

HySMID: An Ischemia Diagnosis System Using Genetic Algorithms and Multicriteria Decision Analysis

C Papaloukas¹, Y Goletsis¹, DI Fotiadis^{1,3,4}, A Likas^{1,3}, LK Michalis^{2,4}

¹Unit of Medical Technology and Intelligent Information Systems,
Department of Computer Science, University of Ioannina, Ioannina, Greece

²Department of Cardiology, Medical School, University of Ioannina, Ioannina, Greece

³Biomedical Research Institute –FORTH, Ioannina, Greece

⁴Michaelidion Cardiac Center, Ioannina, Greece

Abstract

Multicriteria decision analysis is employed in a four-stage (preprocessing, beat classification, window characterization and episode definition) ischemia diagnosis system, named HySMID. Beat classification is realized using a multicriteria approach where each beat is compared to already classified category prototypes and the similarity is computed in a fuzzy way. The comparison is based upon five criteria that include ST segment changes, T wave alterations and patient's age. The difficulty in applying these criteria is the determination of the required parameter values (thresholds and weights). To overcome it, a genetic algorithm is used, which after proper training, automatically calculates the optimum values of the above parameters. HySMID was validated using the European Society of Cardiology ST-T database and performed better than previously reported methods.

1. Introduction

Myocardial ischemia is one of the leading causes of death in modern societies and its early diagnosis and treatment is of great importance [1]. In the electrocardiographic (ECG) signal ischemia is expressed as slow dynamic changes of the ST segment and/or the T wave [2]. A simple and non-invasive method to observe such alterations is long duration electrocardiography, like Holter recordings or continuous ECG monitoring. The accurate ischemic episode detection in the recorded ECG is based on the correct classification of the ischemic cardiac beats. Several techniques have been proposed for this purpose, which evaluate the ST segment changes and the T wave alterations with different methodologies. Their evaluation process is based upon medical knowledge expressed by a set of rules [3], artificial neural

networks (ANNs) [1,4–7], fuzzy logic [8], wavelet analysis [9] or other signal processing techniques [10–12].

In the present work, Multicriteria Decision Analysis (MCDA) is utilized for ischemic beat classification, based on the comparison between prototypes and beats. This MCDA approach avoids using distance measures and is able to utilize data expressed in different units for each criterion. The problem with the MCDA implementation is the difficulty in modelling the doctors' preference process and value system. Doctors' knowledge often needs to be transformed to threshold and weight values, a task that requires a considerable effort from the medical expert. Moreover, it is rather unlikely to achieve similar values from different experts, since each one exhibits differences in his value system. Aiming at overcoming these difficulties and at supporting automated diagnosis we employ a genetic algorithm (GA) in order to properly adjust all necessary parameters of the MCDA model.

2. Materials and methods

The proposed hybrid system for myocardial ischemia diagnosis (HySMID) utilizes a four stage algorithmic schema. In the first stage, the ECG signal is preprocessed to achieve noise removal and extraction of the signal features used in beat classification. In the second stage, each beat is classified as normal, abnormal (ischemic) or artifact. This information is used in the third stage (the window characterisation stage) where each 30-second ECG window is classified as ischemic or not. In the fourth stage the start and end points of each ischemic episode are identified using the concatenation of consecutive ischemic windows. At this stage, a merging procedure is also applied to identify the overall ischemic episodes from the episodes detected in each available

lead. We note that the whole procedure described above is applied in each lead separately.

2.1. ECG preprocessing

In order to classify each cardiac beat as ischemic or normal, all the relevant ECG features (isoelectric line, ST segment and T wave) need to be defined and measured. The feature extraction procedure was implemented using already reported algorithms [13,14]. Due to the presence of noise in the ECG recordings (such as power line interference, electromyographic contamination and baseline wandering) a filtering module was also applied [3]. Briefly, a least square procedure was employed to remove baseline wandering effects, while the A/C interference and the electromyographic noise were handled properly by an averaging technique. It should be noted that when the beginning of the ST segment (J point) cannot be defined, due to substantial amount of noise, the corresponding cardiac beat is rejected at this stage.

2.2. Beat classification

Beat classification is performed comparing each beat with already classified prototype beats under several criteria.

More analytically, let:

A be a finite set of beats,

F be a set of n criteria (features), with $n \geq 1$,

w_j be the weight of each criterion, $\sum_j w_j = 1$,

$C = \{C^1, \dots, C^K\}$ the set of the categories ($K > 1$),

$B^h = \{b_p^h \mid 1, \dots, L^h \text{ and } h=1, \dots, K\}$ the set of reference patterns (prototypes) of the category C^h , where b_p^h represents the p prototype of the category C^h and L^h the number of the prototypes of this category.

Each beat in A and B is characterized by a feature vector \vec{g} containing its feature values for all n criteria of F (i.e. $\forall a \in A, g(a) = (g_1(a), \dots, g_n(a))$ and $\forall b_p^h \in B, g(b_p^h) = (g_1(b_p^h), \dots, g_n(b_p^h))$). Every beat a is compared to each prototype beat b_p^h under each criterion j . A Similarity Index ($SI_j(a, b_p^h)$) is computed for every criterion.

Two thresholds are applied in order to model the similarity system:

-a similarity threshold q_j , representing the maximal value of difference between two beats, so that the two beats to be judged similar under the j^{th} criterion,

-a dissimilarity threshold p_j , representing the minimal value of difference between two beats, so that the two

beats can be considered totally dissimilar under the j^{th} criterion.

SI_j is then computed as:

$$SI_j(a, b_p^h) = \begin{cases} 1 & , \left| g_j(a) - g_j(b_p^h) \right| \leq q_j \\ \frac{\left| g_j(a) - g_j(b_p^h) \right| - p_j}{q_j - p_j} & , q_j < \left| g_j(a) - g_j(b_p^h) \right| < p_j \\ 0 & , \left| g_j(a) - g_j(b_p^h) \right| \geq p_j \end{cases}$$

We can compute a Concordance Index (CI) aggregating the Similarity Indices SI_j . CI indicates the concordance to the hypothesis that beat a is similar to beat b_p^h :

$$CI(a, b_p^h) = \sum_j w_j SI_j(a, b_p^h) \text{ where, } w_j \text{ represents the weight of each criterion and } \sum_j w_j = 1.$$

Having calculated the CI for all the typical objects of category h , we can compute the degree of membership of the beat a to category h :

$$d(a, C^h) = \max\{ CI(a, b_1^h), CI(a, b_2^h), \dots, CI(a, b_{L^h}^h) \}.$$

Finally, repeating this process for all h categories, we can decide that beat a belongs to this category with which it has the maximum membership degree:

$$C(a) = \arg \max_h d(a, C^h).$$

In our multicriteria approach we used five criteria:

1. ST segment deviation,
2. ST segment slope,
3. T wave amplitude,
4. T wave normal amplitude and polarity, and
5. patient's age.

The ST deviation is measured at the J80 point and refers to the amplitude deviation of the ST segment from the isoelectric line, which is the line that defines the zero amplitude level. In cases where the heart rate is faster than 120 beats /min the J60 point is used instead. J80 and J60 are the points that lie 80 ms or 60 ms, respectively, after the J point. The ST segment slope is defined as the slope of the line connecting the J and J80 (or J60) points. The T wave amplitude is defined as the amplitude deviation of the T wave peak from the isoelectric line. Finally, the T wave normal amplitude and polarity expresses the amplitude and polarity of normal beats for a specific ECG recording. It is estimated using the first 30 s of each recording and is computed by the mean value of the T wave amplitudes at this ECG interval.

A set of parameters must be specified for the application of our method. These are the similarity thresholds q_j , the dissimilarity thresholds p_j and the criterion weights w_j for all n criteria ($j=1 \dots n$). We employed a GA to automatically obtain the necessary parameters.

GAs are general purpose search algorithms that use principles inspired by natural population genetics to evolve solutions to problems. The basic idea of a GA is to maintain a population of knowledge structures (called chromosomes) each one representing a candidate solution to the problem. In an iterative process three operators namely, selection, crossover and mutation are applied in order to produce chromosomes (potential solutions) that maximise a user defined fitness function.

In our application, all 15 parameters were encoded in the *chromosome* $\{q_1, \dots, q_5, p_1, \dots, p_5, w_1, \dots, w_5\}$.

Following a series of experiments, the GA we applied demonstrates binary representation, string length 81, two-point crossover, elitism and linear fitness scaling. As far as the parameters of the GA are concerned, the population size was set to 30, the maximum number of generations to 400, the crossover probability to 0.95 and the mutation probability to 0.02.

Aiming at maximising both the sensitivity (Se) and the specificity (Sp) provided by the multicriteria algorithm, the following fitness function was employed:

$$f = \sqrt{\left(\frac{Se + Sp}{2}\right)^2 - (Se - Sp)^2}.$$

2.3. Window characterization

Once every beat in each lead has been classified as ischemic or normal, the next stage is to decide whether a sequence of beats belongs to an ischemic window. An ischemic episode is defined as a time period of no less than 30 s containing ischemic beats. For this reason a sliding adaptive window was implemented to examine whether a sequence of ischemic beats exists that lasts more than 30 s. The sliding technique proceeds moving the window one beat at a time, while always keeping its duration equal to 30 s. This means that we will not have the same number of beats in all the windows but this number is adapted to the heart rate. A percentage threshold is also used for the ischemic beats appearing in a window. Specifically, if the ratio of ischemic beats in a window is above 75%, then the window is considered to belong to an ischemic episode.

2.4. Episode definition

If a series of consecutive ischemic windows is identified then the left boundary of the ischemic episode corresponds to the beginning of the first window in the series and the right boundary to the end of the last window. In order to increase the flexibility of the algorithm, the existence of time intervals of less than 30 s with beats that do not constitute an ischemic window is permitted in the above counting. Once, all the episodes

for each lead are detected, a merging technique is realised to define the overall episodes in the ECG recording.

2.5. HySMID evaluation

A task-specific ECG database was developed in order to adjust all the parameters of HySMID, based on the recordings of the European Society of Cardiology ST-T (ESC ST-T) database [15]. Eleven hours were extracted from 10 representative files of the ESC ST-T database (recordings: e0103, e0104, e0105, e0108, e0113, e0114, e0147, e0159, e0162 and e0206). The excerpts were diagnosed in beat-by-beat mode yielding in a dataset of 86,384 cardiac beats characterised as normal, ischemic or artifact. Those not detected by the QRS detector as well as the artifacts (6,754 beats) were rejected, leaving a total of 76,989 beats. From these, the 37,663 (48.92%) were characterised as ischemic and the rest as normal. The training set for the multicriteria method (i.e. the set of prototypes) was constructed after selecting the first beat out of a sequence of 40 (2.5% of the final dataset). This training set contained 954 ischemic beats and 982 normal, leaving 75,053 beats for training the genetic algorithm and testing the classification system. Approximately the 1/3 of these beats (i.e. 25,000) was selected randomly for training the genetic algorithm while the remaining 50,053 were used as a test set. Finally, all the 90 records of the ESC ST-T database were used to test the ischemic episode detection performance of HySMID.

3. Results

The GA provided the MCDA parameter values achieving a fitness value $f=8295.5$, which corresponds to $Se=91.4\%$ and $Sp=90.8\%$ for the GA training set. Using the obtained by the GA parameter values, the multicriteria method classified the beats in the test set with a Se and a Sp of 91%, both. As for the episode detection performance, the overall system obtained a Se of 91% and a positive predictive accuracy (PPA) of 93%, when tested with all the 90 records of the ESC ST-T database.

In Table 1 the results of the proposed hybrid system are compared to those of other approaches of automated ischemic episode detection, such as ANNs [1,4–6], ANNs combined with principal component analysis (PCA) [4,7], set of rules [3], fuzzy logic [8], PCA [11] or other signal processing techniques [12]. It should be mentioned that other beat classification systems have also been proposed but their performance cannot be judged against the above mentioned methods, since either they have been evaluated with other test sets or they employed different performance measures [9,10].

Table 1. Performance measures of automated ischemia diagnosis systems.

System Description	Se ¹ (%)	PPA ² (%)
ANN [1]	85	88
Rule-based [3]	94	79
ANN & PCA [4]	77	86
Recurrent ANN [4]	77	85
Knowledge-learning ANN [4]	71	66
ANN [5]	84	79
CP-SOM & SVM [6]	83	82
ANN & PCA [7]	90	89
Fuzzy Logic [8]	81	68
PCA [11]	85	86
Geometric method [12]	82	81
HySMID	91	93

¹ Se: Sensitivity ² PPA: Positive Predictive Accuracy

4. Discussion

In the current work, a hybrid system was developed for the automated diagnosis of myocardial ischemia using long duration ECGs. The fuzzy modelization of MCDA eliminates the problem of strict thresholding since it employs a comparison between alternative beats through the use of values in different criteria. The application of thresholds and the pseudocriterion modelization provides a more rational similarity system, since it is not realistic to set a strict value under which there is similarity between two beats and over which strict dissimilarity.

The obtained results are better than those of other similar approaches [1,3–8,11–12] when tested using the data from the ESC ST-T database. Furthermore, the multicriteria approach has the ability to provide interpretations for the decisions made following post-processing (identification of the closest prototype(s)). This is of great importance when designing a device for medical decision support and will assist doctors to reach faster and safer a diagnosis.

HySMID's performance can be further improved by a more accurate definition of the employed ECG features. The presence of noise in the recordings of the ESC ST-T database causes poor detection of the ST segment, the isoelectric line and the T wave. However, modern ECG recorders provide less noisy signals and the proposed system is expected to perform better. Furthermore, in order for HySMID to be used in clinical practice, the developed cardiac beat database should be extended and additional types of ischemic and normal ECG waveform patterns must be included. This can be implemented easily in the MCDA approach since only the beats used as prototypes need to be modified.

The application of the MCDA algorithm was successful and modeled suitably the dynamic characteristics of myocardial ischemia. On the other

hand, the employment of GA assisted in automatically defining the values of the parameters (weights and thresholds) used by the multicriteria method. Still, further testing is needed using data extracted from real clinical conditions to fully reveal the potential of the proposed system.

References

- [1] Silipo R, Taddei A, Marchesi C. Continuous monitoring and detection of ST-T changes in ischemic patients. *Computers in Cardiology* 1994;225–8.
- [2] Goldman MJ. Principles of clinical electrocardiography. Los Altos, CA: LANGE Medical Publications, 1982.
- [3] Papaloukas C, Fotiadis DI, Likas A, Liavas AP, Michalis LK. A knowledge-based technique for automated detection of ischaemic episodes in long duration electrocardiograms. *Med Biol Eng Comput* 2001;39:105–12.
- [4] Silipo R, Marchesi C. Artificial neural networks for automatic ECG analysis. *IEEE Trans Signal Processing* 1998;46:1417–25.
- [5] Stamkopoulos T, Strintzis M, Pappas C, Maglaveras N. One-lead ischaemia detection using a new backpropagation algorithm and the European ST-T database. *Computers in Cardiology* 1992:663–6.
- [6] Papadimitriou S, Mavroudi S, Vladutu L, Bezerianos A. Ischemia detection with a self-organising map supplemented by supervised learning. *IEEE Trans Neur Networks* 2001;12:503–15.
- [7] Papaloukas C, Fotiadis DI, Likas A, Michalis LK. An ischemia detection method based on artificial neural networks. *Artif Intell Med* 2002;24:167–78.
- [8] Vila J, Presedo J, Delgado M, Barro S, Ruiz R, Palacios F. SUTIL: Intelligent ischaemia monitoring system. *Int J Med Informatics* 1997;47:193–214.
- [9] Lemire D, Pharand C, Rajaonah J-C, Dube B, LeBlanc AR. Wavelet time entropy, T wave morphology and myocardial ischaemia. *IEEE Trans Biomed Eng* 2000;47:967–70.
- [10] Badilini F, Merri M, Benhorin J, Moss AJ. Beat-to-beat quantification and analysis of ST displacement from Holter ECGs: a new approach to ischaemia detection. *Computers in Cardiology* 1992:179–82.
- [11] Jager F, Moody G, Mark R. Detection of transient ST segment episodes during ambulatory ECG monitoring. *Comput Biomed Res* 1998;31:305–22.
- [12] Taddei A, Costantino G, Silipo R, Emdin M, Marchesi C. A system for the detection of ischaemic episodes in ambulatory ECG. *Computers in Cardiology* 1995:705–8.
- [13] Tompkins WJ. Biomedical digital signal processing (C-language examples and laboratory experiments for the IBM® PC). Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [14] Daskalov K, Dotsinsky IA, Christov II. Developments in ECG acquisition, preprocessing, parameter measurement, and recording. *IEEE Eng Med Biol* 1998;17:50–8.
- [15] European Society of Cardiology. European ST-T database directory. Pisa: S.T.A.R., 1991.

Address for correspondence:

Dimitrios I. Fotiadis, Unit of Medical Technology and Intelligent Information Systems, Dept. of Computer Science, University of Ioannina, GR 451 10 Ioannina, GREECE
e-mail: fotiadis@cs.uoi.gr