^{\frac{1}{\kappa}}\right)$	$min\left(1,\left(a^{\nu}+b^{\nu}\right)^{\frac{1}{\nu}}\right)$	$\kappa,\nu\in(0,+\infty)$

In the case of a fuzzy model, a cost function must be defined over a dataset $D=\{d^l,c^l\}$, $l=1,\ldots,K$. The minimization of this function leads to an improved model, in terms of its classification ability. A common cost function is the mean square error (MSE) function, which is defined as

$$\text{MSE}(D,\Theta)=\frac{1}{mK}\sum_{l=1}^{K}\left\|M_f\left(d^l,\Theta\right)-c^l\right\|_2^2 \qquad (5)$$

where $c^l\in\{0,1\}^m(c^l=e_i)$; therefore, $M\left(d^l,\Theta\right)\in\{0,1\}^m$. A second approach is to use the trace of the normalized confusion matrix [confusion matrix error (CME)]

$$\text{CME}(D,\Theta)=\frac{1}{m}\text{trace}\,(X)-p \qquad (6)$$

where p is a penalty term, which can be a function of the unclassified rates of each classification category. In this case, $c^l\in\{1,2,\ldots,m\}$ and $M\left(d^l,\Theta\right)\in\{1,2,\ldots,m,m+1\}$. If $M\left(d^l,\Theta\right)\in\{1,2,\ldots,m\}$, then the normalized confusion matrix is defined as $X_{M(d^l,\Theta),c^l}=$ number of patterns in c^l classified to $M\left(d^l,\Theta\right)/$total # of patterns in c^l, and if $M\left(d^l,\Theta\right)=m+1$ the unclassified ratio for each classification category is defined as $\text{uncl}_{c^l}=$ number of patterns in c^l unclassified/total number of patterns in c^l. In this case, the penalty term p can be defined as $p=\sum_{i=i}^{m}\text{uncl}_i/m$.

Fig. 1 presents a flowchart of the above described methodology; using a hypothetical initial set of crisp rules, the three stages of the methodology (crisp model, fuzzy model, and optimization) are shown.

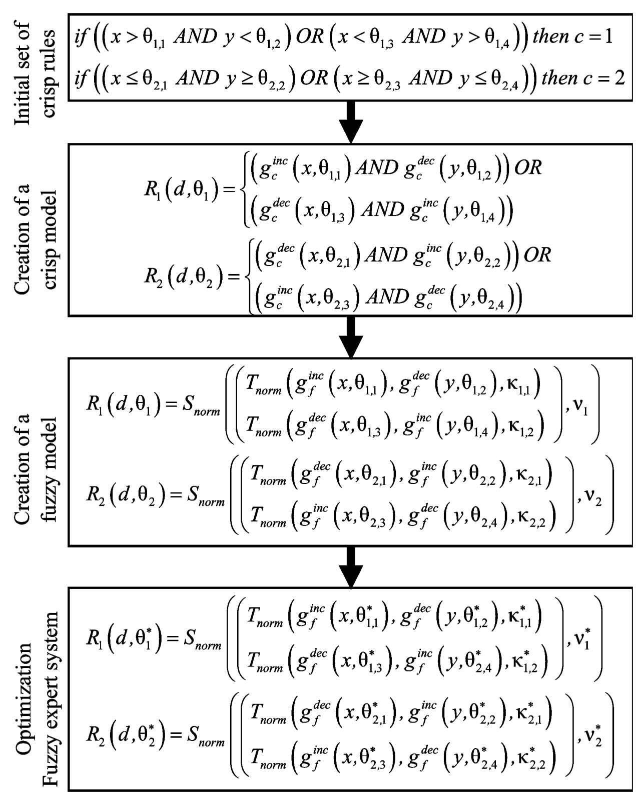


Fig. 1. Flowchart of the proposed methodology and its application on a hypothetical initial set of crisp rules.

III. APPLICATION TO CARDIOVASCULAR DISEASES

The above described framework was applied to two well-known classification problems from the cardiovascular domain, the arrhythmic beat classification and the ischemic beat classification from electrocardiograms (ECGs). For both cases, medical experts determined the initial set of rules, while well-known benchmark databases were used for the creation of the expert systems and their evaluation.

A. Medical Background

1) Arrhythmic Beat Classification: Arrhythmia can be defined as any type of rhythm that deviates from the normal sinus

TABLE III
DESCRIPTION OF DATASETS

Arrhythmic beat classification				Ischemic beat classification			
Rhythm class	Number of beats in D^{arh}	Number of beats in D^{arh}_{train}	Number of beats in D^{arh}_{test}	Beat class	Number of beats in D^{arh}	Number of beats D^{arh}_{train}	Number of beats in D^{arh}_{test}
Ventricular flutter/fibrillation (VF)	484	250	234	Ischemic	37,663	3,766	33,897
Premature ventricular contractions (PVC)	6,183	1,000	5,183				
Normal (N)	102,793	10,000	92,793	Normal	39,326	3,932	35,394
2° heart block (BII)	420	250	170				
Total	**109,880**	**11,500**	**98,380**	**Total**	**76,989**	**7,698**	**69,291**

rhythm. An arrhythmia can be either a single or a group of heartbeats, and it can affect the heart rate causing slow, fast or irregular rhythms [15]. Arrhythmias can take place in a healthy heart and be of minimal consequence but they may also indicate serious cardiac problems [16], [17]. Therefore, automatic arrhythmic beat detection and classification, using the ECG and/or features extracted from it, is a critical task in clinical cardiology, especially when performed in real time. In the later, each beat is classified into several different rhythm types. The techniques for beat classification are based on artificial neural networks [18], [19], "mixture of experts approach" [20], hermite functions combined with self-organizing maps [21], fuzzy neural networks [22], AR models [23], artificial neural networks and fuzzy equivalence data [24], support vector machines [25], ECG morphology and linear discriminates [26], time-frequency analysis combined with knowledge-based systems [27], and rule-based systems [28].

2) Ischemic Beat Classification: Myocardial ischemia is the condition where oxygen deprivation to the heart muscle is accompanied by inadequate removal of metabolites due to reduced blood flow or perfusion. This reduced blood supply to the myocardium causes alterations in the ECG signal, such as deviations in the ST segment and changes in the T wave [29]. The accurate ischemic episode detection, where a sequence of cardiac beats is assessed [30], is based on the correct detection of ischemic beats [31]–[33]. Several techniques that evaluate the ST segment changes and the T-wave alterations have been proposed for ischemic beat detection. More specifically, the use of approaches like parametric modeling [34], wavelet theory [35], set of rules [36], [37], artificial neural networks [30], [38], multicriteria decision analysis and genetic algorithms [7] have been previously reported.

B. Datasets

1) Arrhythmia Dataset: All the records from the MIT-BIH arrhythmia database [42] were used for the training and the evaluation of the arrhythmic beat classification FES. Initially, the RR-interval signal was extracted from the ECG recordings using QRS detection [43], [44], except in the case of VF episodes in record 207, where the actual beats from the annotation of the database were used. Then, windows of three consecutive RR intervals $[RR_{l-1}, RR_l, RR_{l+1}]$, where RR_l is the lth RR interval in the RR interval signal, were defined and both rhythm and beat annotations (defined in the database) were used to specify the class c^l of each window, as follows: if the middle beat of the window (RR_l) belongs to 2 ° heart block episode (rhythm annotation BII in the database), ventricular flutter/fibrillation wave (beat annotations [, !,], respectively, in the database) or it is annotated as premature ventricular contraction (beat annotation V in the database) then $c^l = 4$ or $c^l = 1$ or $c^l = 2$, respectively. Everything else was considered as normal sinus rhythm ($c^l = 3$). Therefore, the dataset was defined as: $D^{\rm arh} = \{d^l, c^l\}$, $l = 1, \ldots, K$, with $d^l = [RR_{l-1}, RR_l, RR_{l+1}]$ being a single pattern with three features, c^l its class with four different classes, and K is the number of patterns (beats in the dataset). The class c^l can be represented either as $c^l = \{1, 2, 3, 4\}$ or $c^l \in \{0, 1\}^4$, where, if d^l belongs to class $i = \{1, 2, 3, 4\}$, then $c^l = e_i$, i.e., $c^l = \{[1, 0, 0, 0], [0, 1, 0, 0], [0, 0, 1, 0], [0, 0, 0, 1]\}$. The cardiac rhythm categories and the number of beats used in each cardiac rhythm category, are shown in Table III.

2) Ischemia Dataset: The European Society of Cardiology (ESC) ST-T database [45] was used for the training and the evaluation of the ischemic beat classification FES; 11 h of two-channel ECG recordings were selected. Those, contain the first hour of the e0103, e0105, e0108, e0113, e0114, e0147, e0159, e0162, and e0206 recordings and the whole e0104 recording. These ten recordings were selected because their ischaemic ECG beats are characterized by significant waveform variability. First, the preprocessing of the recorded ECG signal was performed (for both channels) in order to eliminate noise distortions (e.g., baseline wandering, A/C interference and electromyographic contamination) [7] and locate the isoelectric line and the J point [46]. Then, the following features were extracted from each cardiac beat. i) The ST segment deviation ($ST^{\rm dev}$), which is the amplitude deviation of the ST segment from the isoelectric line. The ST segment changes were measured either 80 ms after the J point (J80) (heart rate ≤ 120 bpm), or 60 ms after the J point (J60) (heart rate > 120 bpm). ii) The ST segment slope ($ST^{\rm slope}$), which is the slope of the line connecting the J and J80 (or J60) points.

iii) The T-wave amplitude (T^{amp}), which is the amplitude deviation of the T-wave peak from the isoelectric line. iv) The T-wave normal amplitude together with its respected polarity (T^{norm}) which refer to the amplitude and polarity of normal beats for a specific ECG lead. It was calculated using the first 30 s of each recording and was computed by using the mean value of the T-wave amplitudes at this interval. In order to define the class c^l of each beat, three medical experts annotated independently each beat as normal, ischemic or artefact. In the case of disagreement, the decision was taken by consensus. After removing the artefacts and the misdetected beats the remaining were diagnosed as normal or ischaemic. Thus, the dataset was defined as: $D^{\mathrm{isch}} = \{d^l, c^l\}$ $l = 1, \ldots, K$, with $d^l = \left[ST_l^{\mathrm{dev}}, ST_l^{\mathrm{slope}}, T_l^{\mathrm{amp}}, T_l^{\mathrm{norm}}\right]$ being the lth feature vector (ST^{dev}, ST^{slope}, T^{amp} and T^{norm} of the lth beat), c^l the class of the beat (normal or ischemic), and K is the number of beats in the dataset. The class is represented either as $c^l = \{1, 2\}$ or $c^l \in \{0, 1\}^2$ i.e., $c^l = [0, 1]$ if the beat is normal and $c^l = [1, 0]$ if the beat is ischaemic. The ischemic beat categories and the number of beats in each category are also shown in Table III.

C. Initial Set of Rules

1) Arrhythmic Set of Rules: The three RR-intervals window was used to classify the middle RR interval (RR_l) into one of the four categories: 1) ventricular flutter/fibrillation (VF), 2) premature ventricular contraction (PVC), 3) normal sinus rhythm (N), and 4) 2 ° heart block (BII). Also, if the classification process fails, the middle RR interval was classified as (5) unclassified. Three rules were used for the classification (see the first equation shown at the bottom of the page). In the case that none of the three rules was true, then the RR_l interval was classified as $N(3)$ while is the case of more than one of the three rules was true the RR_l interval was unclassified (5).

2) Ischemic Set of Rules: In the case of ischemic beat classification, the rules used in [37] were employed. The feature vector $\left[ST_l^{\mathrm{dev}}, ST_l^{\mathrm{slope}}, T_l^{\mathrm{amp}}, T_l^{\mathrm{norm}}\right]$ was used to classify the beat as normal or ischemic (see the second equation shown at the bottom of the page).

D. Crisp Models

1) Arrhythmic Crisp Model: The arrhythmic beat classification crisp model (M_c^{arh}) includes three crisp rules

$$\begin{aligned}
\text{Rule 1 :}\quad & \text{If } ((RR_{l-1} < \theta_{1,1}) \text{ AND } (RR_l < \theta_{1,2}) \text{ AND } (RR_{l+1} < \theta_{1,3})) \text{ OR} \\
& (RR_{l-1} + RR_l + RR_{l+1} < \theta_{1,4}) \text{ then } RR_l \text{ is classified as } \mathrm{VF}(1). \\
\text{Rule 2 :}\quad & \text{If } \left(\left(\frac{RR_{l-1}}{RR_l} > \theta_{2,1}\right) \text{ AND } \left(\frac{RR_{l+1}}{RR_l} > \theta_{2,2}\right)\right) \text{ OR} \\
& \left(\left(\frac{RR_{l+1}}{RR_l} > \theta_{2,3}\right) \text{ AND } \left(\frac{RR_{l-1}}{RR_l} > \theta_{2,4}\right)\right) \text{ OR} \\
& \Bigg((|RR_{l-1} - RR_l| < \theta_{2,5}) \text{ AND } (RR_{l-1} < \theta_{2,6}) \text{ AND } (RR_l < \theta_{2,7}) \\
& \qquad \text{AND } \left(\frac{RR_{l-1} + RR_l}{2RR_{l+1}} < \theta_{2,8}\right)\Bigg) \text{ OR} \\
& \Bigg((|RR_l - RR_{l+1}| < \theta_{2,9}) \text{ AND } (RR_l < \theta_{2,10}) \text{ AND } (RR_{l+1} < \theta_{2,11}) \\
& \qquad \text{AND } \left(\frac{RR_l + RR_{l+1}}{2RR_{l-1}} < \theta_{2,12}\right)\Bigg) \\
& \text{then } RR_l \text{ is classified as } \mathrm{PVC}\ (2). \\
\text{Rule 3 :}\quad & \text{If } ((RR_l \in [\theta_{3,1}, \theta_{3,2}]) \text{ AND } (|RR_{l-1} - RR_l| < \theta_{3,3})) \text{ OR} \\
& ((RR_l \in [\theta_{3,4}, \theta_{3,5}]) \text{ AND } (|RR_l - RR_{l+1}| < \theta_{3,6})) \\
& \text{then } RR_l \text{ is classified as } \mathrm{BII}(4)
\end{aligned}$$

$$\begin{aligned}
\text{Rule 1 :}\quad & \text{If } \left(\left(ST_l^{\mathrm{dev}} < \theta_{1,1}\right) \text{ AND } \left(ST_l^{\mathrm{slope}} > \theta_{1,2}\right)\right) \text{ OR } \left(ST_l^{\mathrm{dev}} > \theta_{1,3}\right) \text{ OR } (T_l^{\mathrm{amp}} \cdot T_l^{\mathrm{norm}} < \theta_{1,4}) \text{ OR} \\
& \left(\frac{|T_l^{\mathrm{amp}}|}{|T_l^{\mathrm{norm}}|} < \theta_{1,6}\right) \text{ OR } \left(\left(\frac{T_l^{\mathrm{amp}}}{T_l^{\mathrm{norm}}} < \theta_{1,6}\right) \text{ AND } (||T_l^{\mathrm{amp}}| - |T_l^{\mathrm{norm}}|| > \theta_{1,7})\right) \\
& \text{then the beat is classified as ischemic } (1) \text{ else the beat is classified as normal } (2)
\end{aligned}$$

[see (7)–(9), shown at the bottom of the page], where $\theta_1 = \{\theta_{1,j}\}$, $j = 1,\ldots,4$, $\theta_2 = \{\theta_{2,j}\}$, $j = 1,\ldots,12$ and $\theta_3 = \{\theta_{3,j}\}$, $j = 1,\ldots,6$. The final decision of the M_c^{arh} was made using the results from all rules, i.e.,: $M_c^{\text{arh}}\left(d^l,\Theta\right) = F_c\left(R_1\left(d^l,\theta_1\right),R_2\left(d^l,\theta_2\right),R_3\left(d^l,\theta_3\right)\right)$, where Θ is a vector containing all thresholds used in the model ($\Theta = \{\theta_1,\theta_2,\theta_3\}$) and $F_c(\cdot)$ is a function that combines the outcomes of all crisp rules and its definition depends on the error function that was used. In the case of the CME function (5), see (10), shown at the bottom of the page.

In a similar way, if the MSE function was used (6), then $F_c(\cdot)$ was defined as (11), shown at the bottom of the page.

2) Ischemic Crisp Model: The ischemic beat classification crisp model (M_c^{isch}) includes one crisp rule [see (12), shown at the bottom of the next page], where $\theta_1 = \{\theta_{1,j}\}$, $j = 1,\ldots,7$. The final decision of the M_c^{isch} was made as: $M_c^{\text{isch}}\left(d^l,\Theta\right) = F_c\left(R_1\left(d^l,\theta_1\right)\right)$, where $\Theta = \{\theta_1\}$ and $F_c(\cdot)$, if the CME function (5) was used for optimization, was defined as

$$F_c\left(R_1\left(d^l,\theta_1\right)\right) = \begin{cases} 1, & \text{if } R_1\left(d^l,\theta_1\right) \text{ is true} \\ 2, & \text{if } R_1\left(d^l,\theta_1\right) \text{ is false} \end{cases} \tag{13}$$

while, if the MSE function (6) was used, then $F_c(\cdot)$ was defined as

$$F_c\left(R_1\left(d^l,\theta_1\right)\right) = \begin{cases} [1 \quad 0], & \text{if } R_1\left(d^l,\theta_1\right) \text{ is true} \\ [0 \quad 1], & \text{if } R_1\left(d^l,\theta_1\right) \text{ is false.} \end{cases} \tag{14}$$

$$\begin{aligned} &R_1\left(d^l,\theta_1\right) \\ &= \left(g_c^{\text{dec}}\left(RR_{l-1},\theta_{1,1}\right) \text{ AND } g_c^{\text{dec}}\left(RR_l,\theta_{1,2}\right) \text{ AND } g_c^{\text{dec}}\left(RR_{l+1}\theta_{1,3}\right)\right) \text{ OR } \left(g_c^{\text{dec}}\left(RR_{l-1}+RR_l+RR_{l+1},\theta_{1,4}\right)\right) \end{aligned} \tag{7}$$

$$\begin{aligned} &R_2\left(d^l,\theta_2\right) \\ &= \begin{cases} \left(g_c^{\text{inc}}\left(\frac{RR_{l-1}}{RR_l},\theta_{2,1}\right) \text{ AND } g_c^{\text{inc}}\left(\frac{RR_{l+1}}{RR_l},\theta_{2,2}\right)\right) & \text{OR} \\ \left(g_c^{\text{inc}}\left(\frac{RR_{l+1}}{RR_{l-1}},\theta_{2,3}\right) \text{ AND } g_c^{\text{inc}}\left(\frac{RR_{l-1}}{RR_l},\theta_{2,4}\right)\right) & \text{OR} \\ \left(g_c^{\text{dec}}\left(\left|RR_{l-1}-RR_l\right|,\theta_{2,5}\right) \text{ AND } g_c^{\text{dec}}\left(RR_{l-1},\theta_{2,6}\right) \text{ AND } g_c^{\text{dec}}\left(RR_l,\theta_{2,7}\right) \text{ AND } g_c^{\text{inc}}\left(\frac{RR_{l-1}+RR_l}{2RR_{l+1}},\theta_{2,8}\right)\right) & \text{OR} \\ \left(g_c^{\text{dec}}\left(\left|RR_l-RR_{l+1}\right|,\theta_{2,9}\right) \text{ AND } g_c^{\text{dec}}\left(RR_l,\theta_{2,10}\right) \text{ AND } g_c^{\text{dec}}\left(RR_{l+1},\theta_{2,11}\right) \text{ AND } g_c^{\text{inc}}\left(\frac{RR_l+RR_{l+1}}{2RR_{l-1}},\theta_{2,12}\right)\right) & \end{cases} \end{aligned} \tag{8}$$

$$\begin{aligned} &R_3\left(d^l,\theta_3\right) \\ &= \begin{cases} \left(g_c^{\text{inc}}\left(RR_l,\theta_{3,1}\right) \text{ AND } g_c^{\text{dec}}\left(RR_l,\theta_{3,2}\right) \text{ AND } g_c^{\text{dec}}\left(\left|RR_{l-1}-RR_l\right|,\theta_{3,3}\right)\right) \quad \text{OR} \\ \left(g_c^{\text{inc}}\left(RR_l,\theta_{3,4}\right) \text{ AND } g_c^{\text{dec}}\left(RR_l,\theta_{3,5}\right) \text{ AND } g_c^{\text{dec}}\left(\left|RR_{l+1}-RR_l\right|,\theta_{3,6}\right)\right) \end{cases} \end{aligned} \tag{9}$$

$$\begin{aligned} &F_c\left(R_1\left(d^l,\theta_1\right),R_2\left(d^l,\theta_2\right),R_3\left(d^l,\theta_3\right)\right) \\ &= \begin{cases} 1, & \text{if only } R_1\left(d^l,\theta_1\right) \text{ is true} \\ 2, & \text{if only } R_2\left(d^l,\theta_2\right) \text{ is true} \\ 3, & \text{if all } R_1\left(d^l,\theta_1\right), R_2\left(d^l,\theta_2\right) \text{ and } R_3\left(d^l,\theta_3\right) \text{ are false} \\ 4, & \text{if only } R_3\left(d^l,\theta_3\right) \text{ is true} \\ 5, & \text{if more than one of the } R_1\left(d^l,\theta_1\right), R_2\left(d^l,\theta_2\right) \text{ or } R_3\left(d^l,\theta_3\right) \text{ are true} \end{cases} \end{aligned} \tag{10}$$

$$\begin{aligned} &F_c\left(R_1\left(d^l,\theta_1\right),R_2\left(d^l,\theta_2\right),R_3\left(d^l,\theta_3\right)\right) \\ &= \begin{cases} e_1, & \text{if only } R_1\left(d^l,\theta_1\right) \text{ is true} \\ e_2, & \text{if only } R_2\left(d^l,\theta_2\right) \text{ is true} \\ e_3, & \text{if all } R_1\left(d^l,\theta_1\right), R_2\left(d^l,\theta_2\right) \text{ and } R_3\left(d^l,\theta_3\right) \text{ are false} \\ e_4, & \text{if only } R_3\left(d^l,\theta_3\right) \text{ is true} \\ e_1+e_2, & \text{if } R_1\left(d^l,\theta_1\right) \text{ and } R_2\left(d^l,\theta_2\right) \text{ are true} \\ e_2+e_4, & \text{if } R_2\left(d^l,\theta_2\right) \text{ and } R_3\left(d^l,\theta_3\right) \text{ are true} \\ e_1+e_4, & \text{if } R_1\left(d^l,\theta_1\right) \text{ and } R_3\left(d^l,\theta_3\right) \text{ are true} \\ e_1+e_2+e_4, & \text{if all } R_1\left(d^l,\theta_1\right) R_2\left(d^l,\theta_2\right) \text{ and } R_3\left(d^l,\theta_3\right) \text{ are true} \end{cases} \end{aligned} \tag{11}$$

E. Fuzzy Models

1) Arrhythmic Fuzzy Models: Several fuzzy models were developed, depending on the selection of the fuzzy membership function and the T_{norm} and S_{norm} definitions. Each fuzzy model ($M^{\mathrm{arh}}_{f,TS}$), where f was the fuzzy membership function and TS were the T_{norm} and S_{norm} definitions had three fuzzy rules, defined as (15)–(17), shown at the bottom of the page, where $\theta_1 = \{\theta_{1,j}, \kappa_{1,1}, \nu_1\}$, $j = 1,\ldots,4$, $\theta_2 = \{\theta_{2,j}, \kappa_{2,k}, \nu_2\}$, $j = 1,\ldots,12$, $k = 1,\ldots,4$ and $\theta_3 = \{\theta_{3,j}, \kappa_{3,k}, \nu_3\}$, $j = 1,\ldots,4$, $k = 1, 2$ (if the T_{norm} and S_{norm} do not need parameters, then $\theta_1 = \{\theta_{1,j}\}$, $j = 1,\ldots,4$, $\theta_2 = \{\theta_{2,j}\}$, $j = 1,\ldots,12$ and $\theta_3 = \{\theta_{3,j}\}$, $j = 1,\ldots,4$). The final decision for each $M^{\mathrm{arh}}_{f,TS}$ was made combining the results of all fuzzy rules: $M^{\mathrm{arh}}_{f,TS}(d^l, \Theta) = F_f\left(R_1(d^l,\theta_1), R_2(d^l,\theta_2), R_3(d^l,\theta_3)\right)$, where Θ is a vector containing all parameters used in the model ($\Theta = \{\theta_1,\theta_2,\theta_3,\theta_4\}$; θ_4 is a parameter defined below) and $F_f(\cdot)$ is the defuzzification function, which combines the outcomes of all $R_i(d^l,\theta_i)$ fuzzy rules and its definition depends on the error function used. In the case of CME, $F_f(\cdot)$ was defined as (18), shown at the bottom of the next page.

The defuzzification function is problem-specific and it is designed so as to reflect the expert's knowledge on this specific domain. Each d^l was considered *a priori* normal sinus rhythm (category 3). Therefore, if the maximum value of the results of the three rules is $\leq \theta_4$, then d^l was classified as normal sinus rhythm (θ_4 is a parameter). If the maximum value of the results of the three rules was $> \theta_4$, then d^l was classified in the category of the rule that had the maximum result, i.e., in category 1 if the $R_1(d^l,\theta_1)$ was the maximum, category 2 if $R_2(d^l,\theta_2)$ was the maximum and category 4 if $R_3(d^l,\theta_3)$ was the maximum. Finally, if the maximum value of the results of the three rules was $> \theta_4$ but two or more of the rules had the maximum value, then d^l was classified as category 5 (unclassified). If MSE was used, $F_f(\cdot)$ was defined as

$$F_f\left(R_1(d^l,\theta_1), R_2(d^l,\theta_2), R_3(d^l,\theta_3), \theta_4\right) = \left[R_1(d^l,\theta_1), R_2(d^l,\theta_2), \theta_4, R_3(d^l,\theta_3)\right]. \quad (19)$$

$$R_1(d^l,\theta_1) = \begin{cases} \left(g_c^{\mathrm{dec}}\left(ST_l^{\mathrm{dev}},\theta_{1,1}\right) \text{ AND } g_c^{\mathrm{inc}}\left(ST_l^{\mathrm{slope}},\theta_{1,2}\right)\right) & \text{OR } \left(g_c^{\mathrm{inc}}\left(ST_l^{\mathrm{dev}},\theta_{1,3}\right)\right) \text{ OR} \\ \left(g_c^{\mathrm{dec}}\left(T_l^{\mathrm{amp}}\cdot T_l^{\mathrm{norm}},\theta_{1,4}\right)\right) & \text{OR } \left(g_c^{\mathrm{dec}}\left(\frac{|T_l^{\mathrm{amp}}|}{|T_l^{\mathrm{norm}}|},\theta_{1,5}\right)\right) \text{ OR} \\ \left(g_c^{\mathrm{dec}}\left(\frac{T_l^{\mathrm{amp}}}{T_l^{\mathrm{norm}}},\theta_{1,6}\right) \text{ AND } g_c^{\mathrm{inc}}\left(\left||T_l^{\mathrm{amp}}| - |T_l^{\mathrm{norm}}|\right|,\theta_{1,7}\right)\right) & \end{cases} \quad (12)$$

$$R_1(d^l,\theta_1) = S_{\mathrm{norm}}\left(\left(T_{\mathrm{norm}}\left(g_f^{\mathrm{dec}}(RR_{l-1},\theta_{1,1}), g_f^{\mathrm{dec}}(RR_l,\theta_{1,2}), g_f^{\mathrm{dec}}(RR_{l+1},\theta_{1,3}), \kappa_{1,1}\right) g_f^{\mathrm{dec}}(RR_{l-1}+RR_l+RR_{l+1},\theta_{1,4})\right), \nu_1\right) \quad (15)$$

$$\begin{aligned} R_2(d^l,\theta_2) = S_{\mathrm{norm}}\Bigg(\Bigg(&T_{\mathrm{norm}}\left(g_f^{\mathrm{inc}}\left(\frac{RR_{l-1}}{RR_l},\theta_{2,1}\right), g_f^{\mathrm{inc}}\left(\frac{RR_{l+1}}{RR_l},\theta_{2,2}\right), \kappa_{2,1}\right) \\ &T_{\mathrm{norm}}\left(g_f^{\mathrm{inc}}\left(\frac{RR_{l+1}}{RR_{l-1}},\theta_{2,3}\right), g_f^{\mathrm{inc}}\left(\frac{RR_{l-1}}{RR_l},\theta_{2,4}\right), \kappa_{2,2}\right) \\ &T_{\mathrm{norm}}\left(g_f^{\mathrm{dec}}(|RR_{l-1}-RR_l|,\theta_{2,5}), g_f^{\mathrm{dec}}(RR_{l-1},\theta_{2,6}), g_f^{\mathrm{dec}}(RR_l,\theta_{2,7}), g_f^{\mathrm{dec}}\left(\frac{RR_{l-1}+RR_l}{2RR_{l+1}},\theta_{2,8}\right), \kappa_{2,3}\right) \\ &T_{\mathrm{norm}}\left(g_f^{\mathrm{dec}}(|RR_l-RR_{l+1}|,\theta_{2,9}), g_f^{\mathrm{dec}}(RR_l,\theta_{2,10}), g_f^{\mathrm{dec}}(RR_{l+1},\theta_{2,11}), g_f^{\mathrm{dec}}\left(\frac{RR_l+RR_{l+1}}{2RR_{l-1}},\theta_{2,12}\right), \kappa_{2,4}\right), \nu_2\Bigg) \end{aligned} \quad (16)$$

$$\begin{aligned} R_3(d^l,\theta_3) = S_{\mathrm{norm}}\Bigg(\Bigg(&T_{\mathrm{norm}}\left(g_f^{\mathrm{inc}}(RR_l,\theta_{3,1}), g_f^{\mathrm{dec}}(RR_l,\theta_{3,2}), g_f^{\mathrm{dec}}(|RR_{l-1}-RR_l|,\theta_{3,3}), \kappa_{3,1}\right) \\ &T_{\mathrm{norm}}\left(g_f^{\mathrm{inc}}(RR_l,\theta_{3,4}), g_f^{\mathrm{dec}}(RR_l,\theta_{3,5}), g_f^{\mathrm{dec}}(|RR_{l+1}-RR_l|,\theta_{3,6}), \kappa_{3,2}\right)\Bigg)\Bigg), \nu_3\Bigg) \end{aligned} \quad (17)$$

2) Ischaemic Fuzzy Models: Again, several fuzzy models were developed, depending on the selection of the fuzzy membership function and the $T_{\rm norm}$ and $S_{\rm norm}$ definitions. Each fuzzy model $(M_{f,TS}^{\rm ish})$ included a single fuzzy rule [see (20), shown at the bottom of the page], where $\theta_1 = \{\theta_{1,j}, \kappa_{1,k}, \nu_1\}$, $j = 1,\ldots,7$, $k = 1, 2$ (if the $T_{\rm norm}$ and $S_{\rm norm}$ does not need parameters, then $\theta_1 = \{\theta_{1,j}\}$, $j = 1,\ldots,7$). Again, the final decision for each fuzzy model was made using a problem-specific defuzzification function: $M_{f,TS}^{\rm ish}(d^l, \Theta) = F_f(R_1(d^l,\theta_1),\theta_2)$, where $\Theta = \{\theta_1, \theta_2\}$ (θ_2 is defined below) and $F_f(\cdot)$ was defined as

$$F_f\left(R_1\left(d^l,\theta_1\right),\theta_2\right) = \begin{cases} 1, & \text{if } R_1\left(d^l,\theta_1\right) > \theta_2 \\ 2, & \text{else} \end{cases} \qquad (21)$$

when the CME function was used, while in the case of using the MSE function, then $F_f(\cdot)$ was defined as

$$F_f\left(R_1\left(d^l,\theta_1\right),\theta_2\right) = \left[R_1\left(d^l,\theta_1\right),\theta_2\right]. \qquad (22)$$

Detailed versions of the equations of the fuzzy rules for both arrhythmic and ischemic fuzzy models, for specific selections of fuzzy membership functions and $T_{\rm norm}$ and $S_{\rm norm}$ definitions are presented in the Appendix.

F. Expert Systems

Once a fuzzy model were created, the parameters entering the model (Θ) must be identified; thus, a cost function was minimized. It should be mentioned that the number of parameters entering each fuzzy model differs significantly, depending on the fuzzy membership function selection and the $T_{\rm norm}$ and $S_{\rm norm}$ definitions. Both mean square error and confusion matrix cost functions, defined in (5) and (6), respectively, were tested. To perform the optimization a training dataset $(D_{\rm train})$ was needed, which was a randomly selected subset of $D^{\rm arh}$ or $D^{\rm isch}$, depending on the problem. In the case of arrhythmic beat classification, the $D_{\rm train}^{\rm arh}$ contained 250 patterns from classes VFL and BII, 1000 patterns from class PVC and 10 000 patterns from class N. Thus, the size of the $D_{\rm train}$ dataset for the arrhythmic beat classification was 11 500 beats. Appropriate weights were used for each class so as there would be no bias for larger classes (i.e., each VFL or BII pattern entered the optimization procedure 40 times and each PVC pattern ten times). In the case of ischemic beat classification the $D_{\rm train}^{\rm isch}$ contained 3766 normal and 3932 ischemic beats, and, thus, its size was 7698 beats (the training set was constructed by selecting iteratively the first beat out of a sequence of ten beats). $D_{\rm train}^{\rm arh}$ and $D_{\rm train}^{\rm isch}$ are shown in Table III.

The optimization method that it was used is a modification of controlled random search (MCRS) [47]. The MCRS is inspired from simplex method for local optimization, because of the irregular simplex comprised from $N+1$ points, which is maintained in each iteration of the method. In the main step of the algorithm, the simplex's points are used to obtain a trial point which, under certain conditions, will replace the previous best from the simplex. Note that if more than one global minima exist, the method will locate only one of them. The MCRS method is described in the Appendix. Given a specific fuzzy model (e.g., $M_{sig,\min\max}^{\rm arh}$ where the sigmoid function was selected as fuzzy membership function and the min-max definition for the $T_{\rm norm}$ and $S_{\rm norm}$), a cost function (e.g., the mean square error), a training dataset $(D_{\rm train})$ and a range where the model's parameters were constrained, the MCRS algorithm was applied for a specified number of iterations or until a stopping criterion was met (see Appendix), and it attempted to optimize the value of the cost function with respect to the parameters Θ entering the fuzzy model (e.g., the parameters used for the fuzzy membership functions and the parameters of the $T_{\rm norm}$ and $S_{\rm norm}$, if any). The FES was formulated by setting the parameters of each model to the best solution found.

$$\begin{aligned} &F_f\left(R_1\left(d^l,\theta_1\right),R_2\left(d^l,\theta_2\right),R_3\left(d^l,\theta_3\right),\theta_4\right) \\ &= \begin{cases} 1, & \text{if } \max\left(R_1\left(d^l,\theta_1\right),R_2\left(d^l,\theta_2\right),R_3\left(d^l,\theta_3\right)\right) = R_1\left(d^l,\theta_1\right) > \theta_4 \\ 2, & \text{if } \max\left(R_1\left(d^l,\theta_1\right),R_2\left(d^l,\theta_2\right),R_3\left(d^l,\theta_3\right)\right) = R_2\left(d^l,\theta_2\right) > \theta_4 \\ 3, & \text{if } \max\left(R_1\left(d^l,\theta_1\right),R_2\left(d^l,\theta_2\right),R_3\left(d^l,\theta_3\right)\right) \le \theta_4 \\ 4, & \text{if } \max\left(R_1\left(d^l,\theta_1\right),R_2\left(d^l,\theta_2\right),R_3\left(d^l,\theta_3\right)\right) = R_3\left(d^l,\theta_3\right) > \theta_4 \\ 5, & \text{if } \max\left(R_1\left(d^l,\theta_1\right),R_2\left(d^l,\theta_2\right),R_3\left(d^l,\theta_3\right)\right) > \theta_4 \\ & \quad \text{and two or more of the } R_1\left(d^l,\theta_1\right), R_2\left(d^l,\theta_2\right) \text{ or } R_3\left(d^l,\theta_3\right) \text{ are equal} \end{cases} \end{aligned} \qquad (18)$$

$$\begin{aligned} R_1\left(d^l,\theta_1\right) = S_{\rm norm}\Bigg(\Bigg(&T_{\rm norm}\left(g_f^{\rm dec}\left(ST_l^{\rm dev},\theta_{1,1}\right), g_f^{\rm inc}\left(ST_l^{\rm slope},\theta_{1,2}\right),\kappa_{1,1}\right) \\ &g_f^{\rm inc}\left(ST_l^{\rm dev},\theta_{1,3}\right), g_f^{\rm dec}\left(T_l^{\rm amp}\cdot T_l^{\rm norm},\theta_{1,4}\right), g_f^{\rm dec}\left(\frac{|T_l^{\rm amp}|}{|T_l^{\rm norm}|},\theta_{1,5}\right) \\ &T_{\rm norm}\left(g_f^{\rm dec}\left(\frac{T_l^{\rm amp}}{T_l^{\rm norm}},\theta_{1,6}\right), g_f^{\rm inc}\left(\left||T_l^{\rm amp}|-|T_l^{\rm norm}|\right|,\theta_{1,7}\right),\kappa_{1,2}\right)\Bigg),\nu_1\Bigg) \end{aligned} \qquad (20)$$

TABLE IV
SENSITIVITY (%), SPECIFICITY (%) AND POSITIVE PREDICTIVE VALUE (%) OF THE CRISP MODEL AND THE AVERAGE CONFUSION MATRICES FOR ALL FUZZY EXPERT SYSTEMS FOR THE ARRHYTHMIC BEAT CLASSIFICATION PROBLEM, USING THE CME COST FUNCTION

	Crisp			
	VF	PVC	N	BII
Se	90.00	82.96	93.00	98.29
Sp	99.49	97.21	94.09	99.91
PPV	98.33	91.06	84.38	99.74

		Linear				Sigmoid				Sum of sigmoid and its gradient				Nested sigmoid				Sum of two sigmoids			
		VF	PVC	N	BII	VF	PVC	N	BII	VF	PVC	N	BII	VF	PVC	N	BII	VF	PVC	N	BII
Minimum & maximum	Se	94.04	89.91	92.89	98.12	97.56	90.99	93.00	98.18	98.29	90.68	92.71	97.94	97.74	90.30	93.08	98.24	97.48	90.68	92.84	98.35
	Sp	99.81	97.19	96.15	99.91	99.81	97.12	96.40	99.91	99.76	97.17	96.38	99.90	99.79	97.16	96.26	99.90	99.81	97.25	96.17	99.89
	PPV	99.41	91.53	89.09	99.74	99.43	91.33	89.60	99.72	99.27	91.43	89.50	99.71	99.37	91.38	89.24	99.70	99.41	91.65	89.00	99.67
Algebraic product & probabilistic OR	Se	95.38	89.90	92.95	98.47	97.74	91.50	93.23	98.35	98.42	91.18	93.38	98.06	97.82	91.07	93.65	98.41	97.61	91.02	93.10	98.47
	Sp	99.81	97.32	96.26	99.92	99.81	97.26	96.62	99.91	99.76	97.43	96.57	99.91	99.80	97.38	96.57	99.91	99.82	97.35	96.33	99.90
	PPV	99.41	91.87	89.34	99.76	99.43	91.75	90.20	99.73	99.29	92.20	90.08	99.73	99.40	92.04	90.10	99.71	99.46	91.98	89.41	99.69
Einstein product and sum	Se	95.21	90.00	92.97	98.47	97.74	91.10	93.23	98.35	98.50	91.71	93.30	98.24	97.91	90.31	93.25	98.41	97.65	90.76	93.41	98.53
	Sp	99.82	97.28	96.26	99.93	99.82	97.25	96.49	99.91	99.78	97.44	96.79	99.91	99.80	97.26	96.33	99.91	99.82	97.47	96.26	99.90
	PPV	99.45	91.77	89.34	99.78	99.45	91.70	89.84	99.74	99.34	92.26	90.64	99.72	99.39	91.66	89.43	99.72	99.44	92.29	89.28	99.70
Dombi class	Se	95.94	90.13	93.04	98.71	98.29	93.17	95.12	98.71	98.21	91.61	93.65	99.00	98.29	93.00	93.85	98.59	97.86	92.25	93.36	98.65
	Sp	99.82	97.41	96.37	99.93	99.86	97.90	97.41	99.92	99.85	97.53	96.85	99.92	99.83	97.65	97.17	99.92	99.82	97.49	96.82	99.90
	PPV	99.45	92.14	89.61	99.79	99.59	93.67	92.45	99.75	99.54	92.53	90.84	99.76	99.49	92.95	91.71	99.77	99.46	92.46	90.74	99.71
Dubois-Prade class	Se	96.97	90.32	93.08	99.00	98.12	92.81	94.78	98.53	98.59	92.58	93.49	98.18	98.21	92.83	95.97	98.71	97.82	92.21	93.17	98.59
	Sp	99.82	97.58	96.59	99.93	99.87	97.74	97.22	99.92	99.79	97.49	97.07	99.92	99.82	98.24	97.24	99.94	99.82	97.41	96.79	99.90
	PPV	99.46	92.61	90.18	99.79	99.61	93.19	91.90	99.75	99.36	92.48	91.42	99.76	99.46	94.63	92.05	99.80	99.46	92.24	90.64	99.70
Yager class	Se	96.07	90.15	93.05	98.71	98.08	94.28	93.54	98.71	98.63	92.40	93.00	98.59	97.91	90.72	93.82	98.53	97.74	91.35	93.89	98.65
	Sp	99.82	97.45	96.39	99.93	99.88	97.44	97.60	99.94	99.83	97.35	97.12	99.91	99.80	97.44	96.51	99.91	99.81	97.65	96.51	99.90
	PPV	99.43	92.26	89.67	99.78	99.65	92.47	92.85	99.83	99.47	92.09	91.50	99.72	99.39	92.19	89.97	99.71	99.43	92.83	89.97	99.69

IV. RESULTS

The crisp model and the FESs (all combinations between the fuzzy membership functions and the $T_{\rm norm}$ and $S_{\rm norm}$ definitions) were tested using both MSE cost function and total accuracy cost function (CME), for both arrhythmic and ischemic beat classification. The test dataset $(D_{\rm test})$ consisted of the remaining patterns of D after selecting $D_{\rm train}$ (the $D_{\rm train}$ selection was made as described above for each problem); both $D_{\rm test}^{\rm arh}$ and $D_{\rm test}^{\rm isch}$ are presented in Table III. Ten different pairs of $D_{\rm train}$ and $D_{\rm test}$ were created. The crisp model of both arrhythmic beat classification and ischemic beat classification problems were evaluated for all $D_{\rm test}$ datasets, resulting to ten normalized confusion matrices which were combined using gross statistics to result to the average confusion matrix. Finally, sensitivity (Se), specificity (Sp), and positive predictive value (PPV), for the average confusion matrix, were calculated. The same procedure was followed for the FESs; they were optimized (using $D_{\rm train}$) and evaluated (using $D_{\rm test}$) with each pair of them. The maximum number of iterations of the MCRS algorithm was set to 10 000; this ensures that the algorithm would stop either when the convergence criterion was satisfied (Appendix, MCRS algorithm, Step 1, third bullet), or when the maximum number of iterations was reached. For the arrhythmic beat classification problem, Se, Sp, and PPV of the crisp model and the FESs created using the CME cost function are presented in Table IV. The results using the MSE cost function are quite similar; the average absolute difference is 0.25% while the maximum absolute difference is 2.25%. For the ischemic beat classification problem, all evaluation results are presented in Table V. In Table VI, accuracy (acc) for the crisp model and all FESs, for both arrhythmic and ischemic beat classification problems, using both me and MSE cost functions are presented, along with the number of parameters entering each fuzzy model.

From the obtained results it is clear that the application of the proposed methodology improved the efficiency of the initial crisp model; the best FES for the arrhythmic beat classification results to 96.43% accuracy, improving by 5.36% the corresponding accuracy of the crisp model, while, in the case of the ischemic beat classification the corresponding improvement is 11.27%. The number of beats in test sets is sufficiently large; thus, the error rates, defined as: $e = 1 - acc$, of the crisp model and the best fuzzy model in both cases (i.e., arrhythmic and ischemic beat classification) can be approximated using normal distributions [48]. If the observed difference in e is defined as: $d = |e_1 - e_2|$, then d is also normally distributed, with variance: $\sigma_d^2 = (acc_c\,(1 - acc_c) + acc_f\,(1 - acc_f))/N$, where N the number of test records (i.e., number of beats), acc_c is the accuracy of the crisp model and acc_f is the accuracy of the best fuzzy model. At 95% confidence level, the upper bound for the standard normal distribution is 1.96, and, thus, the confidence interval for the true difference d_t is given by: $d_t = d \pm 1.96\,\sigma_d$.

TABLE V
SENSITIVITY (%), SPECIFICITY (%) AND POSITIVE PREDICTIVE VALUE (%) OF THE CRISP MODEL AND THE AVERAGE CONFUSION MATRICES FOR ALL FUZZY EXPERT SYSTEMS FOR THE ISCHEMIC BEAT CLASSIFICATION PROBLEM, USING BOTH CME AND MSE COST FUNCTIONS

		Crisp
	Se	71.23
	Sp	64.42
	PPV	65.72

		cme					*mse*				
		Linear	Sigmoid	Sum of sigmoid and its gradient	Nested sigmoid	Sum of two sigmoids	Linear	Sigmoid	Sum of sigmoid and its gradient	Nested sigmoid	Sum of two sigmoids
Minimum & maximum	Se	76.70	78.47	78.92	78.99	78.77	77.01	78.61	79.29	77.21	80.79
	Sp	70.63	72.33	73.32	72.94	73.18	70.90	72.05	74.66	74.08	74.62
	PPV	71.44	73.09	73.91	73.65	73.77	71.71	72.93	74.98	74.04	75.30
Algebraic product & probabilistic OR	Se	78.18	79.65	79.80	79.85	79.60	78.67	80.02	81.16	81.44	78.27
	Sp	69.79	73.32	73.15	73.18	73.40	70.44	72.98	73.88	72.59	71.63
	PPV	71.25	74.09	74.00	74.03	74.13	71.82	73.93	74.84	73.99	72.55
Einstein product and sum	Se	77.88	80.01	80.27	79.79	79.94	78.15	80.81	78.74	80.52	79.94
	Sp	70.92	73.18	72.87	73.35	74.31	70.26	72.40	73.63	72.00	72.14
	PPV	71.95	74.07	73.91	74.14	74.87	71.57	73.72	74.09	73.36	73.32
Dombi class	Se	79.36	80.83	81.03	81.66	82.07	78.77	80.93	81.99	82.46	80.39
	Sp	73.46	75.30	75.38	74.02	74.58	73.86	74.86	75.60	74.67	74.15
	PPV	74.12	75.81	75.92	75.07	75.56	74.26	75.51	76.30	75.72	74.87
Dubois-Prade class	Se	78.77	81.42	82.00	81.06	80.71	79.03	82.71	80.84	80.79	82.59
	Sp	73.18	73.88	74.91	75.14	75.07	73.41	74.48	73.24	74.36	74.07
	PPV	73.77	74.91	75.78	75.74	75.62	74.01	75.64	74.31	75.11	75.31
Yager class	Se	79.03	81.48	81.64	81.19	82.28	79.39	81.59	81.19	82.71	80.97
	Sp	73.32	74.73	73.88	76.14	74.17	73.44	74.97	73.78	75.41	72.20
	PPV	73.94	75.54	74.96	76.52	75.31	74.11	75.74	74.78	76.31	73.61

For arrhythmic beat classification, the confidence interval for d_t at 95% confidence level is $0.0536 \pm 0.2e10^{-5}$, which does not spam the zero value, and, thus, the observed difference is statistically significant. Similarly, for ischemic beat classification, the confidence interval for d_t at 95% confidence level is $0.1127 \pm 0.1e10^{-5}$, which also does not spam the zero value, and, thus, the observed difference is statistically significant. In both cases the observed difference is also statistically significant if the confidence level is set to 99%; in this case, the upper bound for the standard normal distribution is 2.58 and the confidence intervals for d_t are $0.0536 \pm 0.3e10^{-5}$ and $0.1127 \pm 1.3e10^{-5}$ for arrhythmic and ischemic beat classification, respectively.

The selection of the cost function does not have an impact on the obtained results; for both CME and MSE cost functions the results were similar for arrhythmic and ischemic beat classification. The obtained results are slightly improved if a parameter-based approach was incorporated for the $T_{\rm norm}$ and $S_{\rm norm}$ definitions (i.e., Dompi, Dubois–Prade or Yager class), compared to the approaches which are based on parameter-free definitions (i.e., minimum and maximum, algebraic product and probabilistic OR, Einstein product and sum). An average increase 1.2% exists independently of the examined problem or the incorporated cost function or the fuzzy membership function selection. The extra $T_{\rm norm}$ and $S_{\rm norm}$ parameters make the fuzzy models more flexible, and, thus, the optimization results to better FESs. However, with respect to the fuzzy membership function selection, the FESs for the arrhythmic beat classification problem using fuzzy membership functions with less parameters (sigmoid or sum of a sigmoid and its gradient) have slightly better results than the ones with more parameters (nested sigmoid or sum of two sigmoids); the results of the FESs for the arrhythmic beat classification problem when the sigmoid or sum of a sigmoid and its gradient fuzzy membership functions are incorporated are 0.22% on average better than when the nested sigmoid or sum of two sigmoids fuzzy membership functions are used. The FESs, for the ischemic beat classification problem show similar performance. In all cases, the linear fuzzy membership function has the worst results; 1.12% reduction for arrhythmic beat classification and 1.9% for ischemic beat classification.

V. DISCUSSION

In this paper, we describe a methodological framework for the automated generation of FESs, which are based on an initial crisp model that includes a set of rules. The set of rules is represented in DNF, using the crisp memberhip function, formulating a crisp model. Then, the rules of the crisp model are transformed to fuzzy ones, forming the fuzzy model. This fuzzification is based on the use of a fuzzy membership function instead of the crisp one and the use of $T_{\rm norm}$ and $S_{\rm norm}$ instead of the binary operators. The produced fuzzy models are tuned using global optimization. Given an initial set of rules, the pro-

TABLE VI
ACCURACY (%) FOR THE CRISP MODEL AND ALL FUZZY EXPERT SYSTEMS, FOR BOTH ARRHYTHMIC AND ISCHEMIC BEAT CLASSIFICATION PROBLEMS, USING BOTH CME AND MSE COST FUNCTIONS

Arrhythmic beat classification (crisp model's accuracy: 91.07)

	cme					*mse*				
	Linear	Sigmoid	Sum of sigmoid and its gradient	Nested sigmoid	Sum of two sigmoids	Linear	Sigmoid	Sum of sigmoid and its gradient	Nested sigmoid	Sum of two sigmoids
Minimum & maximum	93.73 / 45	94.93 / 45	94.90 / 45	94.84 / 89	94.84 / 111	93.58 / 45	95.01 / 45	95.00 / 45	94.86 / 89	94.74 / 111
Algebraic product & probabilistic OR	94.18 / 45	95.20 / 45	95.26 / 45	95.24 / 89	95.05 / 111	94.22 / 45	95.32 / 45	95.48 / 45	95.07 / 89	95.02 / 111
Einstein product and sum	94.16 / 45	95.10 / 45	95.44 / 45	94.97 / 89	95.09 / 111	94.15 / 45	95.24 / 45	95.33 / 45	95.33 / 89	94.93 / 111
Dombi class	94.45 / 55	96.32 / 55	95.62 / 55	95.93 / 99	95.53 / 121	94.83 / 55	96.25 / 55	95.57 / 55	95.85 / 89	95.14 / 121
Dubois-Prade class	94.84 / 55	96.06 / 55	95.71 / 55	96.43 / 99	95.45 / 121	94.47 / 55	95.87 / 55	95.34 / 55	95.40 / 89	95.28 / 121
Yager class	94.50 / 55	96.15 / 55	95.66 / 55	95.24 / 99	95.40 / 121	94.43 / 55	95.86 / 55	96.18 / 55	96.06 / 89	95.40 / 121

Ischemic beat classification (crisp model's accuracy: 67.75)

	cme					*mse*				
	Linear	Sigmoid	Sum of sigmoid and its gradient	Nested sigmoid	Sum of two sigmoids	Linear	Sigmoid	Sum of sigmoid and its gradient	Nested sigmoid	Sum of two sigmoids
Minimum & maximum	73.60 / 15	75.33 / 15	76.06 / 15	75.90 / 29	75.91 / 36	73.47 / 15	74.69 / 15	76.23 / 15	76.47 / 29	74.99 / 36
Algebraic product & probabilistic OR	73.89 / 15	76.42 / 15	76.40 / 15	76.40 / 29	76.43 / 36	74.06 / 15	76.60 / 15	76.07 / 15	76.07 / 29	75.97 / 36
Eintein product and sum	74.32 / 15	76.52 / 15	76.49 / 15	76.50 / 29	77.06 / 36	74.87 / 15	76.97 / 15	75.98 / 15	75.89 / 29	77.57 / 36
Dombi class	76.34 / 18	78.00 / 18	78.14 / 18	78.14 / 32	78.25 / 39	76.07 / 18	78.79 / 18	76.86 / 18	76.86 / 32	78.71 / 39
Dubois-Prade class	75.91 / 18	77.57 / 18	78.38 / 18	78.04 / 32	77.83 / 39	76.25 / 18	77.09 / 18	79.02 / 18	78.18 / 32	76.89 / 39
Yager class	76.11 / 18	78.03 / 18	77.68 / 18	78.61 / 32	78.13 / 39	76.31 / 18	78.32 / 18	76.41 / 18	77.97 / 32	78.91 / 39

posed methodology can automatically produce a FES, for any problem under consideration. This is a due to: a) the employment of the DNF which can be used to every logical expression; b) the transformation of the crisp rules into fuzzy rules is performed using a fuzzy membership function and an approach for the binary operators and; therefore, it can be carried out for any set of rules written in a DNF in an simple and automated manner; and c) the optimization technique is a "derivative-free" algorithm and it does not need any information other than the objective function.

There are some approaches proposed in the literature which are based on the same philosophy, i.e., proposing a model and optimizing its parameters [5]–[7]. In [6] and [7], the methodology followed is designed for a specific problem, and it is not a general approach. In [5], a framework for the development and optimization of fuzzy models is described, but the initial model is based on the entropy of the data and not on a set of rules. Also the methodology is evaluated using only artificial data. All the optimization techniques used in [5]–[7] are "derivative free." In the proposed methodology, although the optimization technique does not require derivatives, there are available for all fuzzy models, except is the case where the minimum & maximum approach is used for the T_{norm} and S_{norm} definition [49]. We have not tested other optimization methods; MCRS is a recently developed global optimization technique, having several advantages and presenting superior robustness among its peers, although is not the most efficient one [47]. Furthermore, our main concern was to locate the global minimum and not to find the most efficient optimization method. However, this must be examined in a future communication.

This general framework is evaluated in the development of FESs for two cardiovascular domain problems, the arrhythmic and the ischemic beat classification, using ECG recordings. Medical experts provided the initial set of rules, which are represented in a DNF, forming the crisp model. The crisp rules are transformed into fuzzy, using several different combinations of common fuzzy membership functions and approaches for the T_{norm} and S_{norm} definitions. However, for both presented applications, the defuzzifiers are designed based on knowledge provided by experts on the specific domains and they are not common approaches proposed in the literature. These, task-specific defuzzification procedures can be considered as an advantage compared to common defuzzification procedures, since they reflect the experts' knowledge for the specific domains, but they have also the disadvantage that known implicit input-output formulas, proposed in the literature, cannot be applied. The MCRS algorithm is used to tune the parameters of the models (Θ). The MIT-BIH arrhythmia database and the ESC ST-T database are used for the model's optimization, for the arrhythmic and the ischemic beat classification, respectively. In both cases, the obtained results indicate a significant improvement compared to the initial crisp models. Thus, our methodology can be proven to be a useful tool for the improvement of the efficiency of existing knowledge-based systems.

TABLE VII
SUMMARY OF PREVIOUS METHODS FOR ARRHYTHMIC BEAT CLASSIFICATION AND ISCHEMIC BEAT CLASSIFICATION

Arrhythmic beat classification			
Authors	Method	Dataset (beats)	Accuracy (%)
Simon & Eswaran [18]	decision based neural network	1,096	96.03
Langerholm et al. [21]	Hermite functions and self organizing maps	108,963	98.49[2]
Dokur&Olmez [19]	discrete wavelet transform and intersecting spheres network	3,000	95.7
Osowski&Linh [22]	cumulants of the second, third and fourth order and fuzzy hybrid neural network	7,185	96.06
Ge et al. [23]	autoregressive modelling	856	96.84[3]
Osowski et al. [25]	support vector machines	12,785	95.91
Chazal et al. [26]	ECG morphology and linear discriminates	100,000	97.5
Hu et al. [20]	PCA in 29 points from QRS, instantaneous and average RR-interval, QRS complex width and mixture of experts (SOM, LVQ)	49,260	95.52[1]
Tsipouras et al. [27]	knowledge-based system	109,880	94.26
this work	**fuzzy expert system**	**109,880**	**96.43**

Ischemic beat classification				
Authors	Method	Sensitivity (%)	Specificity (%)	Accuracy (%)
Goletsis et al. [7]	GA & multicriteria decision analysis	91	91	
Stamkopoulos et al. [31]	Feed forward ANN and nonlinear PCA	79	75	
Papaloukas et al. [35]	Rule-based	70	63	
Papaloukas et al. [37]	ANN & principal components analysis	90	90	
Maglaveras et al. [38]	Bidirectional associative memories ANN			56
Papadimitriou et al. [39]	ANN (Classification partitioning-SOM & SVM)			80
this work	**fuzzy expert system**	81	73	**79**

[1] Calculated from the results presented in Table VI of [20].
[2] Calculated from the results presented in Table VI of [21].
[3] Calculated from the results presented in Table 2 of [23].

The proposed methodology produced very efficient FESs for the arrhythmic beat classification problem, which present several advantages compared to other methods for arrhythmic beat classification: a) they use only the RR-interval signal, which can be extracted with high accuracy even for noisy or complicated ECG recordings (e.g., the 200 series of the MIT-BIH arrhythmia database), while the extraction of all other ECG features or any other type of ECG analysis is seriously affected by noise; b) they are based on medical knowledge, which is usually ignored in similar systems; c) they perform in real time; d) all other approaches use "closed" datasets, i.e., datasets containing data belonging only to the classes that are classified, which is not possible in real life data. In the proposed method the dataset does not contain data only from the classes that are classified; a more realistic approach is used: three classes (VF, VT, BII) are classified and the remaining data are classified as N; e) they are fully automated; and f) interpretation is available for the decisions made. A limitation is the use of the actual beats instead of QRS detection in the VF episodes of the 207 record. A summary of the results obtained for arrhythmic beat classification by other methods is shown in Table VII—all methods are evaluated using the MIT-BIH arrhythmia database. Most of the approaches are based on the analysis of the ECG signal [18]–[26] while the approaches proposed in [27], [28] and in the present work is based on the analysis of the RR-interval signal only. All methods indicate high performance, 94.26%–98.49%; the proposed FESs results in comparable performance. However, some of the methods are evaluated using very small datasets [18], [19], [22], [23]. In [20], initial labeling of the beats was required and there was no automatic QRS detection. A similar approach was used in [26] for the fiducial points. In [21], the primary objective was to perform clustering with a human expert performing the final beat classification. In our case, the resulted FES is evaluated using all records from the MIT-BIH arrhythmia database. It is fully automated, compared to [20] and [21] and there is no training stage, as in other approaches [18], [22].

Table VII also presents a summary of previous methods for ischemic beat classification. A direct comparison is not feasible, since the evaluation is made with other datasets [34], [35], or different subsets of the ESC ST-T database [32], [39], [40], or employing different performance measures [34], [39], [40]. The FESs produced from the methodology, present several advantages compared to other methods for ischemic beat classification: a) they are based on medical knowledge. In [36] a medical rule-based methodology is employed, however the results are rather poor; b) they performs in real time; c) they are fully automated; and d) they can provide interpretation for the decisions made. Most of the proposed approaches are based on ANNs [32], [38], and, thus, interpretation is not available. The interpretation ability characterizes also the methods proposed in [7] and [36]. In [7], genetic algorithms were utilised for ischemic beat detection; the latter performed better than our approach but it requires high computational effort and processing time to tune the parameters and it is not based on medical knowledge.

The application of the proposed methodology is not limited to medical domain problems and can be extended to other domains for problems having the same structure, i.e., decision based on a set of rules. An additional advantage is that the method can be used along with data mining methods, which usually lead to a set of crisp rules. The incorporation of data mining for the initial set of rules acquisition leads to a fully automated method, where only the diagnostic experience of doctors is needed. However, the proposed methodology has some limitations: a) it is not easy to express the fuzzy inference model to a closed form, b) it is limited to applications that are based on crisp rules, and c) it greatly depends on the selection and the quality of the initial set of rules. The third limitation can be overcame if the initial set of rules is combined with a data mined set of rules.

VI. Conclusion

A methodology for FES creation is proposed. The methodology, which is fully automated, uses an initial set of crisp rules, provided by experts, and produces a FES. The methodology has been tested in the arrhythmic and ischemic beat classification problems and the produced FESs indicate significant improvement of the initial classification system, which is based on expert's knowledge and has the form of a set of rules. The obtained results are also fully interpretable, which is a major advantage compared to other approaches proposed in the literature for the specific problems.

In the proposed methodology, the initial crisp set of rules are determined by experts. Starting from a crisp set of rules and then transforming it into a fuzzy model allows our methodology to be applied in cases where the initial set of rules is strictly crisp. Based on this feature, an alternative is to extract rules from the data, using a data mining technique, instead of a knowledge-based origin of the initial set of rules. In this case, the methodology would be fully automated, data driven, but the knowledge introduced from the experts through the initial set of rules, would be lost. Furthermore, hybrid approaches, combining to expert's knowledge and data-mined rules are also applicable. In this context, also the ability to automatically predetermine some of the fuzzy model's basic aspects, such as the fuzzy membership function and the T_{norm} and S_{norm} definitions based on the natural characteristics of the problem, is a significant field of research. Another important feature is that the gradient is available for some of the fuzzy models; this feature can lead to the use of more efficient optimization methods, which take advantage of the first derivative information.

Appendix

In this Appendix, details regarding the equations of the fuzzy models (15)–(21) will be provided along with a detailed description of the MCRS algorithm.

Substituting in (15)–(17), the sigmoid membership function (Table III) and the minimum and maximum definitions for the T_{norm} and S_{norm}, for the arrhythmic fuzzy rules it, see (23)–(25), shown at the bottom of the page.

In (19), the results of the fuzzy rules are used without the use of a defuzzifier. In (18), $\max\left(R_1\left(d^l,\theta_1\right), R_2\left(d^l,\theta_2\right), R_3\left(d^l,\theta_3\right)\right)$ is used, which can be considered as a separate fuzzy model, using $R_1\left(d^l,\theta_1\right)$, $R_2\left(d^l,\theta_2\right)$ and $R_3\left(d^l,\theta_3\right)$ fuzzy rules and maximum defuzzifier [see (26), shown at the bottom of the next page].

Again, substituting in (20), the sigmoid membership function (Table III) and the minimum & maximum definitions for the T_{norm} and S_{norm}, for the ischemic fuzzy rule, it is (27), shown at the bottom of the next page.

$$R_1\left(d^l,\theta_1\right) = \max\left(\min\left(\frac{1}{1+e^{\theta^a_{1,1}\left(RR_{l-1}-\theta^b_{1,1}\right)}}, \frac{1}{1+e^{\theta^a_{1,2}\left(RR_{l}-\theta^b_{1,2}\right)}}, \frac{1}{1+e^{\theta^a_{1,3}\left(RR_{l+1}-\theta^b_{1,3}\right)}}\right) \frac{1}{1+e^{\theta^a_{1,4}\left(RR_{l-1}+RR_l+RR_{l+1}-\theta^b_{1,4}\right)}}\right) \tag{23}$$

$$\begin{aligned} R_2\left(d^l,\theta_2\right) &= \max\left(\min\left(\frac{1}{1+e^{\theta^a_{2,1}\left(\theta^b_{2,1}-RR_{l-1}/RR_l\right)}}, \frac{1}{1+e^{\theta^a_{2,2}\left(\theta^b_{2,2}-RR_{l+1}/RR_l\right)}}\right),\right.\\ &\min\left(\frac{1}{1+e^{\theta^a_{2,3}\left(\theta^b_{2,3}-RR_{l+1}/RR_{l-1}\right)}}, \frac{1}{1+e^{\theta^a_{2,4}\left(\theta^b_{2,4}-RR_{l-1}/RR_l\right)}}\right)\\ &\min\left(\frac{1}{1+e^{\theta^a_{2,5}\left(\left|RR_{l-1}-RR_l\right|-\theta^b_{2,5}\right)}}, \frac{1}{1+e^{\theta^a_{2,6}\left(RR_{l-1}-\theta^b_{2,6}\right)}}, \frac{1}{1+e^{\theta^a_{2,7}\left(RR_{l}-\theta^b_{2,7}\right)}}, \frac{1}{1+e^{\theta^a_{2,8}\left((RR_{l-1}+RR_l)/(2RR_{l+1})-\theta^b_{2,8}\right)}}\right)\\ &\left.\min\left(\frac{1}{1+e^{\theta^a_{2,9}\left(\left|RR_{l}-RR_{l+1}\right|-\theta^b_{2,9}\right)}}, \frac{1}{1+e^{\theta^a_{2,10}\left(RR_{l}-\theta^b_{2,10}\right)}}, \frac{1}{1+e^{\theta^a_{2,11}\left(RR_{l+1}-\theta^b_{2,11}\right)}}, \frac{1}{1+e^{\theta^a_{2,12}\left((RR_{l}+RR_{l+1})/(2RR_{l-1})-\theta^b_{2,12}\right)}}\right)\right) \end{aligned} \tag{24}$$

$$\begin{aligned} R_3\left(d^l,\theta_3\right) &= \max\left(\min\left(\frac{1}{1+e^{\theta^a_{3,1}\left(\theta^b_{3,1}-RR_l\right)}}, \frac{1}{1+e^{\theta^a_{3,2}\left(RR_{l},-\theta^b_{3,2}\right)}}, \frac{1}{1+e^{\theta^a_{3,3}\left(\left|RR_1-RR_2\right|-\theta^b_{3,3}\right)}}\right)\right.\\ &\left.\min\left(\frac{1}{1+e^{\theta^a_{3,4}\left(\theta^b_{3,4}-RR_{l},\right)}}, \frac{1}{1+e^{\theta^a_{3,5}\left(RR_{l},-\theta^b_{3,5},\right)}}, \frac{1}{1+e^{\theta^a_{3,6}\left(\left|RR_{l+1}-RR_l\right|-\theta^b_{3,6}\right)}}\right)\right) \end{aligned} \tag{25}$$

The MCRS algorithm, which is a modification of the Price's algorithm seeking for one global minimum in a given domain D, is described.

MCRS Algorithm

Input Data

• M, an integer such that $M > N+1$, where N is the space dimension (suggested value: $M > 25N$)

• ε, a small positive constant (suggested value $\varepsilon = 10^{-6}$)

• ω, a rather large positive constant (suggested value $\omega = 1000$)

Step 0:

• Set $k = 0$. Form the initial set $S^k = \left\{x_1^k, x_2^k, \ldots, x_M^k\right\}$ by picking M points randomly from D.

• Evaluate $f_i^k = f\left(x_1^k\right)$ for $i = 1, 2, \ldots, M$.

Step 1:

• $f_{\max}^k = \max\left\{f_i^k\right\}$ and let the corresponding point be denoted as $x_{\max}^k$. Similarly

• $f_{\min}^k = \min\left\{f_i^k\right\}$ and let the corresponding point be denoted as $x_{\min}^k$.

• IF $f_{\max}^k - f_{\max}^k \leq \varepsilon$ polish $f_{\min}^k$ via a local search procedure and STOP.

Step 2:

• Choose random by $N+1$ points $\left\{x_{i0}^k, x_{i1}^k, \ldots, x_{iN}^k\right\}$ from S^k.

• Calculate the weighted centroids: $c_w^k = \sum_{j=1}^N w_j^k x_j^k$, $f_w^k = \sum_{j=1}^N w_j^k f\left(x_j^k\right)$, where $w_j^k = \left(n_j^k / \sum_{j=1}^N n_j^k\right)$, $n_j^k = \left(1/f\left(x_{ij}\right) - f_{\min}^k + \phi^k\right)$ $\phi^k = \omega\left(f_{\max}^k - f_{\min}^k\right)^2 / f_{\max}^0 - f_{\min}^0$

$$
\begin{aligned}
&\max\left(R_1\left(d^l, \theta_1\right), R_2\left(d^l, \theta_2\right), R_3\left(d^l, \theta_3\right)\right) \\
&= \max\Bigg(\min\left(\frac{1}{1+e^{\theta_{1,1}^a\left(RR_{l-1}-\theta_{1,1}^b\right)}}, \frac{1}{1+e^{\theta_{1,2}^a\left(RR_l-\theta_{1,2}^b\right)}}, \frac{1}{1+e^{\theta_{1,3}^a\left(RR_{l+1}-\theta_{1,3}^b\right)}}\right) \\
&\qquad \min\left(\frac{1}{1+e^{\theta_{1,4}^a\left(RR_{l-1}+RR_l+RR_{l+1}-\theta_{1,4}^b\right)}}\right) \\
&\qquad \min\left(\frac{1}{1+e^{\theta_{2,1}^a\left(\theta_{2,1}^b-RR_{l-1}/RR_l\right)}}, \frac{1}{1+e^{\theta_{2,2}^a\left(\theta_{2,2}^b-RR_{l+1}/RR_l\right)}}\right) \\
&\qquad \min\left(\frac{1}{1+e^{\theta_{2,3}^a\left(\theta_{2,3}^b-RR_{l+1}/RR_{l-1}\right)}}, \frac{1}{1+e^{\theta_{2,4}^a\left(\theta_{2,4}^b-RR_{l-1}/RR_l\right)}}\right) \\
&\qquad \min\left(\frac{1}{1+e^{\theta_{2,5}^a\left(\left|RR_{l-1}-RR_l\right|-\theta_{2,5}^b\right)}}, \frac{1}{1+e^{\theta_{2,6}^a\left(RR_{l-1}-\theta_{2,6}^b\right)}}, \frac{1}{1+e^{\theta_{2,7}^a\left(RR_l-\theta_{2,7}^b\right)}}, \frac{1}{1+e^{\theta_{2,8}^a\left((RR_{l-1}+RR_l)/(2RR_{l+1})-\theta_{2,8}^b\right)}}\right) \\
&\qquad \min\left(\frac{1}{1+e^{\theta_{2,9}^a\left(\left|RR_l-RR_{l+1}\right|-\theta_{2,9}^b\right)}}, \frac{1}{1+e^{\theta_{2,10}^a\left(RR_l-\theta_{2,10}^b\right)}}, \frac{1}{1+e^{\theta_{2,11}^a\left(RR_{l-1}-\theta_{2,11}^b\right)}}, \frac{1}{1+e^{\theta_{2,12}^a\left((RR_l+RR_{l+1})/(2RR_{l-1})-\theta_{2,12}^b\right)}}\right) \\
&\qquad \min\left(\frac{1}{1+e^{\theta_{3,1}^a\left(\theta_{3,1}^b,-RR_l\right)}}, \frac{1}{1+e^{\theta_{3,2}^a\left(RR_l-\theta_{3,2}^b\right)}}, \frac{1}{1+e^{\theta_{3,3}^a\left(\left|RR_{l-1}-RR_l\right|-\theta_{3,3}^b\right)}}\right) \\
&\qquad \min\left(\frac{1}{1+e^{\theta_{3,4}^a\left(\theta_{3,4}^b,-RR_l\right)}}, \frac{1}{1+e^{\theta_{3,5}^a\left(RR_l-\theta_{3,5}^b\right)}}, \frac{1}{1+e^{\theta_{3,6}^a\left(\left|RR_{l+1}-RR_l\right|-\theta_{3,6}^b\right)}}\right)\Bigg) \qquad (26)
\end{aligned}
$$

$$
\begin{aligned}
R_1\left(d^l, \theta_1\right) = \max\Bigg(&\min\left(\frac{1}{1+e^{\theta_{1,1}^a\left(ST_l^{\text{dev}}-\theta_{1,1}^b\right)}}, \frac{1}{1+e^{\theta_{1,2}^a\left(\theta_{1,2}^b-ST_l^{\text{slope}}\right)}}\right) \\
&\frac{1}{1+e^{\theta_{1,3}^a\left(\theta_{1,3}^b-ST_l^{\text{dev}}\right)}} \\
&\frac{1}{1+e^{\theta_{1,4}^a\left(T_l^{\text{amp}}\cdot T_l^{\text{norm}}-\theta_{1,4}^b\right)}} \\
&\frac{1}{1+e^{\theta_{1,5}^a\left(\left|T_l^{\text{amp}}\right|/\left|T_l^{\text{norm}}\right|-\theta_{1,5}^b\right)}} \\
&\min\left(\frac{1}{1+e^{\theta_{1,6}^a\left(T_l^{\text{amp}}/T_l^{\text{norm}}-\theta_{1,6}^b\right)}}, \frac{1}{1+e^{\theta_{1,7}^a\left(\theta_{1,7}^b-\left|\left|T_l^{\text{amp}}\right|-\left|T_l^{\text{norm}}\right|\right|\right)}}\right)\Bigg) \qquad (27)
\end{aligned}
$$

• Calculate the trial point $\bar{x}^k$ as: $\bar{x}^k = \left(x_{i0}^k - c_w^k\right) f\left(x_{io}^k\right) - f_w^k / f_{\max}^k - f_{\min}^k + \phi^k + \Delta_w^k$, where $\Delta_w^k = 2c_w^k - x_{i0}^k$ if $f_w^k \leq f\left(x_{i0}^k\right)$ and $\Delta_w^k = 2x_{i0}^k - c_w^k$ if $f_w^k > f\left(x_{i0}^k\right)$

• IF $\bar{x}^k \notin D$ REPEAT Step 2

• Compute $f\left(\bar{x}^k\right)$.

Step 3:

• IF $f\left(\bar{x}^k\right) \geq f_{\max}^k$ THEN

— Calculate the success rate (the fraction of function evaluations that led to a new lower upper bound)

— IF success rate > 0.5 set $S^{k+1} = S^k$, $k = k + 1$ and GOTO Step 2

— Calculate $y^k = (c_w^k + x_{iN}^k/2)$, compute $f_y = f\left(y^k\right)$

— IF $f_y \geq f_{\max}^k$, set $S^{k+1} = S^k$, $k = k + 1$ and GOTO Step 2

— Set $S^{k+1} = S^k \cup \left\{\bar{x}^k\right\} - \left\{x_{\max}^k\right\}$, $k = k + 1$ and GOTO Step 1

• ENDIF

Step 4:

• Set $S^{k+1} = S^k \cup \left\{\bar{x}^k\right\} - \left\{x_{\max}^k\right\}$

• Increment $k = k + 1$ and GOTO Step 1.

ACKNOWLEDGMENT

The authors would like to thank Prof. I. Lagaris for his help and support throughout this research.

REFERENCES

[1] L. H. Tsoukalas and R. E. Uhrig, *Fuzzy and Neuralapproaches in Engineering*. New York: Wiley, 1997.

[2] L. X. Wang, *A Course in Fuzzy Systems and Control*. Upper Saddle River, NJ: Prentice-Hall, 1997.

[3] J. Giartano and G. Riley, *Expert Systems, Principles and Programming*. Boston, MA: PWS, 1998.

[4] S. H. Liao, "Expert system methodologies and applications-a decade review from 1995 to 2004," *Expert Syst. Appl.*, vol. 28, pp. 93–103, 2005.

[5] W. Pedrycz and J. V. d. Oliveira, "An algorithmic framework for development and optimization of fuzzy models," *Fuzzy Sets Syst.*, vol. 80, pp. 37–55, 1996.

[6] J. M. Garibalti and E. I. Ifeachor, "Application of simulated annealing fuzzy model tuning to umbilical cord acid-base interpretation," *IEEE Trans. Fuzzy Syst.*, vol. 7, no. 1, pp. 72–84, Feb. 1999.

[7] Y. Goletsis, C. Papaloucas, D. I. Fotiadis, A. Likas, and L. K. Michalis, "Automated ischemic beat classification using genetic algorithms and multicriteria decision analysis," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 1, pp. 1717–1725, Oct. 2004.

[8] J.-S. R. Jang and C.-T. Sun, "Neuro-fuzzy modeling and control," *Proc. IEEE*, vol. 83, no. 3, pp. 378–406, Mar. 1995.

[9] R. O. Duda, P. E. Hart, and D. G. Hart, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[10] J. C. Bazdek and S. K. Pal, *Fuzzy Models for Pattern Recognition*. New York: IEEE, 1992.

[11] X.-J. Zeng and M. G. Singh, "Approximation accuracy analysis of fuzzy systems as function approximators," *IEEE Trans. Fuzzy Syst.*, vol. 4, no. 1, pp. 44–63, Jan. 1996.

[12] E. K. P. Chong and S. H. Zak, *An Introduction to Optimization*. New York: Wiley, 2001.

[13] E. Hansen, *Global Optimization Using Interval Analysis*. New York: Dekker, 1992.

[14] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.

[15] E. Sandoe and B. Sigurd, *Arrhythmia -A Guide to Clinical Electrocardiology*. Bingen, Germany: Verlags GmbH, 1991.

[16] L. Goldberger and E. Goldberger, *Clinical Electrocardiography*. St. Louis, MO: Mosby, 1977.

[17] P. M. Spooner and M. R. Rosen, *Foundations of Cardiac Arrhythmias*. New York: Dekker, 2001.

[18] B. P. Simon and C. Eswarn, "An ECG beat classifier designed using modified decision based neural networks," *Comput. Biomed. Res.*, vol. 30, pp. 257–272, 1997.

[19] Z. Dokur and T. Olmez, "ECG beat classification by a hybrid neural network," *Comput. Meth. Progr. Biomed.*, vol. 66, pp. 167–181, 2001.

[20] Y. Z. Hu, S. Palreddy, and W. J. Tompkins, "A patient-adaptable ECG beat classifier using a mixture of experts approach," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 9, pp. 891–900, Sep. 1997.

[21] M. Lagerholm, C. Peterson, G. Braccini, L. Ebendrandt, and L. Sornmo, "Clustering ECG complexes using hermite functions and self-organizing maps," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 7, pp. 838–848, Jul. 2000.

[22] S. Osowski and T. H. Linh, "ECG beat recognition using fuzzy hybrid neural network," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 11, pp. 1265–1271, Nov. 2001.

[23] D. Ge, N. Srinivasan, and S. M. Krishnan, "Cardiac arrhythmia classification using autoregressive modering," *Biomed. Eng.*, vol. 1, no. 5, 2002.

[24] U. R. Acharya, P. S. Bhat, S. S. Iyengar, A. Rao, and S. Dua, "Classification of heart rate data usingartificial neural network and fuzzy equivalence relation," *Pattern Recognit.*, vol. 36, pp. 61–68, 2003.

[25] S. Osowski, T. H. Linh, and T. Markiewicz, "Support vector machine-based expert system for reliable heartbeat recognition," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 4, pp. 582–589, Apr. 2004.

[26] P. D. Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196–1206, Jul. 2004.

[27] M. G. Tsipouras, D. I. Fotiadis, and D. Sideris, "Arrhythmia classification using the RR-interval duration signal," in *Proc. Computers in Cardiology*, 2002, vol. 29, pp. 485–488.

[28] M. G. Tsipouras, D. I. Fotiadis, and D. Sideris, "An arrhythmia classification system based on the RR interval signal," *Artif. Intell. Med.*, vol. 33, pp. 237–250, 2005.

[29] M. J. Goldman, *Principles of Clinical Electrocardiography*, 11th ed. Los Altos, CA: Lange, 1982.

[30] A. Taddei, G. Distante, M. Emdin, P. Pisani, G. B. Moody, C. Zeelenberg, and C. Marchesi, "The European ST-T database: Standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography," *Eur. Heart J.*, vol. 13, pp. 1164–1172, 1992.

[31] R. Silipo, P. Laguna, C. Marchesi, and R. G. Mark, "ST-T segment change recognition using artificial neural networks and principal component analysis," in *Proc. Computers in Cardiology*, 1995, vol. 22, pp. 213–216.

[32] T. Stamkopoulos, K. Diamantaras, N. Maglaveras, and M. Strintzis, "ECG analysis using nonlinear pca neural networks for ischemia detection," *IEEE Trans. Signal Process.*, vol. 46, no. 11, pp. 3058–3067, Nov. 1998.

[33] F. Jager, R. G. Mark, G. B. Moody, and S. Divjak, "Analysis of transient ST segment changes during ambulatory monitoring using the Karhunen-Loeve transform," in *Proc. Computers in Cardiology*, 1992, vol. 19, pp. 691–694.

[34] I. Pitas, M. G. Strintzis, S. Grippas, and C. Xerostylides, "Machine classification of ischemic electrocardiograms," presented at the IEEE Mediterranean Electrotechnical Conf., Athens, Greece, 1983.

[35] L. Senhadji, G. Carrault, J. J. Bellanger, and G. Passariello, "Comparing wavelet transforms for recognizing cardiac patterns," *IEEE Eng. Med. Biol. Mag.*, vol. 14, no. 2, pp. 167–173, Mar./Apr. 1995.

[36] C. Papaloukas, D. I. Fotiadis, A. P. Liavas, A. Likas, and L. K. Michalis, "A knowledge-based technique for automated detection of ischemic episodes in long duration electrocardiograms," *Med. Biol. Eng. Comput.*, vol. 39, pp. 105–112, 2001.

[37] C. Papaloukas, D. I. Fotiadis, A. Likas, C. S. Stroumbis, and L. K. Michalis, "Use of a novel rule-based expert system in the detection of changes in the ST segment and the T wave in long duration ECGs," *J. Electrocardiol.*, vol. 35, pp. 27–34, Jan. 2002.

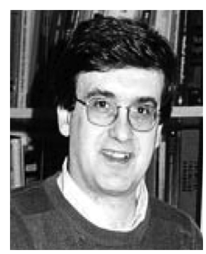
[38] C. Papaloukas, D. I. Fotiadis, A. Likas, and L. K. Michalis, "An ischemia detection method based on artificial neural networks," *Artif. Intell. Med.*, vol. 24, pp. 167–178, 2002.
[39] N. Maglaveras, T. Stamkopoulos, C. Pappas, and M. Strintzis, "ECG processing techniques based on neural networks and bidirectional associative memories," *J. Med. Eng. Technol.*, vol. 22, pp. 106–111, 1998.
[40] S. Papadimitriou, S. Mavroudi, L. Vladutu, and A. Bezerianos, "Ischemia detection with a self-organizing map supplemented by supervised learning," *IEEE Trans. Neural Netw.*, vol. 12, no. 3, pp. 503–515, May 2001.
[41] N. Maglaveras, T. Stamkopoulos, C. Pappas, and M. G. Strintzis, "An adaptive backpropagation neural network for real-time ischemia episodes detection: Development and performance analysis using the European ST-T database," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 7, pp. 805–813, Jul. 1998.
[42] *MIT-BMH Arrhythmia Database*. Cambridge, MA: MIT Press, 1997, Harvard-MIT Div. Health Sci. Technol.
[43] J. Pan and W. J. Thompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. 32, no. 3, pp. 230–236, Mar. 1985.
[44] P. S. Hamilton and W. J. Tompkins, "Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database," *IEEE Trans. Biomed. Eng.*, vol. 33, no. 12, pp. 1157–1165, Dec. 1986.
[45] *European ST-T Database Directory*. Pisa, Italy: S.T.A.R., 1991.
[46] K. Daskalov, I. A. Dotsinsky, and I. I. Christov, "Developments in ECG acquisition, preprocessing, parameter measurement, and recording," *IEEE Eng. Med. Biol.*, vol. 17, no. 1, pp. 50–58, Jan. 1998.
[47] F. V. Theos, I. E. Lagaris, and D. G. Papageorgiou, "PANMIN: Sequential and parallel global optimization procedures with a variety of options for the local search strategy," *CPC*, vol. 159, pp. 63–69, 2004.
[48] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Reading, MA: Pearson Addison Wesley, 2006.
[49] M. G. Tsipouras, C. Voglis, I. A. Lagaris, and D. I. Fotiadis, "A framework for fuzzy expert system creation," in *Proc. 7th Int. Workshop Math. Meth. in Scat. Theory Biomed. Eng.*, 2006, pp. 423–430.

Markos G. Tsipouras (S'07) was born in Athens, Greece, in 1977. He received the Diploma degree and the M.Sc. degree in computer science from the University of Ioannina, Ioannina, Greece, in 1999 and 2002, respectively. He is currently pursuing the Ph.D. degree in the Department of Computer Science, University of Ioannina.

His research interests include biomedical engineering, decision support and medical expert systems, and biomedical applications.

Costas Voglis was born in Ioannina, Greece, in 1978. He received the Diploma degree and the M.Sc. degree in computer science from the University of Ioannina in 1999 and 2002, respectively. He is currently pursuing the Ph.D. degree in the Department of Computer Science, University of Ioannina.

His research interests include local and global optimization.



Dimitrios I. Fotiadis (M'01–SM'07) was born in Ioannina, Greece, in 1961. He received the Diploma degree in chemical engineering from the National Technical University of Athens, Athens, Greece, and the Ph.D. degree in chemical engineering from the University of Minnesota, Minneapolis.

Since 1995, he has been with the Department of Computer Science, University of Ioannina, where he is currently an Associate Professor and Director of the Unit of Medical Technology and Intelligent Information Systems. His research interests include biomedical technology, biomechanics, scientific computing, and intelligent information systems.