Artificial Intelligence in Medicine (2007) 40, 187—200

ARTIFICIAL INTELLIGENCE IN MEDICINE

http://www.intl.elsevierhealth.com/journals/aiim

A methodology for the automated creation of fuzzy expert systems for ischaemic and arrhythmic beat classification based on a set of rules obtained by a decision tree

Themis P. Exarchos [a,b], Markos G. Tsipouras [a], Costas P. Exarchos [a], Costas Papaloukas [a,c], Dimitrios I. Fotiadis [a,d,e,*], Lampros K. Michalis [e,f]

[a] *Unit of Medical Technology and Intelligent Information Systems, Department of Computer Science, University of Ioannina, GR 45110 Ioannina, Greece*
[b] *Department of Medical Physics, Medical School, University of Ioannina, GR 45110 Ioannina, Greece*
[c] *Department of Biological Applications and Technology, University of Ioannina, GR 45110 Ioannina, Greece*
[d] *Biomedical Research Institute-FORTH, GR 45110 Ioannina, Greece*
[e] *Michaelideion Cardiology Centre, GR 45110 Ioannina, Greece*
[f] *Department of Cardiology, Medical School, University of Ioannina, GR 45110 Ioannina, Greece*

Received 6 October 2006; received in revised form 19 February 2007; accepted 5 April 2007

KEYWORDS
Fuzzy expert system;
Data mining;
Optimization;
Ischaemia;
Arrhythmia

Summary

Objective: In the current work we propose a methodology for the automated creation of fuzzy expert systems, applied in ischaemic and arrhythmic beat classification.
Methods: The proposed methodology automatically creates a fuzzy expert system from an initial training dataset. The approach consists of three stages: (a) extraction of a crisp set of rules from a decision tree induced from the training dataset, (b) transformation of the crisp set of rules into a fuzzy model and (c) optimization of the fuzzy model's parameters using global optimization.
Material: The above methodology is employed in order to create fuzzy expert systems for ischaemic and arrhythmic beat classification in ECG recordings. The fuzzy expert system for ischaemic beat detection is evaluated in a cardiac beat dataset that was constructed using recordings from the European Society of Cardiology ST-T database. The arrhythmic beat classification fuzzy expert system is evaluated using the MIT-BIH arrhythmia database.

* Corresponding author at: Unit of Medical Technology and Intelligent Information Systems, Department of Computer Science, University of Ioannina, PO Box 1186, GR 45110 Ioannina, Greece. Tel.: +30 26510 98803; fax: +30 26510 98889.
E-mail address: fotiadis@cs.uoi.gr (D.I. Fotiadis).

doi:10.1016/j.artmed.2007.04.001

Results: The fuzzy expert system for ischaemic beat classification reported 91% sensitivity and 92% specificity. The arrhythmic beat classification fuzzy expert system reported 96% average sensitivity and 99% average specificity for all categories.
Conclusion: The proposed methodology provides high accuracy and the ability to interpret the decisions made. The fuzzy expert systems for ischaemic and arrhythmic beat classification compare well with previously reported results, indicating that they could be part of an overall clinical system for ECG analysis and diagnosis.

1. Introduction

Cardiovascular diseases are the leading cause of death in many countries worldwide. The multifaceted nature of the diseases, combined with a wide variety of treatments and outcomes and complex relationships with other diseases, have made diagnosis of cardiovascular diseases a highly complex and important task, even for experienced cardiologists. Two of the most common cardiovascular diseases are myocardial ischaemia and cardiac arrhythmias.

Myocardial ischaemia is the most common cardiac disorder and its early diagnosis is of great importance. It is defined by a reduced blood flow to parts of the myocardium which causes alterations in the ECG signal, such as deviations in the ST segment and changes in the T wave [1]. Several techniques, which automate the detection and assist in the diagnosis of ischaemia in long duration ECGs have been proposed [2—16]. All these techniques can be described as a sequence of two tasks: ischaemic beat detection and ischaemic episode definition. The first is related to the classification of beats as normal or ischaemic, which is a key process for the definition of the ischaemic episodes in the ECG signal. Several techniques have been proposed for ischaemic beat detection which evaluate the ST segment changes and the T-wave alterations by different methodologies. More specifically, the use of approaches like statistical signal processing [2—4], fuzzy theory [5], wavelet theory [6], set of rules [7,8], artificial neural networks [9—13], multicriteria decision analysis [14], genetic algorithms [15] and association rule mining [16] have been previously reported. Signal processing [2—6] and neural networks [9—13] based approaches have resulted in high performance but require further post processing of the input parameters along with their weights in order to provide useful information. Rule-based approaches exhibit the highly desirable feature of interpreting the decisions but their performance is lower.

In what concerns cardiac arrhythmia, it can be defined as either an irregular single heartbeat (arrhythmic beat), or as an irregular group of heartbeats (arrhythmic episode). Arrhythmias can affect the heart rate causing irregular rhythms, such as slow or fast heartbeat. Arrhythmias can take place in a healthy heart and be of minimal consequence, but they may also indicate serious cardiovascular problems, which may lead to stroke or sudden cardiac death [17]. The ECG beat-by-beat analysis and classification can provide important information regarding the subject's cardiac condition. Several methods have been proposed in the literature for arrhythmic beat classification, where each beat is classified into several different rhythm types utilizing "mixture of experts approach" [18], hermite functions combined with self-organizing maps [19], artificial neural networks [20,21], fuzzy neural networks [22], autoregressive modelling [23], time-frequency analysis combined with knowledge-based systems [24], support vector machines [25], ECG morphology [26] and rule-based systems [27].

Expert systems are a branch of artificial intelligence that makes extensive use of specialized knowledge to solve problems at the level of a human expert. This knowledge is represented by a set of rules [28]. An area where expert systems are widely employed is the medical domain. Several parameters must be taken into consideration in order to create a medical expert system; the representation of medical knowledge and expertise, the decision making, and the choice and adaptation of a suitable model, are some of them. Also, uncertainty and imprecision, inherited in medical problems, is treated incorporating fuzzy logic [29,30]. Fuzzy expert systems (FES) include a set of fuzzy rules comprising a fuzzy model, while some of the model's parameters can be adjusted using global (or local) optimization techniques.

In this context, several approaches have been proposed in the literature: optimization of fuzzy rules with genetic algorithms [15,31] or simulated annealing [32]. Neuro-fuzzy algorithms have also been proposed [33]. In the latter, the fuzzy rules are modelled using an artificial neural network (ANN) and training techniques are employed. Also, a great effort has been made in the induction of decision trees using fuzzy partitions (fuzzy decision trees) and optimization of the parameters entering these trees [34—39]. Most of the works in this field employ genetic algorithms for the optimization of the fuzzy partitions [37—39]. In all the above research attempts, it has been shown that fuzzy decision trees and fuzzy rules, after the optimization of the parameters used, increase the accuracy of the respective crisp models significantly.

In this study, a methodology for the automated creation of fuzzy expert systems (FES) is proposed, that involves three stages: (i) extraction of a set of rules using data mining, (ii) generation of a fuzzy model, and (iii) optimization of the fuzzy model's parameters. Specifically, a set of rules is extracted from a decision tree, developed from a training set. In the second stage, the set of crisp rules is fuzzified, resulting into a fuzzy model. Finally, all the parameters entering the fuzzy model are tuned with respect to the classification accuracy of the fuzzy model, using global optimization. The fuzzy model with the optimized parameters composes the final FES. The generated FESs are able to provide interpretation for their decisions since they are based on sets of rules. In the first stage of the methodology, the use of data mining in the form of decision trees has the advantage of discovering new knowledge [40,41]. More specifically, the initial set of rules is extracted from a decision tree, which is considered a very effective technique for classification [37,42,43]. Furthermore, the development of the fuzzy model from the initial set of rules and the optimization of its parameters, improve the results obtained by the decision tree, while the incorporation of fuzzy logic addresses the uncertainty inherent in several classification problems [29].

We have employed the above methodology in two medical problems: ischaemic and arrhythmic beat classification. Those problems are considered very important in the context of clinical cardiology. In both cases, representative features are extracted from the cardiac beats. The QT interval, along with features extracted from the ST—T interval were used for ischaemic beat detection while features from the tachogram were employed for arrhythmic beat classification. In what concerns ischaemia, the QT interval and more specifically the corrected QT (QTc) interval has great clinical significance and is widely used in clinical practice since it is affected by various clinical conditions such as myocardial ischaemia or infarction with deep T wave inversion [44,45]. Also, the ST—T characteristics are known for their ischaemic diagnostic ability [46,47], while the tachogram can be used to characterize several types of cardiac arrhythmias [27,48]. In the case of ischaemia the classification output for each beat is normal (Norm) or ischaemic (Isch), while in the case of arrhythmia four classes are considered: beats belonging to ventricular flutter/fibrillation episodes (VF), premature ventricular contractions (PVC), normal beats (N) and beats belonging to 2° heart block episodes. The classification is performed using data from two task specific cardiac beat databases and the obtained results indicate that the proposed methodology is very effective and performs well both in terms of sensitivity and specificity.

In the following, we describe the proposed methodology in detail, the employed datasets, the preprocessing of the electrocardiographic (ECG) signal and the features used to create the FESs, for the two medical applications. Next the results of the evaluation are presented. The advantages and the disadvantages of the proposed methodology are given in Section 5. Comparison with previous works, as well as, possible further improvements are also discussed.

2. Materials and methods

The methodology automatically generates a FES, using an initial annotated dataset. The methodology involves three stages: (i) creation of a rule-based classifier using the annotated dataset, (ii) development of a fuzzy model, and (iii) optimization of the fuzzy model's parameters. The flowchart of the methodology is shown in Fig. 1; all stages are described below in detail. Briefly, an initial set of (crisp) rules is extracted from a decision tree, induced by the annotated dataset. The set of rules is transformed to a fuzzy model using a fuzzy membership function and fuzzy equivalents of the binary AND and OR operators. Finally, the fuzzy model is optimized with respect to its parameters, using the annotated dataset.

2.1. Extraction of a set of rules

In order to extract an initial set of rules from an annotated dataset, a rule mining technique must be employed. In our approach we used decision trees, however any rule mining technique could be employed. The construction of the decision tree is implemented using the C4.5 inductive algorithm [42], which is an effective and widely used decision tree induction algorithm and requires low computational effort [37,43]. C4.5 generates a decision tree from the training data that minimizes the expected value of the number of tests for data classification. Each internal node of the tree corresponds to a feature (a_j), while each outgoing branch corresponds to a feature test ($a_j \mathrm{op} \theta_j$), where ($a_j \theta_j$) is a feature—threshold pair and op is a comparison operator chosen from the set $\{=, \neq, <, >, \leq, \geq\}$. Each feature test j forms a (crisp) conjunct $c_j(a_j, \theta_j)$,

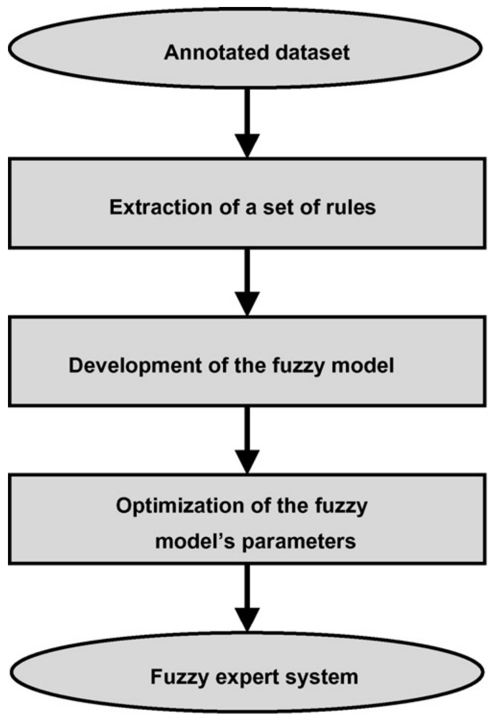

Figure 1 The proposed three-stage methodology for the automated generation of a fuzzy expert system (FES).

which, if op $\in \{\leq, >\}$ it is expressed as: $c_j(a_j, \theta_j) = g_c(a_j, \theta_j)$, where g_c is the crisp membership function, defined as

$$g_c^{inc}(a,\theta) = \begin{cases} 0 & a \leq \theta \\ 1 & a > \theta \end{cases} \text{(increasing)} \text{ or}$$
$$g_c^{dec}(a,\theta) = \begin{cases} 1 & a \leq \theta \\ 0 & a > \theta \end{cases} \text{(decreasing)}. \quad (1)$$

The leaf nodes represent the class to be assigned to a sample. The most important factor in the C4.5 algorithm is its ability to automatically select the feature, which is appropriate at each node. The feature of each node is selected in order to divide input samples effectively. *Information gain* [42] is used as a measure of effectiveness. After the induction of the decision tree, we apply a pruning method to reduce the tree's size and complexity. There exist two common methods for pruning [42]: prepruning and post-pruning. In our problem we followed the post-pruning method. Post-pruning tends to give better results than prepruning since it makes pruning decisions based on a fully grown tree, unlike prepruning, which can suffer from early termination of the tree growing process. In our case, post-pruning is performed by replacing a subtree with a new leaf node whose class label is determined from the majority class of records associated with the subtree (subtree replacement). The subtree replacement was performed by calculating the pessimistic error. The confidence factor for pruning was set to 0.25.

The produced tree can be easily transformed into a set of rules, as follows:

(a) One condition (Cond_i) is created for every leaf of the tree, by parsing the tree from the root node to that leaf. The feature tests encountered along the path form the conjuncts of the condition:

$$\text{Cond}_i(A,\Theta) = c_{\text{root}}(a_{\text{root}},\theta_{\text{root}}) \wedge c_{n_j}(a_{n_j},\theta_{n_j}) \wedge \cdots \wedge c_{n_k}(a_{n_k},\theta_{n_k}), \quad (2)$$

where Cond_i is a condition, $A = \{a_1, a_2, \ldots, a_{n_f}\}$ is the feature vector, $\Theta = \{\theta_1, \theta_2, \ldots, \theta_{n_t}\}$ is a vector containing all thresholds, n_f is the number of features characterizing a record, n_t is the total number of thresholds used in the decision tree. The class label at the leaf node is assigned to the rule consequent: $\text{Cond}_i(A, \Theta) \rightarrow y$, where y is the class.

(b) A general rule (R_y) is created for each class, using all the conditions $\text{Cond}_i(A, \Theta)$ having as consequent this class:

$$R_y(A,\Theta) = \text{Cond}_{j_1}(A,\Theta) \vee \text{Cond}_{j_2}(A,\Theta) \vee \cdots \vee \text{Cond}_{j_n}(A,\Theta), \quad (3)$$

where y is the class. These general rules comprise the crisp set of rules, which are in a disjunctive normal form.

2.2. Development of a fuzzy model

A fuzzy model is based on three fundamental aspects: the fuzzification method, the inference engine and the deffuzification [49]. Different combinations of the realizations of the above aspects result to different fuzzy models. In our approach, the crisp set of rules is transformed into a fuzzy model using a fuzzy membership function instead of the crisp one. The sigmoid function, defined as

$$g_s^{inc}(a,\theta_1,\theta_2) = \frac{1}{1+e^{\theta_1(\theta_2-a)}} \text{(increasing)} \text{ or}$$
$$g_s^{dec}(a,\theta_1,\theta_2) = \frac{1}{1+e^{\theta_1(a-\theta_2)}} \text{(decreasing)}, \quad (4)$$

is used as fuzzy membership function, for the fuzzification of the inputs. According to this, the crisp conjuncts are transformed to fuzzy ones as: $c_j^f(a_j,\theta_{1,j},\theta_{2,j}) = g_s(a_j,\theta_{1,j},\theta_{2,j})$. The fuzzy inference engine is defined establishing the T and S norms definitions (among the several definitions and classes that have been proposed in the literature) as long as the inference procedure between the fuzzy rules. In our approach, the minimum and maximum operators are used as T and S norms

[49]; thus the crisp conditions are transformed to fuzzy ones:

$$\mathrm{Cond}_i^{\mathrm{f}}(A,\Theta^{\mathrm{f}}) = \min\left\{\begin{array}{c} c_{\mathrm{root}}^{\mathrm{f}}(a_{\mathrm{root}},\theta_{1,\mathrm{root}},\theta_{2,\mathrm{root}}), \\ c_{n_j}^{\mathrm{f}}(a_{n_j},\theta_{1,n_j},\theta_{2,n_j}),\ldots, \\ c_{n_k}^{\mathrm{f}}(a_{n_k},\theta_{1,n_k},\theta_{2,n_k}) \end{array}\right\}, \quad (5)$$

where $\Theta^{\mathrm{f}} = \{\theta_{1,\mathrm{root}},\theta_{2,\mathrm{root}},\theta_{1,1},\theta_{2,1},\ldots,\theta_{1,n_t},\theta_{2,n_t}\}$ is a vector containing all parameters used in the fuzzy model.

We define a rule evaluation metric, the likelihood ratio, in order to measure how ''strong'' a rule is [40]:

$$p_i = 2\sum_{j=1}^{n_y} \mathrm{fr}_{i,j}\log\left(\frac{\mathrm{fr}_{i,j}}{e_{i,j}}\right), \quad (6)$$

where n_y is the number of classes, $\mathrm{fr}_{i,j}$ is the observed frequency of class j records, which are covered by a rule $\mathrm{Cond}_i(A, \Theta) \rightarrow y$, and $e_{i,j}$ is the expected frequency of a rule that makes random predictions. A large p_i suggests that the number of correct predictions made by the rule is significantly larger than that expected by random guessing. Other metrics for rule evaluation could be considered, however this was preferred since it takes into account both the accuracy and the coverage of the rules. This metric is applied to each $\mathrm{Cond}_i^{\mathrm{f}}$. Having $p = [p_1, p_2, \ldots, p_{n_c}]$ and $\mathrm{Cond}^{\mathrm{f}} = [\mathrm{Cond}_1^{\mathrm{f}}, \mathrm{Cond}_2^{\mathrm{f}}, \ldots, \mathrm{Cond}_{n_c}^{\mathrm{f}}]$ the general crisp rules are transformed to fuzzy ones:

$$R_y^{\mathrm{f}}(A,\Theta^{\mathrm{f}}) = \max\{\mathrm{diag}\{p^{\mathrm{T}}\mathrm{Cond}^{\mathrm{f}}\}\}, \quad (7)$$

where n_c is the number of conditions (cond). Eq. (7) defines the inference procedure between the fuzzy conditions of the same class. These fuzzy general rules comprise the fuzzy model:

$$M^{\mathrm{f}}(A,\Theta^{\mathrm{f}}) = \arg\max_{y=1,\ldots,n_y}(R_y^{\mathrm{f}}(A,\Theta^{\mathrm{f}})). \quad (8)$$

As it is shown in Eq. (8), for each feature vector A, the fuzzy general rule with the higher value defines its class. Eq. (8) defines the defuzzification procedure.

2.3. Fuzzy model's parameters optimization

The fuzzy model $M^{\mathrm{f}}(A, \Theta^{\mathrm{f}})$ is optimized with respect to its parameters Θ^{f}, using a training dataset (D_{train}). For every conjunct, a parameter θ_1 (analogous to the slope φ) and the centre θ_2 of the fuzzy membership function (sigmoid) are optimized (Fig. 2). If X is the normalized confusion matrix:

$$X_{M^{\mathrm{f}}(A,\Theta^{\mathrm{f}}),y} = \frac{\text{of patterns in } y \text{ classified to } M^{\mathrm{f}}\,(A,\Theta^{\mathrm{f}})}{\text{total of patterns in } y}, \quad (9)$$

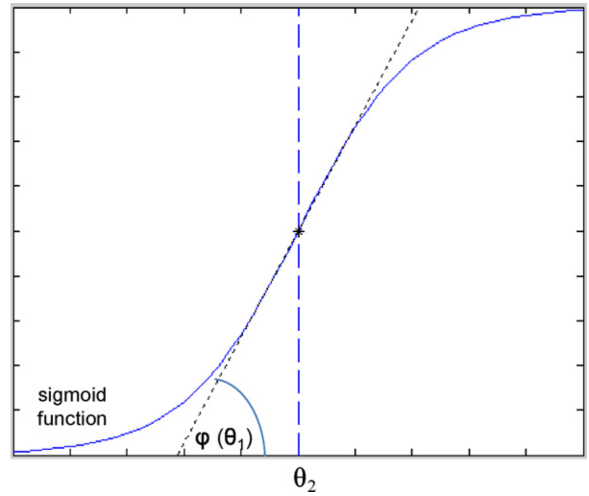

Figure 2 Optimization parameters for the fuzzy membership function (sigmoid — increasing).

then the cost function, used for the optimization, is defined as

$$F(\Theta, D_{\mathrm{train}}) = 1 - \frac{1}{|D_{\mathrm{train}}|}\sum_{i=1}^{n_y} X_{i,i}. \quad (10)$$

The optimization method used is the Healed Topographical Multilevel Single Linkage (HTMLSL) [50], a stochastic algorithm based on MLSL. The algorithm attempts iteratively to find all local minima of an objective function $F(x)$ inside a bounded set $S \subset \mathbb{R}^n$, which are potentially global. These local minima are obtained by a local-search procedure, starting from suitably chosen points in a properly maintained sample. At the kth iteration:

1. Construct a sample selecting at random N points from S and evaluate the objective function at each point;
2. Choose from the sample a subset of points to be used as starting points for local searches;
3. Perform a local search from each starting point. If a new minimum is discovered store it;
4. Determine whether to stop or not. If not, repeat, starting from step 1.

From the stored local minima the one with the lowest value is considered to be the global minimum.

An example of the proposed methodology is presented in Appendix A.

3. Datasets

To create the initial set of rules an annotated dataset is needed. In this work, we have tested the proposed methodology, using two widely known

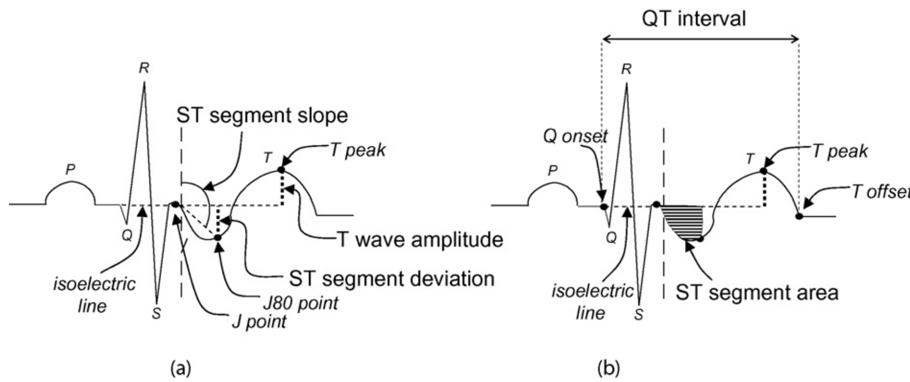

Figure 3 The features extracted from the recordings for ischaemic beat detection: (a) ST segment deviation, ST segment slope and T wave amplitude (b) ST segment area and QT interval.

medical problems: the ischaemic and arrhythmic beat classification. Two benchmark databases were used, the European Society of Cardiology (ESC) ST-T database [51] and the MIT-BIH arrhythmia database [52].

3.1. Signal pre-processing

In some cases, ECG recordings contain significant amount of noise. In order to detect all the relevant ECG characteristics needed to estimate the subsequent features, noise handling must be performed. The QRS complex, which is the most prominent wave in the ECG, is detected for every cardiac beat using the QRS detection method proposed by Tompkins [53,54]. Then, pre-processing of the recorded ECG signal is performed (separately for each lead) in order to eliminate noise distortions (e.g. baseline wandering, A/C interference and electromyographic contamination). Noise elimination is achieved by filtering each recorded cardiac beat separately [16]. Baseline wandering is removed by subtracting from the recorded signal the first-order polynomial that best fits the cardiac beat. A/C interference and electromyographic contamination are not removed from the recorded signal but are handled properly for the detection of the J point. More specifically, for these two types of noise, a 20 ms averaging filter was applied around J point. The exact location of the J point is detected using a technique based on an edge-detection algorithm [55].

3.2. Ischaemic beat classification dataset

In order to construct the dataset for training and testing the ischaemic FES, 11 h of two-channel ECG recordings from the ESC ST-T database [51] are used. Those, contain the whole e0104 recording and the first hour of the e0103, e0105, e0108, e0113, e0114, e0147, e0159, e0162, and e0206 recordings. These 10 recordings are selected because their ischaemic ECG beats are characterized by significant waveform variability. Three medical experts annotated independently each beat as normal, ischaemic or artefact. In case of disagreement the three medical experts reviewed the relevant beat and a decision was taken by consensus. After removing the artefacts and the misdetected beats, the final dataset contained 76,989 cardiac beats, diagnosed as normal or ischaemic.

Several features were extracted from each cardiac beat (Fig. 3). These features were selected according to expert cardiologists [8,44,56]:

- The ST segment deviation (Fig. 3a) refers to the amplitude deviation of the ST segment from the isoelectric line, which is the line defining the level of zero amplitude. The ST segment changes are measured either 80 ms after the J point (J80) (heart rate $\leq$ 120 bpm), or 60 ms after the J point (J60) (heart rate $>$ 120 bpm). Following the ESC recommendations [57] the ST segment deviation is measured relative to a reference waveform for each subject. The reference waveform is calculated using the first 30 s of each recording and is computed by the mean value of the ST segment deviations at this interval respectively.
- The ST segment slope (Fig. 3a) is the slope of the line connecting the J and J80 (or J60) points.
- The ST segment area (Fig. 3a) is the area between the ECG trace, the isoelectric line and the points J and J80 (or J60).
- The T-wave amplitude (Fig. 3b) is the amplitude deviation of the T-wave peak from the isoelectric line. Similarly with the ST segment deviation, the T wave amplitude is measured relative to a reference waveform for each subject which is selected from the first 30 s of each database record.
- The QT interval (Fig. 3b) which is the interval from the beginning of the Q wave (Q_{onset}) to the end of

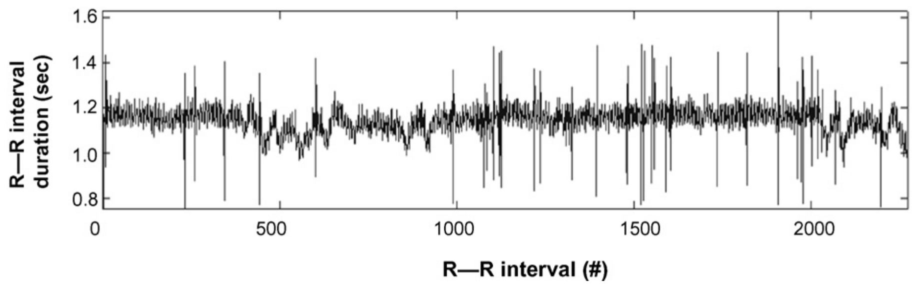

Figure 4 Heart rate variability (HRV) signal (tachogram).

the T wave (T_{offset}). The beginning of the Q wave is determined using the edge detection algorithm mentioned before. For the detection of the T wave end, a 5th order polynomial is fitted to the interval between the peak of the T wave and 0.3*RR seconds after it. Based on the derivative of the fitted function, we can detect the T_{offset} [58]. In order to handle properly the biphasic T waves, a rule followed by Daskalov and Christov [59] has also been considered. Furthermore, the obtained QT has been corrected using an efficient QT correction formula, based on the heart rate variability [45]. The above QT delineator has been tested in the CSE database [60] and reported comparable performance with the method of Daskalov and Christov [59].

In addition to these features a sixth one, the age of the patient, is used. All the above features are considered very relevant for the detection of ischaemic beats. These features are used to create the dataset: $D_{isch} = \{d^l, c^l\}$ with d^l, the lth feature vector and c^l the class of the beat. The class c^l is represented as $c^l \in \{0, 1\}^2$, i.e. $c^l = [0, 1]$ if the beat is normal and $c^l = [1, 0]$ if the beat is ischaemic.

3.3. Arrhythmic beat classification dataset

For training and testing the arrhythmic FES, all beats from all records from the MIT-BIH arrhythmia database [52] are used for the creation of the dataset. Having detected the R waves, the tachogram (Fig. 4) is extracted measuring the time intervals between consecutive R waves. A three RR interval sliding window, is used (RR_1, RR_2, RR_3) as well as functions of those intervals, to create the dataset $D_{arrh} = \{d^l, c^l\}$ with $d^l = (RR_1, RR_2, RR_3, RR_1 + RR_2 + RR_3, RR_1/RR_2, RR_3/RR_1, RR_3/RR_2, |RR_1 - RR_2|, |RR_2 - RR_3|, 2RR_3)/(RR_1 + RR_2), 2RR_1/(RR_2 + RR_3)$, the lth feature vector and c^l the class of the middle RR interval (RR_2). These functions provide useful information of the non-linear relations between the three consecutive RR intervals, related to specific cardiac rhythm patterns, and thus being important for the classification process. The functions have been proposed by expert cardiologists and have been used in previous research attempts [24,27]. The class c^l is represented as $c^l \in \{0, 1\}^4$, where, if d^l belongs to class i, then $c^l = e_i$. Both rhythm and beat annotations from the database are used to specify the class, following the scheme: if RR_2 is annotated as ventricular flutter/fibrillation (VF), then $c^l = [1, 0, 0, 0]$, else if RR_2 is annotated as premature ventricular contraction (PVC)[1] then $c^l = [0, 1, 0, 0]$, else if RR_2 belongs to 2° heart block episode (BII), then $c^l = [0, 0, 0, 1]$, else RR_2 is considered as normal (N) and $c^l = [0, 0, 1, 0]$. The above resulted in 109,880 beats.

4. Results

In the case of ischaemic beat detection, from the 76,989 beats, we used 1936 beats (954 ischaemic and 982 normal) for training the ischaemic FES and the rest 75,053 (36,709 ischaemic and 38,344 normal) beats for testing it (Table 1). The sampling of the 76,989 beats for acquiring the training ones was performed by selecting iteratively the first beat out of a sequence of 40 ones. In this way, beats from all recordings were used both for training and testing (global training).

For training and testing the arrhythmic FES, we followed a different strategy due to the large imbalance in the distribution of classes. In order to select training and test sets in highly imbalanced datasets, three approaches can be followed: oversampling, undersampling or hybrid sampling. However, both oversampling and hybrid sampling tend to give overfitted models [40]. For this reason, in order to train the arrhythmic FES, undersampling was employed. Three hundred beats from each category, randomly selected, were used for training the arrhythmic FES (1200 beats) and the remaining beats from all categories for testing it (108,680 beats). Table 1 presents

[1] Isolated PVCs, as well as, runs of PVCs are included.

Table 1 Number of beats used for training and testing the FESs for ischaemia and arrhythmia

Disease	Classes	Train	Test	Overall
Ischaemia	Ischaemic	954	36,709	37,663
	Normal	982	38,344	39,326
	Overall	1936	75,053	76,989
Arrhythmia	VF	300	184	484
	PVC	300	5,883	6183
	N	300	102,493	102,793
	BII	300	120	420
	Overall	1200	108,680	109,880

the training and test sets for each class. As it is mentioned above, the training beats were used both for the crisp model development and the parameter optimization.

The above-described datasets are used to evaluate our methodology. In the first stage of the methodology, the set of rules extracted from the decision tree consists of 53 rules (ischaemic FES), from which 27 predicted normal beats and the rest 26 predicted ischaemic beats. In the case of the arrhythmic FES, 17 rules are generated: 2 of them have as consequent the VF category, 7 the PVC category, 7 the N category and one rule predicted the BII category (Table 2). Indicative crisp rules from both application domains are presented below (one rule for each class of the classification problems):

Indicative rules for Ischaemia[2]

$$\text{if}\left\{\begin{array}{c}\text{ST segment area} > 0.9705 \text{ AND ST segment area} \leq 1.6802 \text{ AND}\\ \text{T wave amplitute} > -0.1917 \text{ AND T wave amplitude} \leq 0.218 \text{ AND}\\ \text{ST segment slope} > 53.45 \text{ AND}\\ \text{Age} > 47 \text{ AND Age} \leq 65\end{array}\right\}$$

then {Beat is Isch}

$$\text{if}\left\{\begin{array}{c}\text{T wave amplitude} > -1.3307 \text{ AND T wave amplitude} \leq 0.1628 \text{ AND}\\ \text{ST segment area} > -0.7349 \text{ AND ST segment area} \leq 0.9705 \text{ AND}\\ \text{ST segment deviation} > -0.0103 \text{ AND QT interval} \leq 1.427 \text{ AND}\\ \text{Age} > 60 \text{ AND Age} \leq 62\end{array}\right\}$$

then {Bean is Norm}

Indicative rules for Arrhythmia

$$\text{if}\left\{\begin{array}{c}\text{RR}_2 \leq 1.464 \text{ AND}\\ \text{RR}_1 + \text{RR}_2 + \text{RR}_3 \leq 1.377\end{array}\right\}\text{then}\{\text{Beat is } VF\}$$

$$\text{if}\left\{\begin{array}{c}\text{RR}_1 + \text{RR}_2 + \text{RR}_3 > 1.377 \text{ AND}\\ \text{RR}_2 > 0.358 \text{ AND RR}_2 \leq 0.656\\ \text{RR}_3/\text{RR}_1 > 1.1484 \text{ AND}\end{array}\right\}\text{then}\{\text{Beat is } PVC\}$$

$$\text{if}\left\{\begin{array}{c}\text{RR}_2 \leq 1.464 \text{ AND}\\ \text{RR}_3/\text{RR}_1 \leq 1.1484 \text{ AND}\\ \text{RR}_3/\text{RR}_1 > 1.1484 \text{ AND}\\ 2\text{RR}_3/(\text{RR}_1 + \text{RR}_2) \leq 1.14173 \text{ AND}\\ \text{RR}_1 + \text{RR}_2 + \text{RR}_3 > 1.722\end{array}\right\}\text{then}\{\text{Beat is } N\}$$

if $\{\text{RR}_2 > 1.461\}$ then {Bean is BII}

Table 3 displays the normalized confusion matrix for ischaemic beat classification, performed using only the initial set of rules extracted from the decision tree. The obtained sensitivity (Se) and specificity

[2] ST segment deviation and T wave amplitude are measured in millivolt, ST segment slope is measured in degrees, ST Segment area in milivolt seconds, QT interval and RR interval in seconds and age is measured in years.

Table 2 Number of rules extracted from the decision trees for ischaemia and arrhythmia FESs

Disease	Classes	No. rules
Ischaemia	Ischaemic	26
	Normal	27
	Overall	53
Arrhythmia	VF	2
	PVC	7
	N	7
	BII	1
	Overall	17

Table 3 Confusion matrix, sensitivity (Se) and specificity (Sp) for ischaemic beat detection using only the 1st stage (decision tree) and the ischaemic FES

	First stage only classified as		Three-stage methodology classified as	
	Isch	Norm	Isch	Norm
Database				
Isch	0.907	0.093	0.912	0.088
Norm	0.100	0.900	0.078	0.922
Metrics (%)				
Se	90.7		91.2	
Sp	90.0		92.2	
Acc	90.4		91.7	

(Sp) are 90.7% and 90%, respectively. In addition, Table 3 presents the normalized confusion matrix for the ischaemic FES; in the latter, the sensitivity and specificity are increased to 91.2% and 92.2%, respectively. The application of the methodology in the ischaemic beat detection problem misclassified 3226 ischaemic and 2989 normal beats.

Table 4 presents the normalized confusion matrix for arrhythmic beat classification, again employing only the initial set of rules and then using the three stage methodology (arrhythmic FES). Using only the initial set of rules, the sensitivity and specificity is 97.3% and 98.8% for the VF category, 89.1% and 96.6% for the PVC category, 91.9% and 97% for the N category, 98.3% and 99.8% for the BII category, respectively. The above results are improved when all stages of the methodology are used. More specifically, the sensitivity and specificity is 98.9% and 99.3% for the VF category, 92.4% and 97.6% for the PVC category, 93.6% and 97.7% for the N category, 98.3% and 99.9% for the BII category, respectively. The results for the VF and BII categories are very high, while there is high misclassification rate between the PVC and N categories; 362 PVC beats were misclassified as N (6.15%) and 5453N beats were misclassified as PVC (5.32%).

From the obtained results it is clear that the application of the proposed methodology improved the efficiency of the induced decision trees, for both ischaemic and arrhythmic beat classification. The ischaemic FES improved the accuracy of the decision tree by 1.3%, while the respective improvement for the arrhythmic FES is 1.6%. The number of beats in the test set is sufficiently large, thus the error rates, defined as: $e = 1 - \text{acc}$, of the decision trees and the FESs in both cases (i.e. ischaemic and arrhythmic beat classification) can be approximated using normal distributions [40]. If the observed difference in e is defined as $d = |e_{\text{FES}} - e_{\text{DT}}|$, where e_{FES} is the error rate of the FES and e_{DT} is the error rate of the decision tree, then d is also normally distributed, with variance: $\sigma_d^2 = (\text{acc}_{\text{DT}}(1 - \text{acc}_{\text{DT}}) + \text{acc}_{\text{FES}}(1 - \text{acc}_{\text{FES}}))/N$, where N the number of test records (i.e. number of beats), acc_{DT} is the accuracy of the decision tree and acc_{FES} is the accuracy of the FES. 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: $d_t = d \pm 1.96\sigma_d$. For ischaemic beat classification, the confidence interval for d_t at 95% confidence level is

Table 4 Confusion matrix, sensitivity (Se) and specificity (Sp) for all categories of the arrhythmic beat classification using only the 1st stage (decision tree) and the arrhythmic FES

	First stage only classified as				Three-stage methodology classified as			
	VF	PVC	N	BII	VF	PVC	N	BII
Database								
VF	0.973	0.027	0.000	0.000	0.989	0.011	0.000	0.000
PVC	0.026	0.891	0.083	0.000	0.014	0.924	0.062	0.000
N	0.009	0.065	0.919	0.006	0.006	0.053	0.936	0.004
BII	0.000	0.008	0.008	0.983	0.000	0.008	0.008	0.983
Metrics (%)								
Se	97.3	89.1	91.9	98.3	98.9	92.4	93.6	98.3
Sp	98.8	96.6	97.0	99.8	99.3	97.6	97.7	99.9
Acc	94.2				95.8			

1.3 ± 0.43, which does not spam the zero value and thus the observed difference is statistically significant. Similarly, for arrhythmic beat classification, the confidence interval for d_t at 95% confidence level is 1.6 ± 0.32, which also does not spam the zero value and thus the observed difference is statistically significant.

5. Discussion

In this study, we propose a methodology for the automated creation of fuzzy expert systems that consists of three stages: (i) extraction of a set of rules using a decision tree, (ii) transformation of the set of rules into a fuzzy model, and (iii) optimization of the fuzzy model's parameters using global optimization. The proposed methodology has been evaluated in the detection of ischaemic cardiac beats in ECG recordings using data from the ESC ST-T database. Also, it has been evaluated in arrhythmic beat classification, using data from the MIT-BIH arrhythmia database. In both cases high classification results were obtained; the accuracy (Acc) is 92% and 96% for the ischaemic and arrhythmic FES, respectively.

The proposed methodology is innovative since it combines data mining techniques with fuzzy modelling and introduces several novelties. It is generic and thus it can be applied to any classification domain; given an initial annotated dataset, it can automatically generate a FES. This FES is based on a set of fuzzy rules and thus it is able to provide interpretation for its decisions. This is a highly desirable feature, since the ability to explain the reason for a decision is of great value for the domain experts. In addition, the employment of data mining (decision trees) in the first stage of the methodology has the advantage of discovering new knowledge [40,41]. It should me mentioned that the proposed methodology can incorporate in the first stage any rule mining technique. In the current work we employed decision trees with the C4.5 algorithm which is widely used and is considered as a very effective approach for classification. Also, the introduction of the fuzzy models addresses the uncertainty inherent in several medical problems [30]. The development of the fuzzy model from the initial set of rules and the optimization of its parameters improves the efficiency of the decision tree.

Thus, in the case of ischaemic beat detection, the performance (accuracy) is improved by 1.3% and in the case of arrhythmic beat classification the performance is improved by 1.6%. Finally, representative features from the cardiac beats are extracted and they are used for both FESs: features from the ST-T interval, which is of known ischaemic diagnostic value, are used for ischaemic beat detection while features from the tachogram, which is appropriate to characterize the types of arrhythmias that are under consideration in this study [27], are employed for arrhythmic beat classification.

In what concerns ischaemic beat detection, in Table 5 the results of the proposed ischaemic FES are compared to those of other similar approaches; our approach shows slightly better performance. These methods were tested using data form the ESC ST-T database, which is a standard reference for myocardial ischaemia detection [2,3]. However, some of the results reported in the literature refer to different subsets of ECG recordings of the ESC ST-T database [10,11,13] or have used different databases for their evaluation [5,6], and thus, their performance cannot be directly compared. It should be noted that in Ref. [13] a different subset of the ESC ST-T database was employed to evaluate the ischaemic beat classifier. More specifically, it was considered that each annotated episode in the database contains only ischaemic beats. In addition, most of these techniques are based on neural or signal processing approaches; such methods exhibit a serious drawback compared to our rule-based approach, due to

Table 5 Comparison of the performance of several methods for ischaemic beat detection evaluated using the ESC ST-T database

Method	Se (%)	Sp (%)	Acc (%)
Rule-based [7]	70	63	
ANN & PCA [9]	90	90	
Bidirectional associative memories ANN [10]			56
ANN (classification partitioning-SOM & SVM) [11]			80
Feed forward ANN and nonlinear PCA [13]	79	75	
Multicriteria decision analysis [14]	90	89	
Genetic algorithms & multicriteria decision analysis [15]	91	91	
Association rule mining [16]	87	93	90
Current work	91	92	92

Table 6 Comparison of the performance of several methods for arrhythmic beat classification evaluated using the MIT-BIH arrhythmia database

Method	Acc (%)
PCA & mixture of experts approach (SOM, LVQ) [18]	95.5
Hermite functions & SOM [19]	98.5
Discrete wavelet transform & intersecting spheres network [20]	96
Second, third and fourth order cumulants & hybrid ANN [22]	96
Autoregressive modelling [23]	97
SVM [25]	96
ECG morphology & linear discriminates [26]	97.5
Knowledge-based system [27]	94
Current work	96

their inability to provide clear and direct explanations for their classification decisions [61]. This is of great importance when developing medical decision support systems that will assist physicians in the diagnosis.

Table 6 presents several methods proposed in the literature for arrhythmic beat classification, along with the reported accuracy. The accuracy obtained from those methods is in the range from 94% to 98.5%. The methods reported in Refs. [18—23,25] are based on ''black box'' approaches, such as neural networks and support vector machines. Therefore, there is no exact interpretation for their results [61]. In our approach each decision can be interpreted in a medical manner. In the proposed methodology, only QRS detection was performed, on the ECG signal and the analysis is based on the RR intervals. Several of the methods proposed in the literature are based on the analysis of the ECG signal (Dokur et al. [20], Osowski et al. [22,25], Hu et al. [18], Lagerholm et al. [19]), which is much more time-consuming than the proposed method. Also, it is advantageous compared to other approaches which use morphological ECG features [26], which are not feasible in cases of high noise. In [18] initial labelling of the beats was required and there was no automatic QRS detection—the points of the database annotation were used. The method was evaluated using the last 25 min of the records in the 200 series, apart from records 212, 217, 220, 222 and 232. In Ref. [19] all MIT-BIH arrhythmia database records were used for evaluation but the primary objective was to perform clustering with an expert performing the final beat classification. In the present work four beat categories are automatically classified, without any human interference, in contrast to Refs. [18,19]. In addition, some of the proposed methods have been tested on small subsets of the MIT-BIH arrhythmia database [20,22,23,25], while our results were obtained using all records from the MIT-BIH arrhythmia database for evaluation.

A limitation of our methodology is the requirement of a representative training set in order to extract reliable rules and thus create a reliable fuzzy model. In addition, the utilization of decision rules for classification, besides finding valid, causal relationships in the clinical data, will also find all of the spurious and particular relationships among the data in the specific dataset. For this reason, results of any data mining procedure should be considered as exploratory and hypothesis-generating. Regarding the arrhythmic beat classification problem, the RR interval signal was used, thus limiting the arrhythmic categories to only those that affect the physiological RR intervals. Future work will also include other types of arrhythmias, i.e. atrial arrhythmias.

Since the proposed methodology is generic, different approaches can be employed for all three stages. Future work will focus on the use of other rule mining techniques (C5.0, association rule mining), different definition of the fuzzy model (other fuzzification functions, inference engines and defuzzification approaches) and employment of alternative optimization techniques (global or local).

6. Conclusions

We presented a novel methodology for the automated creation of FESs. The main advantage of the methodology is the combination of high accuracy with the ability to provide interpretation for the decisions made. The generated FESs for ischaemic and arrhythmic beat classification compare well with previously reported results, indicating that they could be part of an overall clinical system for ECG analysis and diagnosis. However, more clinical testing is needed in order to be fully evaluated.

Acknowledgments

This research is part funded 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 and the European Social Fund.

Appendix A

In this appendix we provide a working example of our methodology.

In the first stage, having the initial annotated dataset, with three features ($n_f = 3$): $A = \{a_1, a_2, a_3\}$ and two classes (($n_y = 2$)), we create a decision tree, parse it and create the following set of rules:

$$\text{if}(a_1 > \theta_1 \wedge a_2 \leq \theta_2) \quad \text{then} \;\; c = 1, \qquad \text{if}(a_2 > \theta_3 \wedge a_3 \leq \theta_4) \quad \text{then} \;\; c = 1,$$
$$\text{if}(a_1 > \theta_5 \wedge a_3 > \theta_6) \quad \text{then} \;\; c = 2,$$

where $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6\}$ is the vector containing all thresholds used in the tree (without loss of generality we have not included a_{root} and θ_{root}). The crisp model contains three conditions:

$$\begin{aligned}
\text{Cond}_1(A, \Theta) &: (g_c^{\text{inc}}(a_1, \theta_1) \wedge g_c^{\text{dec}}(a_2, \theta_2)), \quad \text{with Cond}_1(A, \Theta) \to 1, \qquad \text{Cond}_2(A, \Theta) \\
&: (g_c^{\text{inc}}(a_2, \theta_3) \wedge g_c^{\text{dec}}(a_3, \theta_4)), \quad \text{with Cond}_2(A, \Theta) \to 1, \qquad \text{Cond}_3(A, \Theta) \\
&: (g_c^{\text{inc}}(a_1, \theta_5) \wedge g_c^{\text{inc}}(a_3, \theta_6)), \quad \text{with Cond}_3(A, \Theta) \to 2.
\end{aligned}$$

Therefore, the crisp model contains two general crisp rules (one for each class):

$$R_1(A, \Theta) = (g_c^{\text{inc}}(a_1, \theta_1) \wedge g_c^{\text{dec}}(a_2, \theta_2)) \vee (g_c^{\text{inc}}(a_2, \theta_3) \wedge g_c^{\text{dec}}(a_3, \theta_4)), \qquad R_2(A, \Theta) = (g_c^{\text{inc}}(a_1, \theta_5) \wedge g_c^{\text{inc}}(a_3, \theta_6)).$$

In the second stage, the fuzzy model is created, fuzzifying the crisp conditions:

$$\begin{aligned}
\text{Cond}_1^f(A, \Theta^f) &: \min(g_s^{\text{inc}}(a_1, \theta_{1,1}, \theta_{2,1}), g_s^{\text{dec}}(a_2, \theta_{1,2}, \theta_{2,2})), \qquad \text{Cond}_2^f(A, \Theta^f) \\
&: \min(g_s^{\text{inc}}(a_2, \theta_{1,3}, \theta_{2,3}), g_s^{\text{dec}}(a_3, \theta_{1,4}, \theta_{2,4})), \qquad \text{Cond}_3^f(A, \Theta^f) \\
&: \min(g_s^{\text{inc}}(a_1, \theta_{1,5}, \theta_{2,5}), g_s^{\text{inc}}(a_3, \theta_{1,6}, \theta_{2,6})),
\end{aligned}$$

and thus, the general fuzzy rules are:

$$\begin{aligned}
R_1^f(A, \Theta^f) &= \max\begin{pmatrix} p_1 \min(g_s^{\text{inc}}(a_1, \theta_{1,1}, \theta_{2,1}), g_s^{\text{dec}}(a_2, \theta_{1,2}, \theta_{2,2})), \\ p_2 \min(g_s^{\text{inc}}(a_2, \theta_{1,3}, \theta_{2,3}), g_s^{\text{dec}}(a_3, \theta_{1,4}, \theta_{2,4})) \end{pmatrix}, \qquad R_2^f(A, \Theta^f) \\
&= p_3 \min(g_s^{\text{inc}}(a_1, \theta_{1,5}, \theta_{2,5}), g_s^{\text{inc}}(a_3, \theta_{1,6}, \theta_{2,6})),
\end{aligned}$$

with $\Theta^f = \{\theta_{1,1}, \theta_{2,1}, \ldots, \theta_{1,6}, \theta_{2,6}\}$ being the parameter set of the fuzzy model and $p = [p_1, p_2, p_3]$ the likelihood ratio of each rule. The fuzzy model is then created as follows:

$$M^f(A, \Theta^f) = \arg \max_{y=1,\ldots,n_y} (R_y^f(A, \Theta^f)) = \arg\max \begin{pmatrix} \max\begin{pmatrix} p_1 \min(g_s^{\text{inc}}(a_1, \theta_{1,1}, \theta_{2,1}), g_s^{\text{dec}}(a_2, \theta_{1,2}, \theta_{2,2})), \\ p_2 \min(g_s^{\text{inc}}(a_2, \theta_{1,3}, \theta_{2,3}), g_s^{\text{dec}}(a_3, \theta_{1,4}, \theta_{2,4})) \end{pmatrix}, \\ p_3 \min(g_s^{\text{inc}}(a_1, \theta_{1,5}, \theta_{2,5}), g_s^{\text{inc}}(a_3, \theta_{1,6}, \theta_{2,6})) \end{pmatrix}.$$

Finally, in the third stage, $M^f(A, \Theta^f)$ is optimized with respect to Θ^f and the fuzzy expert system is defined as follows:

$$M^f(A, \Theta^{f*}) = \arg\max \begin{pmatrix} \max\begin{pmatrix} p_1 \min(g_s^{\text{inc}}(a_1, \theta_{1,1}^*, \theta_{2,1}^*), g_s^{\text{dec}}(a_2, \theta_{1,2}^*, \theta_{2,2}^*)), \\ p_2 \min\left(g_s^{\text{inc}}(a_2, \theta_{1,3}^*, \theta_{2,3}^*), g_s^{\text{dec}}(a_3, \theta_{1,4}^*, \theta_{2,4}^*)\right) \end{pmatrix}, \\ p_3 \min(g_s^{\text{inc}}(a_1, \theta_{1,5}^*, \theta_{2,5}^*), g_s^{\text{inc}}(a_3, \theta_{1,6}^*, \theta_{2,6}^*)) \end{pmatrix},$$

where $\Theta^{f*} = \{\theta_{1,1}^*, \theta_{2,1}^*, \ldots, \theta_{1,6}^*, \theta_{2,6}^*\}$ is the set of the optimized parameters.

References

[1] Goldman MJ. Principles of clinical electrocardiography, 11th ed., Los Altos, CA: LANGE Medical Publications; 1982.

[2] Jager F, Moody GB, Mark RG. Characterization of transient ischemic and non-ischemic ST segment changes. In: Murray A, editor. Proceedings of the computers in cardiology. 1995. p. 721—4.

[3] Jager F, Mark RG, Moody GB, Divjak S. Analysis of transient events ST segment changes during ambulatory monitoring using the Karhunen—Loeve transform. In: Murray A, editor. Proceedings of the computers in cardiology. 1992. p. 691—4.

[4] Langley P, Bowers ET, Wild J, Drinnan MJ, Allen J, Sims AJ, et al. An algorithm to distinguish ischaemic and non-ischaemic ST changes in the Holter ECG. In: Murray A, editor. Proceedings of the computers in cardiology. 2003. p. 235—8.

[5] Zahan S. A fuzzy approach to computer-assisted myocardial ischemia diagnosis. Artif Intell Med 2001;21:271—5.
[6] Senhadji L, Carrault G, Bellanger JJ, Passariello G. Comparing wavelet transforms for recognizing cardiac patterns. IEEE Eng Med Biol Mag 1995;14:167—73.
[7] Papaloukas C, Fotiadis DI, Liavas AP, Likas A, Michalis LK. A knowledge-based technique for automated detection of ischemic episodes in long duration electrocardiograms. Med Biol Eng Comput 2001;39:105—12.
[8] Papaloukas C, Fotiadis DI, Likas ACS, Stroumbis CS, Michalis LK. 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 2002;35:27—34.
[9] Papaloukas C, Fotiadis DI, Likas A, Michalis LK. An ischemia detection method based on artificial neural networks. Artif Intell Med 2002;24:167—78.
[10] Maglaveras N, Stamkopoulos T, Pappas C, Strintzis M. ECG processing techniques based on neural networks and bidirectional associative memories. J Med Eng Technol 1998;22:106—11.
[11] Papadimitriou S, Mavroudi S, Vladutu L, Bezerianos A. Ischemia detection with a self-organizing map supplemented by supervised learning. IEEE Trans Neural Networks 2001;12:503—15.
[12] Maglaveras N, Stamkopoulos T, Pappas C, Strintzis M. 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 1998;45(7):805—13.
[13] Stamkopoulos T, Diamantaras K, Maglaveras N, Strintzis M. ECG analysis using nonlinear PCA neural networks for ischemia detection. IEEE Trans Signal Process 1998;46:3058—67.
[14] Goletsis Y, Papaloukas C, Fotiadis DI, Likas A, Michalis LK. A multicriteria decision based approach for ischemia detection in long duration ECGs. In: Proceedings of the IEEE EMBS 4th international conference on information technology applications in biomedicine; 2003. p. 230—3.
[15] Goletsis Y, Papaloukas C, Fotiadis DI, Likas A, Michalis LK. Automated ischemic beat classification using genetic algorithms and multicriteria decision analysis. IEEE Trans Biom Eng 2004;51:1717—25.
[16] Exarchos TP, Papaloukas C, Fotiadis DI, Michalis LK. An association rule mining based methodology for the automated detection if ischemic ECG beats. IEEE Trans Biomed Eng 2006;53(8):531—1540.
[17] Sandoe E, Sigurd B. Arrhythmia—a guide to clinical electrocardiology. Bingen: Publishing Partners Verlags GmbH; 1991.
[18] Hu YZ, Palreddy S, Tompkins WJ. A patient-adaptable ECG beat classifier using a mixture of experts approach. IEEE Trans Biomed Eng 1997;44:891—900.
[19] Lagerholm M, Peterson C, Braccini G, Ebendrandt L, Sornmo L. Clustering ECG complexes using Hermite functions and self-organizing maps 2000. IEEE Trans Biomed Eng 2000;47:838—48.
[20] Dokur Z, Olmez T. ECG beat classification by a hybrid neural network. Comput Meth Prog Biomed 2001;66:167—81.
[21] Acharya UR, Bhat PS, Iyengar SS, Rao A, Dua S. Classification of heart rate data using artificial neural network and fuzzy equivalence relation. Pat Recogn 2003;36:61—8.
[22] Osowski S, Linh TH. ECG beat recognition using fuzzy hybrid neural network. IEEE Trans Biom Eng 2001;48:1265—71.
[23] Ge D, Srinivasan N, Krishnan SM. Cardiac arrhythmia classification using autoregressive modelling. Biomed Eng OnLine 2002;1:5.
[24] Tsipouras MG, Fotiadis DI, Sideris D. Arrhythmia classification using the RR-interval duration signal. In: Murray A, editor. Proceedings of the computers in cardiology. 2002. p. 485—8.
[25] Osowski S, Linh TH, Markiewicz T. Support vector machine based expert system for reliable heartbeat recognition. IEEE Trans Biom Eng 2004;51:582—9.
[26] Chazal P, O'Dwyer M, Reilly RB. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. IEEE Trans Biom Eng 2004;51:1196—206.
[27] Tsipouras MG, Fotiadis DI, Sideris D. An arrhythmia classification system based on the RR interval signal. Artif Intell Med 2005;33:237—50.
[28] Giartano J, Riley G.In: Expert systems, principles and programming3rd ed., PWS Publishing Company; 1998.
[29] Tsoukalas LH, Uhrig RE. Fuzzy and neural approaches in engineering. John Wiley & Sons; 1997.
[30] Akay M. Nonlinear biomedical signal processing. Vol. 1: fuzzy logic, neural networks and new algorithm. IEEE Press series on biomedical engineering. New York, USA: John Wiley & Sons; 2000.
[31] Pedrycz W, Oliveira JV. An algorithmic framework for development and optimization of fuzzy models. Fuzzy Sets Syst 1996;80:37—55.
[32] Garibalti JM, Ifeachor EI. Application of simulated annealing fuzzy model tuning to umbilical cord acid—base interpretation. IEEE Trans Fuzzy Syst 1999;7:72—84.
[33] Babuska R, Verbruggen H. Neuro-fuzzy methods for nonlinear system identification. Annu Rev Control 2003;27:73—85.
[34] Janikow CZ. Fuzzy decision trees: issues and methods. IEEE Trans Syst Man Cybernet 1998;28(1):1—14.
[35] Olaru C, Wehenkel L. A complete fuzzy decision tree technique. Fuzzy Sets Syst 2003;138:221—54.
[36] Janikow CZ. FID4.1. Fuzzy information processing. NAFIPS 2004;2:877—81.
[37] Pedrycz W, Sosnowski ZA. Genetically optimized fuzzy decision trees. IEEE Trans Syst Man Cybernet B Cybernet 2005;35:633—41.
[38] Crockett K, Bandar Z, O'Shea J, Mclean D. On constructing a fuzzy inference framework using crisp decision trees. Fuzzy Sets Syst 2006;157:2809—32.
[39] Crockett K, Bandar Z, O'Shea J, Fowdar J. Genetic tuning of fuzzy inference within fuzzy classifier systems. Int J Expert Syst 2006;23.
[40] Tan PN, Steinbach M, Kumar V. Introduction to data mining. Addison Wesley; 2005.
[41] Kantardzic M. Data mining: concepts, models, methods, and algorithms. Wiley—IEEE Press; 2002.
[42] Quinlan JR. C4.5: programs for machine learning. California: Morgan Kauffman; 1993.
[43] Quinlan JR. Improved use of continuous attributes in C4.5. J Artif Intell Res 1996;4:77—90.
[44] O'keefe JH, Hammill SC, Freed M. The ECG criteria and ACLS handbook. Physician's Press; 1998/2002.
[45] Luo S, Michler K, Johnston P, Macfarlane PW. A comparison of commonly used QT correction formulae: the effect of heart rate on the QTc of normal ECGs. J Electrocardiol 2004;37:81—90.
[46] Silipo R, Laguna P, Marchesi C, Mark RG. ST-T segment change recognition using artificial neural networks and principal component analysis. In: Murray A, editor. Proceedings of the IEEE computers in cardiology. 1995. p. 213—6.
[47] Jager F, Moody GB, Taddei A, Mark RG. Performance measures for algorithms to detect transient ischemic ST segment changes. In: Murray A, editor. Proceedings of the IEEE computers in cardiology. 1991. p. 369—72.
[48] Malik M, Camm AJ. Heart rate variability. Armonk, New York: Futura Publishing Company; 1995.

[49] Wang LX. A course in fuzzy systems and control. Prentice-Hall; 1986.
[50] Theos FV, Lagaris IE, Papageorgiou DG. PANMIN: sequential and parallel global optimization procedures with a variety of options for the local search strategy. Comp Phys Commun 2004;159:63—9.
[51] European Society of Cardiology, European ST-T database directory. Pisa: S.T.A.R; 1991.
[52] MIT-BIH arrhythmia database CD-ROM, 3rd ed., Harvard-MIT Division of Health Sciences and Technology; 1994.
[53] Tompkins WJ. Biomedical digital signal processing (C-language examples and laboratory experiments for the IBM® PC). New Jersey: Prentice-Hall, Englewood Cliffs; 1993.
[54] Hamilton PS, Tompkins WJ. Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database. IEEE Trans Biom Eng 1986;33:1157—65.
[55] Daskalov K, Dotsinsky IA, Christov II. Developments in ECG acquisition, preprocessing, parameter measurement, and recording. IEEE Eng Med Biol 1998;17:50—8.
[56] Macfarlane PW, Browne D, Devine B, Clark E, Miller E, Seyal J, et al. Effect of age and gender on diagnostic accuracy of ECG diagnosis of acute myocardial infarction. In: Murray A, editor. Proceedings of the computers in cardiology. 2004. p. 165—8.
[57] Taddei A, Benassi A, Bongiorni MG, Contini C, Distante G, Landucci L, et al. ST-T changes analysis in ECG ambulatory monitoring: a European standard for performance evaluation. In: Proceedings of the IEEE computers in cardiology; 1988. p. 63—8.
[58] Exarchos CP. Automatic calculation of QT interval, Diploma Thesis, Department of Computer Science, University of Ioannina; 2006.
[59] Daskalov IK, Christov II. Automatic detection of the electrocardiogram T-wave end. Med Biol Eng Comput 1999;37:348—53.
[60] Willems JL, Arnaud P, Van Bemmel JH, Bourdillon PJ, Degani R, Denis B, et al. A reference database for multilead electrocardiographic computer measurement programs. J Am Coll Cardiol 1987;10:1313.
[61] Kecman V. Learning and soft computing: support vector machines, neural networks and fuzzy logic models. Cambridge, MA: MIT press; 2000.