Examining cross-database global training to evaluate five different methods for ventricular beat classification

2009 Physiol. Meas. 30 661

(http://iopscience.iop.org/0967-3334/30/7/010)

IOP Publishing Physiological Measurement

Physiol. Meas. **30** (2009) 661–677 doi:10.1088/0967-3334/30/7/010

Examining cross-database global training to evaluate five different methods for ventricular beat classification

V Chudáček[1], G Georgoulas[2], L Lhotská[1], C Stylios[2], M Petrík[1] and M Čepek[3]

[1] Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic
[2] Department of Informatics and Telecommunication Technology, TEI of Epirus, Arta, Greece
[3] Department of Computer Science and Engineering, Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic

E-mail: chudacv@fel.cvut.cz

Received 6 March 2009, accepted for publication 14 May 2009
Published 12 June 2009
Online at stacks.iop.org/PM/30/661

Abstract

The detection of ventricular beats in the holter recording is a task of great importance since it can direct clinicians toward the parts of the electrocardiogram record that might be crucial for determining the final diagnosis. Although there already exists a fair amount of research work dealing with ventricular beat detection in holter recordings, the vast majority uses a local training approach, which is highly disputable from the point of view of any practical—real-life—application. In this paper, we compare five well-known methods: a classical decision tree approach and its variant with fuzzy rules, a self-organizing map clustering method with template matching for classification, a back-propagation neural network and a support vector machine classifier, all examined using the same global cross-database approach for training and testing. For this task two databases were used—the MIT–BIH database and the AHA database. Both databases are required for testing any newly developed algorithms for holter beat classification that is going to be deployed in the EU market. According to cross-database global training, when the classifier is trained with the beats from the records of one database then the records from the other database are used for testing. The results of all the methods are compared and evaluated using the measures of sensitivity and specificity. The support vector machine classifier is the best classifier from the five we tested, achieving an average sensitivity of 87.20% and an average specificity of 91.57%, which outperforms nearly all the published algorithms when applied in the context of a similar global training approach.

Keywords: ECG, classification, holter, premature ventricular complex, rule-based tree, fuzzy rule-based tree, neural networks, self-organizing map, support vector machines

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Long-term holter monitoring (Holter 1961) is usually used for patients with suspected heart problems such as arrhythmias. Identification of heart beats with unusual timing or unusual electrocardiogram (ECG) morphology is necessary for early diagnosis of hearts with pathological electrophysiology. The problem of identifying such beats is often transformed into the problem of discriminating between normal 'N' and premature ventricular 'V' beats.

Solving this problem is of major importance for proper diagnostics of heart problems since the magnitude of incidence of ventricular beats can indicate the possibility of onset of ventricular fibrillation and sequential sudden cardiac death. An advanced system with automatic classification and analysis capabilities is needed in order to process the large amounts of data the holter measurement creates.

Many different approaches have been taken to represent the beats using features such as energy of QRS complex (Acar 2000), beat-shape description parameters (Chazal *et al* 2004, Christov 2004, Christov *et al* 2006), features obtained from heartbeat interval measurements (Tsipouras *et al* 2002), fractal features (Bakardijan 1992), shape descriptive parameters transformed with the Karhunen–Loeve transform (Moody and Mark 1989, Jager 2002) or Hermite polynomials (Lagerholm *et al* 2000). Other works have used independent components (Yu and Chou 2009) or time-frequency features (Christov *et al* 2006, Minhas and Arif 2008) for beat representation. An example of the ECG beat with beat-shape description parameters is depicted in figure 1.

A wide variety of methods have been proposed for classification of holter beats such as linear discriminants (Bortolan *et al* 2005), nearest-neighbor classifiers (Christov *et al* 2005), neural networks (Hosseini *et al* 2006, Ceylan *et al* 2009), self-organizing maps (SOMs) (Lagerholm *et al* 2000), support vector machines (SVMs) (Tsipouras *et al* 2005), clustering (Cuesta-Frau *et al* 2007), genetic algorithms (Olmez *et al* 1997), inductive logic programming (Carrault *et al* 2003) and fuzzy expert systems (Tsipouras *et al* 2007).

To be able to compare the results described in different papers one needs to know the exact settings that each of the authors used for his/her classifier training. There are in general two ways of tackling the training phase. Local training on the ground that the algorithm will be used in such a setting where each patient will have at least part of his/her holter-ECG record annotated by an expert clinician, and it is employed by the majority of the papers (Tsipouras *et al* 2005, Christov *et al* 2006). The disputable presumption of the preceding expert annotation might be nevertheless fulfilled in the telemedicine application setting as reported in Kyriacou *et al* (2003). The second option is global training, an approach where the records used for training the classifier are distinct from those used for testing ensuring therefore better, if the data are representative, generalization. The global classifier was used e.g. in Chazal *et al* (2004), Dotsinsky and Stoyanov (2004) and Jekova *et al* (2008) and it usually leads to worse overall results for the classifier performance. On the other hand, these classifiers should be more robust performing evenly on any other unknown data. Since the

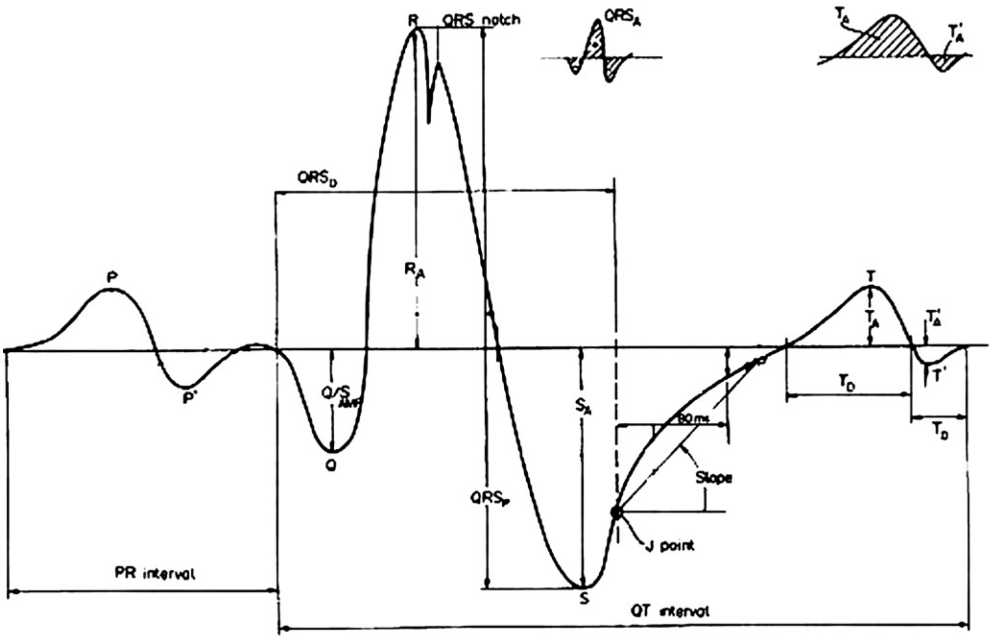

Figure 1. Example of one ECG beat with some of the possible features used for its description.

morphology of 'N' and 'V' beats differs not only from patient to patient but also according to the position of the ECG leads, even on the same person, the global classifiers might even finally become a better choice for telemedicine applications where keeping the exact lead position cannot be guaranteed (Papouchado *et al* 2001).

Additionally, published works differ also in the results presentation. Some papers use the accuracy measure (Hu *et al* 1997, Minhas and Arif 2008) that is common in the field of classical artificial intelligence (AI). Others use the measures of sensitivity and specificity (Chazal *et al* 2004), the common way of reporting results in the medical field. While the accuracy measure gives an overall insight to the correctness of the methods, the sensitivity and the specificity 'evaluate' the methods from the clinical point of view; they are therefore more suitable measures to report classification results of uneven classes where pathology is far less common than the normal state.

In this paper, we attempt to provide a thorough investigation of detection of ventricular beats using different pattern analysis techniques. We use a global training approach based on training using two different databases: the MIT–BIH arrhythmia database from Physionet (Mark *et al* 1997, Goldberger *et al* 2000) and the AHA (AHA-db 1997) database, pushing the generality of our proposed approach even further. The detailed results for each recording are presented using the notions of sensitivity and specificity. The overall methodology is shown in figure 2.

The rest of the paper is structured as follows: in section 2 we present several approaches for the detection of 'V' and 'N' beats. These approaches are tested using the MIT–BIH and AHA databases which are described in section 3. Records from both databases have undergone the same preprocessing steps, and the same features have been computed. In section 4, the measures of sensitivity and specificity are also introduced for the fuzzy case and in section 5 the results are presented and commented. Finally, section 6 concludes the paper presenting a brief discussion and some hints for future work.

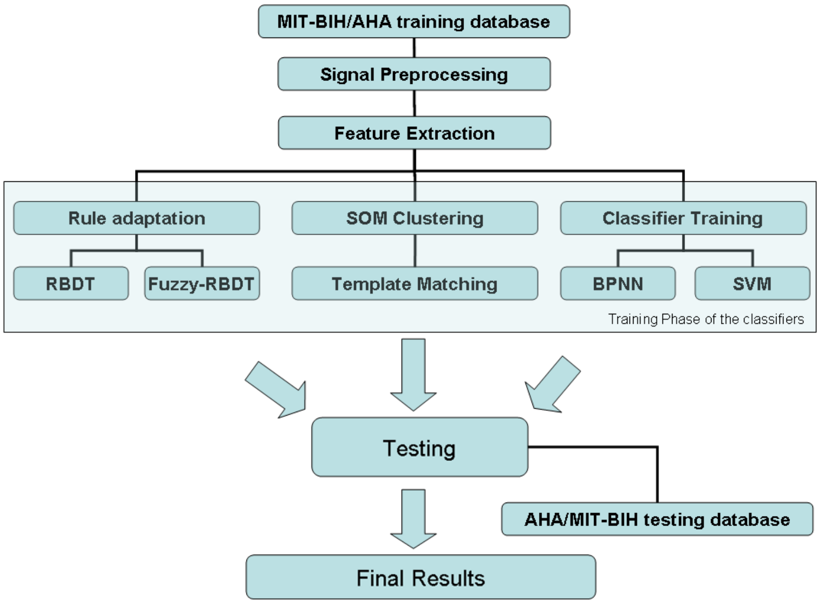

Figure 2. The overall proposed methodology of the holter beat classification presented in the paper. Rule-based decision tree (RBDT), its fuzzy variation (fuzzy RBDT), self-organizing maps and template matching, back-propagation neural network (BPNN) and support vector machine (SVM).

2. Methods

2.1. Data used for experiments

For training and testing the aforementioned methods, 30 min long segments annotated by experts are used from the MIT–BIH and AHA databases. Since we focus on the discrimination between ventricular and normal beats, for the classification purposes all beats labeled as 'V' or 'N' are selected. Additionally, it is necessary to consider some special features regarding the databases as follows.

2.1.1. MIT–BIH database. The beats annotated as right (R) or left (L) bundle branch blocks (BBB) are relabeled as 'N' since the annotations 'R' or 'L' represent morphology of the beat instead of the site of the beat origin, which we are interested in. This approach also complies with the ANSI/AAMI EC57:1998 standard (AAMI 1998) used e.g. in Chazal *et al* (2004).

Our final set contains 89 724 'N' and 6895 'V' beats from the MIT database.

2.1.2. AHA database. It is important to note that the AHA database does not have the supra-ventricular complexes labeled 'S', which are available in the MIT database but they are labeled as normal 'N' instead. The records from the AHA database are divided into eight classes of ten records each, according to the highest level of ventricular ectopy present:

- no ventricular ectopy (records 1001–1010)
- isolated unifocal premature ventricular contractions (PVC) (2001–2010)
- isolated multifocal PVCs (3001–3010)
- ventricular bi- and trigeminy (4001–4010)
- R-on-T PVCs (5001–5010)

Table 1. Features used for ECG beat description.

Amplitude of wave peaks	Ratios of peak amplitudes	Width of intervals	Other features
ampR	ratRT	intQRS	Prematurity of the beat
ampS	ratRS	intQTc	
ampQ	ratQR		
ampTpos			
ampTneg			

- ventricular couplets (6001–6010)
- ventricular tachycardia (7001–7010)
- ventricular flutter/fibrillation (8001–8010).

We select only these records from the AHA database whose final structure would be similar to that of the MIT–BIH database. Therefore, we exclude all records from the last two classes (7001–8010) and also the difficult to measure R-on-T recordings (5000–5010). In total, we choose 110 083 'N' and 8333 'V' beats from the AHA database.

2.2. *Data preprocessing*

Power line interference is filtered using an adaptive filter. Low-frequency drift is filtered during the preprocessing phase using a high-pass filter with cut-off frequency at 0.66 Hz (Daskalov *et al* 1998).

We use a QRS detection algorithm based on the work of Christov (2004) and afterwards the preprocessed signal is analyzed and the typical points on the ECG curve are detected using gradient methods on appropriately filtered signals. We detect the beginning and the end of the QRS complex, along with the maxima of the complex (the highest R peak). The end of T-wave (Toff) position is determined and the P-wave existence measurement is carried out.

2.3. *Extracted features*

We compute nine parameters shown in table 1 that characterize the shape of each ECG beat. The selection of the features is based on our previous work (Chudáček *et al* 2006) and it is motivated also by the requirement to use only such features that will also be usable in the decision tree algorithm, i.e. there is some statistical background for the rule creation. An additional reason which justifies our selection is that most of these features are widely accepted and understood by the medical community.

The amplitude features represent the maxima of the amplitudes. The ratio features represent the ratios of the amplitudes of the main deflections. Two well-known intervals, width of the QRS complex and the corrected QT interval (using the Fridericia equation (Fridericia 1920)), are computed as well. The prematurity feature is considered crucial in distinguishing especially between BBB (part of the 'N' class) and 'V' beats.

2.4. *Training and testing sets*

Although the local training paradigm is used in the majority of papers, we consider that as a slightly controversial practice since the local training needs a rather complicated *a priori* annotation of the part of the signal, as described in the introduction.

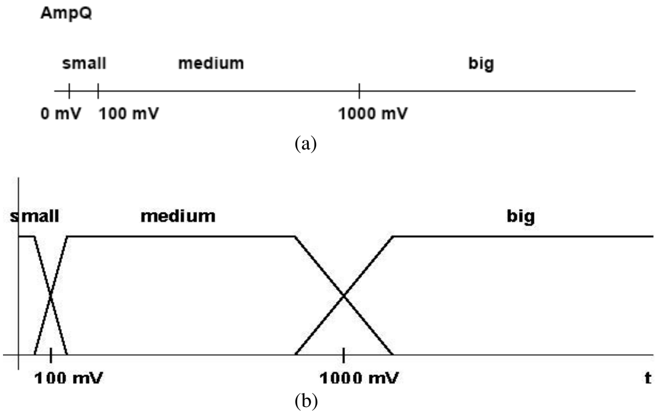

Figure 3. (a) The crisp input for the rule-based decision tree. (b) The membership function for the fuzzy rule-based decision tree.

Therefore, we have decided to use the global training approach and in our case even enforced by using cross-database training. It means that one of the MIT–BIH or AHA databases is fully used for training the classifier and the other is used for testing and vice versa.

Of course, the above-mentioned approach had to be tweaked to meet the needs of the classifiers.

- The decision tree and its fuzzy variant are not trained in the classical sense; the rules are derived based on known relations in the ECG holter data shown in statistical studies and the 'training' phase consists of scaling the rule borders based on histograms as in Castro *et al* (2007).
- The back-propagation network and the SVM approach 'require' even representation of the classes in the training set; therefore, even representations of the V and N beats are randomly selected from the training database. For testing, the whole testing database is used.
- For the template-matching technique following the SOM stage, 25 templates are randomly selected for each 'N' and 'V' class using the rule of having not more than one 'N' and one 'V' class template from each recording in the database.

2.5. Rule-based decision tree

In the case of the rule-based decision tree (RBDT), the domain of each feature (i.e. a value measured on each ECG beat) can be divided into several intervals (see figure 3(a)), which are created based on empirical statistical studies (MacFarlane *et al* 1989) and sometimes based on experience and knowledge of clinicians. The rules that are in each node of the decision tree have a structure such as follows.

```
if(interval_QRS >= 100)& (interval_QRS <= 110)
  pass_status.QRS = 'Widening';
elseif(interval_QRS >= 140)& (interval_QRS <= 160)
  pass_status.QRS = 'Wide > 140'
elseif(interval_QRS >= 110))&  (interval_QRS < 140)
  pass_status.QRS = 'Wide'
```

```
elseif (interval_QRS < 100)& (interval_QRS>50)
    pass_status.QRS = 'Normal'
else
    pass_status.QRS = 'Artefact'
end
goto next_rule
```

The ECG beats are sorted into 'clusters' defined by the Cartesian product of all the passed intervals. Such an approach has, at least for the clinicians, great advantage over the 'black box' methods because the final decision can be easily explained. Moreover, such an approach enables us to use the classification result for further stages of classification (e.g. more detailed) without the need to recompute certain parameters.

For the final decision the rules are arranged into sequences such as follows:

```
if pass_status.Prematurity = YES & ((pass_status.QRS = WIDE) ||
(pass_status.QRS = WIDE>140))
  if pass_status.QRS = WIDE>140
   RESULT = Premature ventricular contraction (V)
else
   if BundleBranchCheck = NEGATIVE
     RESULT = Premature ventricular contraction (V)
   else
     RESULT = BUNDLEBRANCHBLOCK?
   end
  end
end
```

The decision tree algorithm is based on the preset rules as described above. There is nevertheless still room for the training phase, which consists of adapting the intervals to the histogram of the features from the training set, to cover the possible changes of the spread of occurrence of the data in the dataset approach adopted from Castro *et al* (2007).

2.6. Rule-based decision tree with fuzzyfied intervals

The RBDT described in the previous section has a drawback when dealing with border cases; e.g. two very similar ECG beats can be classified differently only because their parameters are very close to the border of an interval resulting in categorizing each of them to a different class. For this reason we propose a generalization of the RBDT and we replace the crisp intervals by fuzzy intervals.

Fuzzy sets (Zadeh 1965) are a generalization of the classical sets. The membership function in the case of a crisp set can take the value of 1 ('belongs to') and 0 ('does not belong to'). The idea of fuzzy sets is to extend this set of truth values, i.e. the output of the membership function, to take values in the interval [0, 1]. This allows us to express the state of a partial belonging to a set which is useful in modeling vague values such as e.g. 'small', 'medium', 'great', etc. Interested readers may find more information e.g. in Klir and Yuan (1995) and Nguyen and Walker (1997).

When we describe the generalization of the RBDT to fuzzy sets (fuzzy RBDT in short), the situation is similar to the RBDT with the only difference that the intervals are fuzzyfied and the fuzzy intervals may be overlapping. Thus, for example, an ECG beat whose value of ampQ is near to a border will partially belong to both intervals (figure 3(b)).

2.7. Self-organizing map ANN

The SOM artificial neural network (ANN) is an unsupervised clustering method (Kohonen 1995). The SOM consists of neurons organized on a regular low-dimensional grid. Each neuron is represented by a d-dimensional weight vector, where d is equal to the dimension of the input vectors. Each neuron is connected to adjacent neurons by a neighborhood relation, which determines the structure of the map.

The SOM is trained iteratively. For each training step, one sample is presented to the SOM input level. The distance between the sample and all the weight vectors of the SOM is calculated using a pre-selected distance measure. After finding the closest neuron, called the best-matching unit (BMU), the weights in the SOM are updated so that the BMU moves closer to the input vector in the input space. This is repeated until the stopping criterion—for example in our case, 1000 learning steps-–is reached.

In our particular implementation of the SOM in the MATLAB's SOMToolbox (Vesanto *et al* 2000), we have selected a SOM consisting of 15×9 neurons in a hexagonal grid arrangement after intensive testing of different sizes of the grid. Additionally, the whole cluster is represented by the median beat from the cluster enabling us to carry out the classification step—template matching—described in section 2.8.

2.8. Template matching

For the SOM clustering method, we had to carry out the classification step separately using the template-matching method. This approach has been used widely and some recent experiments are presented in Chiu *et al* (2005).

The template-matching method in general tries to compare several templates (in our case representatives of the 'N' and 'V' classes) with the investigated part of a signal. Their similarity is then usually described using a distance ('dissimilarity') measure.

The median of each cluster created by the SOM is compared with 50 different 'N' and 'V' templates-–25 for each class. These are obtained by selecting randomly at most one 'N' and one 'V' template from each recording in the database used for training. Finally, correlation coefficients are computed using the classical approach.

For the final decision on the cluster medians, the appropriate coefficients are sorted in a descending manner. Then the majority 'two out of the first three' rule is used to assign the class to the cluster median. All the beats in the cluster represented by the median are classified according to the classification of the corresponding median.

2.9. Support vector machines

Support vector machines (SVMs) are learning systems that are trained using an algorithm based on optimization theory (Burges 1998, Muller *et al* 2001). The SVM solution finds a hyperplane in the feature space that keeps the empirical error small while maximizing the margin between the hyperplane and the instances closest to it. Every new pattern $\mathbf{x}$ is classified to either one of the two categories (in the case of dichotomizing problems $y_i \in \{-1, 1\}$) through

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^{n} y_i a_i K(\mathbf{x}, \mathbf{x}_i) + b\right), \tag{1}$$

where b is a threshold parameter. The coefficients a_i are found by solving a maximization quadratic programming problem which is 'controlled' by a penalty factor C and are assigned to each of the training patterns $\mathbf{x}_i$. The points for which $a_i > 0$ are called support vectors and are the points lying closest to the hyperplane.

The kernel function K implicitly performs the mapping from the input space to the feature space. Among others, the most popular kernels are the polynomial, the radial basis function networks and the two-layer perceptrons. In our experimental procedure, we employ the radial basis function machines:

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x} - \mathbf{x}_i\|^2\right), \qquad i = 1, ..., l. \tag{2}$$

The width σ and the penalty factor C are selected through a cross-validation procedure in an attempt to achieve balanced sensitivity and specificity.

2.10. Back-propagation ANN

Back propagation (BP) (Rumelhart *et al* 1986) is probably the most well-known supervised learning technique used for training ANNs. It can be used to train feed-forward networks with one or more hidden layers.

During the recall phase, a sample is presented to the network and values are propagated from inputs to outputs of the ANN. The difference between desired and actual outputs is calculated formulating the overall network's error. This error is propagated backward from output neurons toward inputs. For each neuron its contribution to the output error is calculated and the weights of its connections are adjusted accordingly. The weights are adjusted using the gradient descent algorithm, which has the disadvantage of getting trapped in local minimum. To overcome this, techniques such as the addition of a momentum term or the delta–bar–delta rule are used.

The back-propagation network used in this study (selected after thorough testing of different configurations) had seven neurons in the first hidden layer and four neurons in the second. For training of the ANN, the standard BP algorithm is implemented in WEKA (Witten and Eibe 2005) with momentum and decreasing learning rate.

2.11. Sensitivity and specificity analysis

In this section, sensitivity and specificity analysis for crisp and fuzzy sets are defined. The computation of these two metrics is based on the 'true positives', 'true negatives', 'false positives', and 'false negatives' elements (i.e. ECG beats) of clusters. The descriptions of these notions are as follows:

- 'true positives': correctly classified as abnormal (the group 'V' in our case)
- 'true negatives': correctly classified as normal (the group 'N' in our case)
- 'false positives': incorrectly classified as abnormal
- 'false negatives': incorrectly classified as normal.

Definition 1: *The cardinality C(S) of a finite set S is the number of elements it contains.*
Definition 2: *Let T_N be a finite set of true negatives and let F_P be a finite set of false positives. The specificity* (Sp) *is a real number in the interval* [0, 1] *defined as follows*:

$$\mathrm{Sp}(T_N, F_P) = \frac{C(T_N)}{C(T_N) + C(F_P)}. \tag{3}$$

Table 2. Results on testing sets for all used methods: RBDT—rule-based decision tree, fuzzy RBDT—rule-based decision tree with fuzzyfied rules, BPNN, SOM, SVM.

Method	MIT sensitivity (%)	MIT specificity (%)	AHA sensitivity (%)	AHA specificity (%)
RBDT	65.75	68.19	58.18	74.63
Fuzzy RBDT	66.11	68.28	80.16	90.43
BP NN	54.14	57.64	51.07	59.56
SOM	63.65	60.26	77.64	62.31
SVM	86.22	85.71	88.17	97.42

Table 3. Results on training sets for all used methods: RBDT—rule-based decision tree, fuzzy RBDT—rule-based decision tree with fuzzyfied rules, BPNN, SOM, SVM.

Method	MIT sensitivity (%)	MIT specificity (%)	AHA sensitivity (%)	AHA specificity (%)
RBDT	84.68	82.40	77.66	82.33
Fuzzy RBDT	88.49	85.88	80.17	81.02
BP NN	91.21	91.12	90.71	96.70
SOM	86.27	75.27	87.80	76.43
SVM	97.30	94.01	98.50	93.47

Definition 3: *Let T_P be a finite set of true positives and let F_N be a finite set of false negatives. The sensitivity* (Se) *is a real number in the interval* [0, 1] *defined as follows*:

$$\mathrm{Se}(T_P, F_N) = \frac{C(T_P)}{C(T_P) + C(F_N)}. \tag{4}$$

Definition 4: *Let T_N be a finite fuzzy set of true negatives and let F_P be a finite fuzzy set of false positives. The fuzzy specificity* Sp_F *is a real number in the interval* [0, 1] *defined as follows*:

$$\mathrm{Sp}_F(T_N, F_P) = \frac{C_S(T_N)}{C_S(T_N) + C_S(F_P)} \tag{5}$$

Definition 5: *Let T_P be a finite fuzzy set of true positives and let F_N be a finite fuzzy set of false negatives. The fuzzy sensitivity* Se_F *is a real number in the interval* [0, 1] *defined as follows*:

$$\mathrm{Se}_F(T_P, F_N) = \frac{C_S(T_P)}{C_S(T_P) + C_S(F_N)}. \tag{6}$$

Remark 6: The sensitivity and the specificity measures, both classical and fuzzy, are defined as real numbers between 0 and 1. Nevertheless, percent values are easier to understand. Therefore, these values are usually multiplied by 100%.

Remark 7: A high value of the sensitivity means that most of the abnormal elements are classified correctly. A high value of the specificity means that most of the normal elements are classified correctly. Thus, high values of both of them mean that the classification method is excellent.

3. Results

The overall results for the globally trained classifiers tested on the whole databases are presented in table 2 and are illustrated in figure 4 and the box plots in figure 5. For comparison of quality of the learning process, results for training sets are presented in table 3.

More detailed results for the three best performing classifiers are given in tables 4 and 5 for the MIT and AHA testing sets respectively. There are only few reported research works

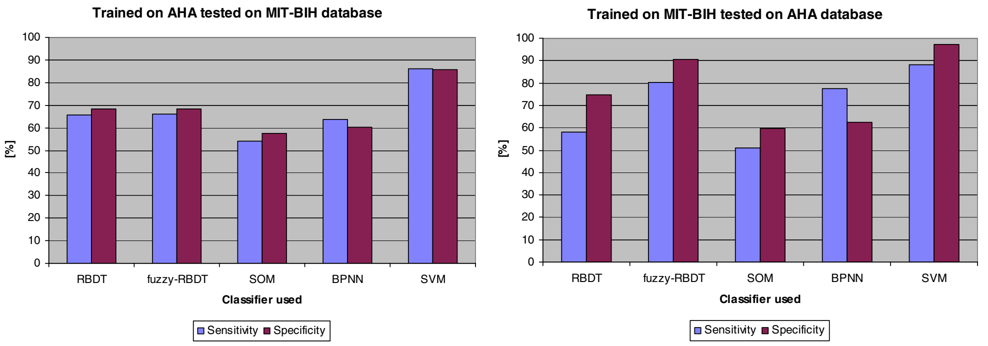

Figure 4. The overall performance of the examined methods using MIT (left) and AHA (right) databases as a testing set.

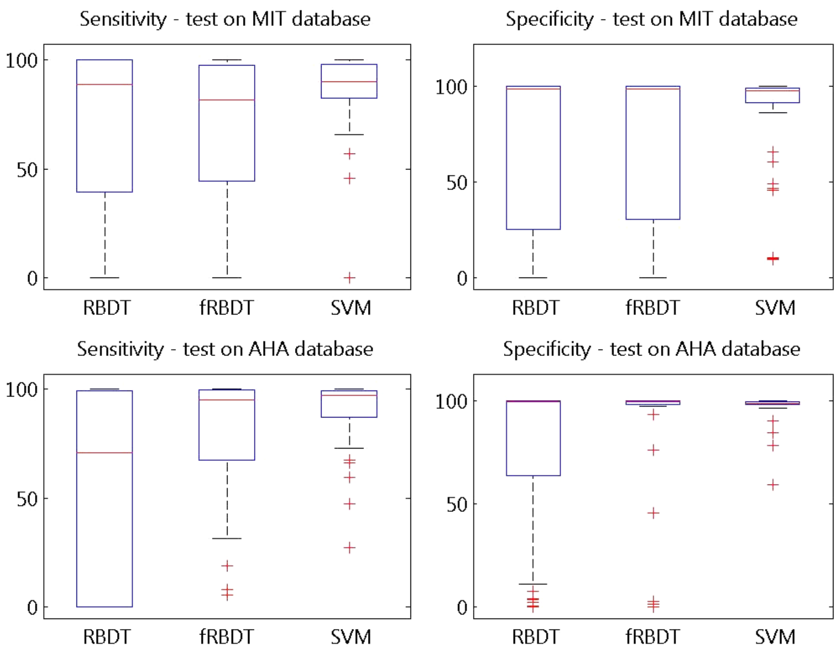

Figure 5. Confidence intervals for the three best performing methods in the form of box plots.

dealing with global classifiers. We mention those which use an approach close to ours using the MIT–BIH database, usually with slight modifications in the way of obtaining the global training/testing for each of them. Hu *et al* (1997) achieved global accuracy of 62.2% for distinguishing 'N' and 'V' beats. The sensitivity and specificity achieved by Bortolan *et al* (2005) are about 80%. Jekova *et al* (2008) report 78.79% sensitivity and 80.61% specificity on the global training set when distinguishing also right and left bundle branch blocks. While pursuing the more difficult task of distinguishing holter beats into five classes using the MIT database, lower numbers were reported by Chazal *et al* (2004) with 86.7% specificity and 67.3% sensitivity for ventricular beats. Dokur and Olmez (2001) had reported so far the

Table 4. Results for the RBDT, fuzzy RBDT and SVM classifiers trained on the MIT and tested on the AHA database.

	Number of beats		RBDT		Fuzzy RBDT		SVM	
Record	N	V	Se (%)	Sp (%)	Se (%)	Sp (%)	Se (%)	Sp (%)
100	2231	1	100	100	100	100	100	97.31
101	1852	0	–	100	–	100	–	98.11
103	2074	0	–	100	–	100	–	95.66
105	2518	41	7.32	98.25	7.32	98.53	90.24	96.82
106	1501	518	91.51	99.2	87.07	98.73	90.15	99
108	1732	16	25	25.12	62.5	19.05	87.5	95.44
109	2484	38	100	0	73.68	0.04	92.11	96.62
111	2115	1	100	0	0	0.05	100	10.59
112	2529	0	–	0.32	–	1.3	–	9.49
113	1781	0	–	100	–	100	–	98.32
114	1812	43	74.42	63.91	88.37	49.39	88.37	97.3
115	1944	0	–	100	–	100	–	97.48
116	2295	109	99.08	100	100	99.3	98.17	99.65
117	1526	0	–	45.81	–	29.36	–	45.87
118	2160	16	100	0	50	0.14	81.25	9.91
119	1535	444	100	0	100	100	99.1	99.87
121	1853	1	0	99.84	0	99.62	0	99.78
122	2468	0	–	100	–	99.96	–	99.43
123	1507	3	100	49.17	100	31.92	100	90.25
124	1523	47	82.98	25.61	80.85	28.17	87.23	92.78
200	1735	826	8.84	1.9	14.16	0.75	91.77	97.75
201	1619	198	100	99.57	100	99.88	98.48	99.75
202	2053	19	42.11	99.76	52.63	99.95	89.47	98.68
203	2521	444	52.03	16.22	46.17	8.96	80.18	97.86
205	2563	71	4.23	100	0	100	0	98.52
207	1539	105	61.9	96.82	36.19	94.02	45.71	96.82
208	1579	991	99.39	99.43	99.29	99.24	96.77	97.15
209	2613	1	100	100	100	100	100	100
210	2416	193	67.36	6.71	64.77	1.7	82.9	49.34
212	2740	0	–	50.99	–	35.66	–	65.91
213	2633	220	97.27	99.77	95.45	99.92	98.18	99.32
214	1995	256	42.58	99.75	38.67	99.8	99.61	98.15
215	3188	164	89.02	59.94	97.56	58.59	98.78	86.39
219	2075	63	95.24	98.36	95.24	98.17	98.41	99.71
220	1946	0	–	100	–	99.9	–	99.13
221	2023	396	96.97	99.41	94.19	99.95	96.72	99.36
222	2054	0	–	100	–	100	–	99.12
223	2021	473	34.25	3.27	62.58	1.14	87.74	46.91
228	1681	361	88.92	99.35	77.01	99.11	78.12	99.82
230	2257	1	100	61.63	100	61.28	100	86.89
231	1560	2	100	20.51	100	36.73	100	60.38
232	392	0	–	100	–	99.74	–	98.98
233	2223	830	33.61	97.66	32.29	97.53	56.99	98.65
234	2692	3	0	100	100	99.96	100	99.89

Table 5. Results for the RBDT, fuzzy RBDT and SVM classifiers trained on the MIT and tested on the AHA database.

Record	Number of beats		RBDT		Fuzzy RBDT		SVM	
	N	V	Se (%)	Sp (%)	Se (%)	Sp (%)	Se (%)	Sp (%)
1001	1615	0	–	100	–	99.94	–	99.75
1002	2588	0	–	100	–	99.96	–	99.92
1003	2173	0	–	99.82	–	99.91	–	99.36
1004	2967	0	–	96.66	–	98.52	–	99.6
1005	2546	0	–	95.8	–	97.6	–	98.86
1006	2115	0	–	100	–	99.91	–	99.48
1007	1528	0	–	100	–	100	–	99.93
1008	2440	0	–	63.73	–	100	–	99.59
1009	3575	0	–	11.64	–	99.41	–	96.39
1010	1985	0	–	0.15	–	97.78	–	98.89
2001	2795	73	0	99.36	98.63	99.82	95.89	90.23
2002	345	54	0	100	100	97.68	100	98.84
2003	2393	13	0	100	100	99.92	100	99
2004	3465	38	0	2.48	63.16	93.3	89.47	97.4
2005	1561	59	100	11.4	94.92	99.62	96.61	98.14
2006	1337	268	100	3.66	100	99.78	100	99.78
2007	2982	276	100	0.13	100	99.87	99.64	99.93
2008	2536	309	100	99.92	100	100	100	100
2009	2263	144	99.31	0.09	99.31	100	99.31	99.82
2010	2450	79	93.67	98.33	100	99.63	100	99.8
3001	2125	26	38.46	100	73.08	99.86	88.46	99.44
3002	2878	58	93.1	99.86	89.66	99.9	96.55	97.39
3003	1907	35	0	99.74	85.71	99.42	94.29	84.58
3004	1791	79	32.91	99.94	44.3	99.83	73.42	98.32
3005	1762	14	57.14	4.2	50	45.69	64.29	99.72
3006	3123	113	0	100	95.58	100	99.12	99.62
3007	2290	26	100	100	92.31	99.83	96.15	99.78
3008	2302	115	100	99.96	94.78	99.74	98.26	98.44
3009	2514	62	0	99.12	96.77	99.68	98.39	98.89
3010	2383	80	87.5	97.82	98.75	98.32	100	97.99
4001	1484	441	99.77	99.73	99.77	99.87	98.87	98.25
4002	2247	120	55	99.6	60.83	99.82	85.83	99.87
4003	2101	473	0	100	8.03	99.86	79.28	99.71
4004	2140	109	86.24	99.67	80.73	99.53	98.17	97.38
4005	1297	145	81.38	99.77	98.62	99.77	98.62	98.77
4006	1767	156	0.64	0.11	99.36	0	96.15	98.36
4007	2821	641	99.22	99.22	98.6	99.47	99.84	98.23
4008	1845	25	68	99.67	72	97.83	88	98.27
4009	1539	825	0	100	99.52	99.87	97.94	98.7
4010	2208	683	0	100	18.89	98.82	47.29	98.23
6001	2447	46	100	0.41	100	0.08	100	99.26
6002	1709	235	93.19	99.77	31.49	76.36	66.38	78.53
6003	2534	157	64.97	99.76	50.32	99.84	59.24	99.72
6004	1999	136	91.91	0	59.56	2.8	89.71	59.38
6005	2083	204	75	99.9	80.39	99.86	97.06	98.27
6006	2364	360	38.06	7.61	99.72	1.4	93.89	99.15
6007	1573	463	39.31	100	98.92	99.87	73	99.87
6008	2300	51	17.65	100	5.88	100	27.45	98.13
6009	1725	750	99.73	99.3	99.73	99.42	99.73	99.25
6010	2869	392	80.87	99.16	80.1	99.72	80.1	99.3

best results using the global classification with 100% specificity and 91.3% sensitivity when dealing with ventricular detection (although the specification of the training/testing sets is quite vague). The only reported work using both MIT–BIH and AHA databases is by Dotsinsky and Stoyanov (2004) reporting an overall sensitivity of 99.04% and a specificity of 99.62%. Although their algorithm is comparable to our rule-based decision tree or fuzzy RBDT, the authors state that the thresholds in the algorithm were adjusted for performance using both databases. The algorithm was adjusted for these two databases and no independent testing set was used for evaluation; therefore, it resembles more a local training approach than the general global approach that we are pursuing.

4. Discussion

The main purpose of this study was to compare and evaluate different approaches for classification of ventricular beats. Since there is no generally used standard, at least when it comes to scientific papers, for the methodology of obtaining and reporting the results for classification of ventricular beats and holter classification in general, we have given large emphasis on the description and critical reasoning for all the steps undertaken prior to the actual classification. The final overall results are presented in table 2 and illustrated in figure 4.

To summarize, we have preprocessed the signals from both AHA and MIT databases in the same way. As described in section 2.2, we have re-sampled all recordings to 500 Hz, we have detected the QRS complexes and located the important points of each ECG beat. We have then selected, partly based on performance and partly based on the necessity to get features that had been statistically processed e.g. in MacFarlane *et al* (1989), nine features described in section 2.3. After that the crucial task of creating the training and testing set is undertaken and described and justified in section 2.4.

In the following paragraphs, we discuss the results of the three best performing methods, i.e. the RBDT, the fuzzy RBDT and the SVM classifier. The RBDT is based on general rules originating from statistical knowledge of intervals for normality of the features used. The main advantage of the RBDT approach is its larger independence from the training set. It does not require any training; nevertheless, in order to set the conditions more similar to other algorithms used we have adapted the rules based on derivation of thresholds from intervals (Castro *et al* 2007) of the training set. The fuzzy variant of the RBDT uses fuzzyfied rules that in general not only are closer to human decision making but also lead to greater generality of the decision making as well.

The SVM algorithm on the other hand is a ‘black box’ technique giving no ‘justification’ of the underlying mechanism involved in coming up with a decision. Nevertheless SVMs have been proven to be once more quite robust in dealing with ‘unknown’ cases, i.e. being able to generalize well-achieving results comparable to the state-of-the-art methods in almost all fields of pattern recognition.

The comparison of our results to works with comparable approaches was done in section 3. The proposed approach with SVMs has achieved results comparable with the best reported in Dokur and Olmez (2001). Only the results of Dotsinsky and Stoyanov (2004) are better than ours. They employ a decision tree-based algorithm, which is in its general concept comparable to our RBDT. Nevertheless, the authors use rules fully adjusted to both AHA and MIT databases—thus, although they do use a general algorithm the results are hard to compare.

From the detailed examination in tables 4 and 5, it is possible to come up with several conclusions.

- All methods perform better when the MIT database is used for training and the AHA database is used for testing. The reason for this behavior lies probably in the larger variety of recordings in the MIT database.
- Comparing the fuzzy RBDT to the classical RBDT, the fuzzy RBDT performs better on both databases. The reason seems to be the fuzzyfication of the rules that we believe lessens the impact of measurement errors on the final decision. The tree works with a given set of rules; it just adjusts them to the training database feature histograms.
- The performance of the SVM method is the best overall in any of the tested databases. It is rather good and balanced on both data sets but the time needed for training is quite high.

The detailed inspection of the results reveals some interesting aspects on the algorithm performance.

- Algorithms fail when dealing with MIT recordings 112 that have significant ST depression causing problems to the detection algorithm and 117. In the case of 117 all the N beats are measured too wide—thus classified as V.
- The performance on the AHA database was generally better, but sensitivity was mediocre on recordings 6002 with two types of normal beats where only one of them is classified correctly while 6004 was again measured too wide. Specificity on the recording 6007 is affected by large variability of ventricular beats in the recording; on the recording 4010 there are ventricular beats too narrow, very much similar to the normal beats.

To sum up, the nine features extracted from both databases seem to be adequate for the discrimination between 'N' and 'V' beats in a global fashion approach. In future work, we will conduct a more thorough investigation dealing with more classes-–a much more demanding task investigating if needed features can be extracted using recently developed techniques in the field of biomedical signal processing.

Acknowledgments

This work has been supported partially by the research program 'Information Society' under grant no 1ET201210527 'Knowledge-based support of diagnostics and prediction in cardiology'. We would also like to thank Martin Hanuliak from BTL Medical Devices, Prague, Czech Republic for many valuable insightful comments.

References

AAMI—Association for the Advancement of Medical Instrumentation 1998 *Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms 1998*

Acar B, Savelieva I, Hemingway H and Malik M 2000 Automatic ectopic beat elimination in short-term heart rate variability measurement *Comput. Methods Programs Biomed.* **63** 123–31

AHA-DB 1997 *AHA database series 1* American Heart Association

Bakardijan H 1992 Ventricular beat classifier using fractal number clustering *Med. Biol. Eng. Comput.* **30** 495–502

Bortolan G, Jekova I and Christov I 2005 Comparison of four methods for premature ventricular contraction and normal beat clustering *Comput. Cardiol.* **32** 921–4

Buhmann M D and Ablowitz M J 2003 *Radial Basis Functions: Theory and Implementations* (Cambridge: Cambridge University Press)

Burges C J C 1998 A tutorial on support vector machines for pattern recognition *Data Mining and Knowledge Discovery* **2** 121–67

Carrault G, Cordier M O, Quiniou R and Wang F 2003 Temporal abstraction and inductive logic programming for arrhythmia recognition from electrocardiograms *Artif. Intell. Med.* **28** 231–63

Castro A, Bveda C and Arcay B 2007 A new method for the initialization of clustering algorithms based on histogram analysis *Visualization, Imaging, and Image Processing 2007* (Palma de Mallorca: Acta Press)

Ceylan R, Özbay Y and Karlik B 2009 A novel approach for classification of ECG arrhythmias: type-2 fuzzy clustering neural network *Expert Syst. Appl.* **36** 6721–6

Chazal P D, O'Dwyer O and Reilly R B 2004 Automatic classification of heartbeats using ECG morphology and heartbeat interval features *IEEE Trans. Biomed. Eng.* **51** 1196–206

Chiu C, Lin T and Liau B 2005 Using correlation coefficient in ECG waveform for arrhythmia detection *Biomed. Eng. Appl. Basis Commun.* **17** 147–52

Christov I I 2004 Real time electrocardiogram QRS detection using combined adaptive threshold *Biomed. Eng. Online* **3** 28

Christov I and Bortolan G 2004 Ranking of pattern recognition parameters for premature ventricular contractions classification by neural networks *Physiol. Meas.* **25** 1281–90

Christov I, Gomez-Herrero G, Krasteva V, Jekova I, Gotchev A and Egiazarian K 2006 Comparative study of morphological and time-frequency ECG descriptors for heartbeat classification *Med. Eng. Phys.* **28** 876–87

Christov I, Jekova I and Bortolan G 2005 Premature ventricular contraction classification by the *K*th nearest-neighbours rule *Physiol. Meas.* **26** 123–30

Chudáček V, Lhotská L, Stylios C, Georgoulas G and Staviař M 2006 Comparison of methods for premature ventricular beat detection *Int. Special Topics ITAB Conf.* (Piscataway, NJ: IEEE)

Cuesta-Frau D, Biagetti M, Mico-Tormos R Q and Aboy M 2007 Unsupervised classification of ventricular extrasystoles using bounded clustering algorithms and morphology matching *Med. Biol. Eng. Comput.* **45** 229–39

Cuesta-Frau D, Perez-Cortes J C and Andreu-Garcıa G 2001 Clustering of electrocardiograph signals in computer-aided Holter analysis *Comput. Methods Programs Biomed.* **72** 179–96

Daskalov I K, Dotsinsky I A and Christov I I 1998 Developments in ECG acquisition, preprocessing, parameter measurement, and recording *IEEE Eng. Med. Biol. Mag.* **17** 50–8

Dokur Z and Olmez T 2001 ECG beat classification by a novel hybrid neural network *Comput. Methods Programs Biomed.* **66** 167–81

Dotsinsky I A and Stoyanov T V 2004 Ventricular beat detection in single channel electrocardiograms *Biomed. Eng. Online* **3** 3

Fridericia L S 1920 Die Systolendauer im Elektrokardiogramm bei normalen Menschen und bei erzkranken *Acta Medica Scand. Stockh.* **57** 469–86

Goldberger A L, Amaral L, Glass L, Hausdorf J M, Ivanov P C, Mark R G, Mietus J E, Moody G B, Peng C K and Stanley H E 2000 PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals *Circulation* **101** e215–20

Haykin S 1999 *Neural Networks: A Comprehensive Foundation* (Englewood Cliffs, NJ: Prentice-Hall)

Holter N J 1961 New method for heart studies *Science* **134** 1214–20

Hosseini G, Luo D and Reynolds K J 2006 The comparison of different feed forward neural network architectures for ECG signal diagnosis *Med. Eng. Phys.* **28** 372–8

Hu Y H, Palreddy S and Tompkins W J 1997 A patient-adaptable ECG beat classifier using a mixture of experts approach *IEEE Trans. Biomed. Eng.* **44** 891–900

Jager F 2002 Feature extraction and shape representation of ambulatory electrocardiogram using the Karhunen–Loeve transform *Elektroteh. Vestn.* **69** 83–89

Jekova I, Bortolan G and Christov I 2008 Assessment and comparison of different methods for heartbeat classification *Med. Eng. Phys.* **30** 248–57

Klir G J and Yuan B 1995 *Fuzzy Sets and Fuzzy Logic: Theory and Applications* (Englewood Cliffs, NJ: Prentice-Hall)

Kohonen T 1995 *Self-Organizing Maps* (Berlin: Springer)

Kyriacou E *et al* 2003 Multi-purpose health care telemedicine systems with mobile communication link support *BioMedical Engineering OnLine* **2** 7

Lagerholm M, Peterson C, Braccini G, Edenbrandt L and Sornmo L 2000 Clustering ECG complexes using Hermite functions and self-organizing maps *IEEE Trans. Biomed. Eng.* **47** 838–48

Macfarlane P W, van Oosterom A, Pahlm O, Kligfield P, Janse M and Camm J 1989 *Comprehensive Electrocardiology* (Oxford: Pergamon)

Madala H and Ivakhnenko A 1994 *Inductive Learning Algorithm for Complex System Modelling* (Boca Raton, FL: CRC Press)

Mark R and Moody G 1997 MIT-BIH Arrhythmia Database (Online) http://ecg.mit.edu/dbinfo.html

Minhas F A and Arif M 2008 Robust electrocardiogram (ECG) beat classification using discrete wavelet transform *Physiol. Meas.* **29** 555–70

Moody G and Mark R 1989 QRS morphology representation and noise estimation using the Karhunen–Loeve transform *Comput. Cardiol.* **16** 269–72

Moreas J C, Seixas M O, Vilani F N and Costa E V 2002 A real time QRS complex classification method using Mahalanobis distance *Comput. Cardiol.* **29** 201–4

Muller K R, Mika S, Ratsch G, Tsuda K and Scholkopf B 2001 An introduction to Kernel-based learning algorithms *IEEE Trans. Neural Netw.* **2** 181–201

Nguyen H T and Walker E A 1997 *A First Course in Fuzzy Logic* (Boca Raton, FL: CRC Press)

Olmez T, Dokur Z and Yazgan E 1997 Classification of ECG waveform using genetic algorithms *Proc. 19th Int. Conf. IEEE/EMBS*

Papouchado M, Cox H, Bailey J, White W and Spreadbury T 2001 Early experience with transmission of data from moving ambulances to improve the care of patients with myocardial infarction *J. Telemed. Telecare* **7** 27–8

Rumelhart D E, Hinton G E and Williams R J 1986 Learning representations by back-propagating errors *Nature* **323** 533–6

Tsipouras M G, Fotiadis D I and Sideris D 2002 Arrhythmia classification using the RR-interval duration signal *Comput. Cardiol.* **29** 485–8

Tsipouras M G, Voglis C and Fotiadis D I 2007 A framework for fuzzy expert system creation—application to cardiovascular diseases *IEEE Trans. Biomed. Eng.* **54** 11

Tsipouras M G, Voglis C, Lagaris I E and Fotiadis D I 2005 Cardiag arrhytmia classification using support vector machines *3rd European Medical and Biological Engineering Conf.*

Vesanto J, Himberg J, Alhoniemi E and Parhankangas J 2000 Self-organizing map in Matlab: the SOM Toolbox, TOOLMET

Witten I and Eibe F 2005 *Data Mining: Practical Machine Learning Tools and Techniques* 2nd edn (San Francisco, CA: Morgan Kaufmann)

Yu S N and Chou K 2009 Selection of significant independent components for ECG beat classification *Expert Syst. Appl.* **36** 2088–96

Zadeh L A 1965 Fuzzy sets *Inf. Control* **8** 338–53