A NOVEL METHODOLOGY FOR MYOCARDIAL ISCHAEMIA DIAGNOSIS USING ASSOCIATION RULES

T.P. Exarchos*, C. Papaloukas** and D.I. Fotiadis*,***

*Unit of Medical Technology and Intelligent Information Systems,
Department of Computer Science, University of Ioannina, Ioannina, Greece
**Department of Biological Applications and Technology, University of Ioannina, Ioannina, Greece
***Biomedical Research Institute, Ioannina, Greece

me01238@cc.uoi.gr, papalouk@cc.uoi.gr, fotiadis@cs.uoi.gr

Abstract: In this work, an automated methodology based on association rules is presented for the identification of ischaemic beats in long duration electrocardiographic (ECG) recordings. The proposed methodology consists of three stages: (a) The ECG signal is pre-processed and noise removal takes place along with the extraction of all the necessary ECG features, (b) The extracted continuous valued features are transformed to categorical ones using discretization techniques and (c) Association rules are extracted from the discretized dataset which are used for classification. Our methodology was evaluated using a cardiac beat dataset, constructed using several recordings of the European Society of Cardiology (ESC) ST-T database. The obtained sensitivity (Se) and specificity (Sp) was 88% and 88%, respectively. The major advantage of the proposed methodology is the ability to provide interpretation for the decisions made since it is based on a set of association rules.

Keywords: Automated ischaemic beat detection, data mining, classification using association rules

Introduction

Myocardial ischaemia is caused by the lack of oxygen and nutrients to the contractile cells. Frequently, it may lead to myocardial infarction and consequently to heart failure. In the case of myocardial ischaemia, alterations are observed in the electrocardiographic (ECG) signal like deviations in the ST segment and changes in the T wave. The detection and assessment of those alterations in long duration ECGs is a simple and non-invasive method for the diagnosis of ischaemia [1].

Myocardial ischaemia diagnosis using the ECG signal can be described as a sequence of two tasks: ischaemic beat detection and ischaemic episode definition. The first is the classification of beats as normal or ischaemic. Several techniques have been proposed for ischaemic beat classification that take into consideration the ST segment changes and the T-wave alterations using different approaches. More specifically, researchers in the field use parametric modelling [2,3], wavelet theory [4], set of rules [5,6], artificial neural networks [7-10], multicriteria decision analysis [11] and genetic algorithms [12].

Cardiac beat detection and classification is a key process for the definition of the ischaemic episodes in the ECG signal. The accuracy of the beat classification influences ischaemic episode definition where sequences of ischaemic beats need to be identified. Various methods have been proposed for ischaemic episode detection based on a set of rules [5,13], artificial neural networks [8,10,14], fuzzy logic [15] and other signal analysis techniques [16,17].

The most common techniques for the beat classification problem are the neural and the rule based ones. Neural based approaches have resulted in high performance but exhibit an important drawback due to their inability to provide explanations for the classification decisions. Rule based approaches exhibit the highly desirable feature of interpreting the decisions but their performance is not equally satisfactory in terms of accuracy.

On the other hand, data mining and more precisely classification using association rules [18] can deal with both aspects: high accuracy and interpretation for the decisions made. This methodology has been used in medicine for medical image categorization [19], analysis of hospitalized patient flows [20] and electroencephalographic transient event detection and classification [21].

In this work, a classification methodology employing association rules is proposed for cardiac beat classification as normal and ischaemic. Our methodology is based on a three stage schema (Fig. 1) and involves a pre-processing module (noise removal and feature extraction), a module for discretizing the continuous feature values and a module for association rule mining and classification. In order to train and test the efficiency of the proposed methodology a specific dataset was constructed using recordings of the ESC ST-T database. Five representative features were extracted from every cardiac beat. The employment of association rules for ECG analysis offers the potential of discovering new knowledge in the form of rules and is able to provide explanation for the decisions made.

In the following we describe briefly data mining as well as the process of discretization and classification

ISSN: 1727-1983

using association rules. Next, the three stage schema of the beat classification methodology as well as the employed dataset are analysed. The experiments carried out for the evaluation of our classification system are presented. The advantages and disadvantages of the methodology are given in the discussion section. Possible further improvements are discussed too.

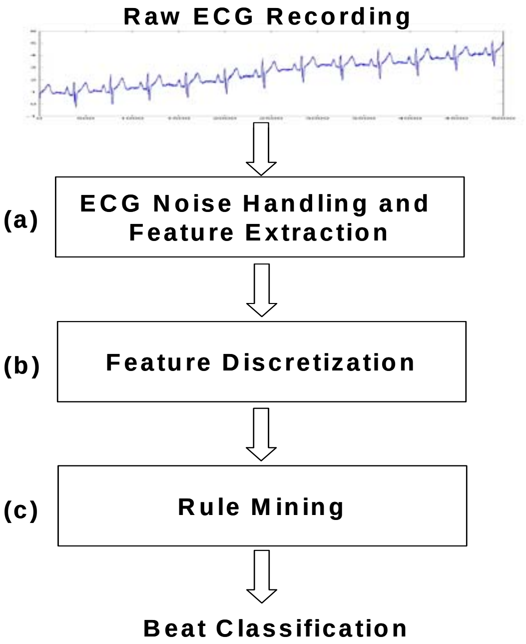

Figure 1: The proposed three-stage methodology.

Materials and Methods

A methodology based on a three-stage schema was developed for ischaemic beat detection (Fig. 1). The three stages correspond to a) noise handling and ECG feature extraction, b) feature discretization and c) rule mining and beat classification. In the first stage, pre-processing of the ECG recording was performed in order to remove noise and extract all the necessary signal features. In the second stage every continuous valued feature was discretized in order to be utilized in the next stage. In the third stage, an association rule mining algorithm was applied to generate association rules, which were used for cardiac beat classification.

A . Pre-processing and feature extraction

In the first stage of the methodology, pre-processing of the recorded ECG signal was performed in order to remove noise (e.g. baseline wandering, A/C interference and electromyographic contamination). Noise removal was achieved by filtering each recorded cardiac beat separately using ECG filtering [12]. Baseline wandering was removed by subtracting from the recorded signal the first-order polynomial that best fits the cardiac beat. A/C interference and electromyographic contamination were not removed from the recorded signal but were handled properly for the detection of the J point. More specifically, for these two types of noise, averaging filters of 20 ms were applied around J. The location of the J point was detected using an edge-detection algorithm [22].

After noise removal and J point detection, the following features were extracted from each cardiac beat (Fig. 2):

- ST segment deviation,
- ST segment slope,
- ST segment area,
- T-wave amplitude and
- T-wave normal amplitude.

In addition to these 5 features a sixth one (age of the patient) was employed. This feature was given in the demographic data provided by the ESC ST-T database.

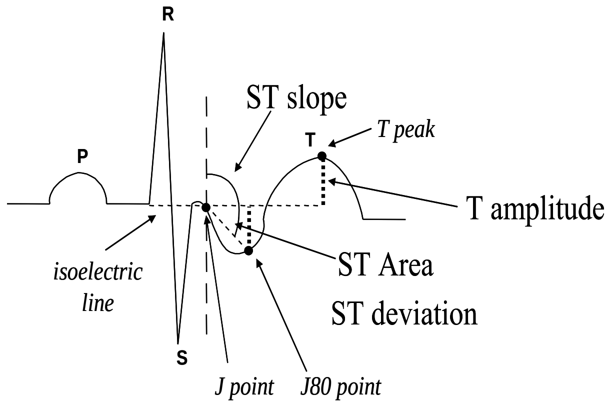

Figure 2: The ECG features extracted from the recordings: ST segment deviation, ST segment slope, ST segment area, and T-wave amplitude

In the ST segment the measurements are taken 80 milliseconds after the J point (J80) if the heart rate does not exceed 120 beats/min, or 60 milliseconds after the J point (J60), otherwise. The ST segment deviation 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 slope is the slope of the line connecting the J and J80 (or J60) points. The ST segment area is the area between the ECG trace, the isoelectric line and the points J and J80 (or J60). The T-wave amplitude is the amplitude deviation of the T-wave peak from the isoelectric line. The T-wave normal amplitude together with its respected polarity refers to the amplitude and polarity of normal beats for a specific ECG lead. It is calculated using the first 30 s of each recording and is computed by the mean value of the T-wave amplitudes at this interval. T-wave amplitude and T-wave normal amplitude are merged into a single new feature, T-wave – T-normal amplitude, which is the difference between the two features.

ISSN: 1727-1983 © 2005 IFMBE

B. Discretization

Datasets usually have continuous features which make them unsuitable for certain data mining algorithms dealing mainly with nominal features, such as naive Bayes classifiers and association rule mining algorithms. To use such algorithms, continuous features must be replaced with nominal features representing intervals of continuous domains with discrete values. This process is known as discretization and involves the transformation of a quantitative variable (continuous feature) into a qualitative one (nominal feature) [23].

Equal depth binning (or frequency partitioning) [23] is perhaps one of the simplest methods for data discretization and it has been applied to produce nominal values from continuous ones. It involves division of a continuous variable A into b bins (sets). The number of bins b is a parameter provided by the user. Given d instances, each bin contains d/b (possibly duplicated) adjacent values. The method is applied to each continuous feature independently. Furthermore, since it makes no use of instance class information is an unsupervised discretization method. Equal depth binning is an efficient discretization method when data are skewed as is in our case due to the presence of noise.

C. Data Mining and Classification using Association Rules

The process of discovering valuable information from large amounts of data stored in databases, data warehouses, or other information repositories is called data mining. This valuable information can have the form of associations, patterns, changes, anomalies and significant structures [24]. That is, data mining attempts to extract potentially useful knowledge from data. The field of data mining draws upon extensive work in a variety of areas to discover interesting and previously unknown information existing in the data.

In general, data mining tasks can be classified in two categories: *descriptive data mining* and *predictive data mining*. The first describes the dataset in a concise and brief manner and presents general properties of the data; whereas the second constructs one, or a set of, models, performs inference on the available dataset, and attempts to predict the behaviour of new datasets. Data mining is used for class description, association analysis, classification, prediction, clustering, time-series analysis and outlier analysis.

One of the well known data mining techniques is association rule mining [25] and is perhaps the most common form of local-pattern discovery in unsupervised learning systems. Association rule mining discovers association relationships or correlations among a set of items. An association rule has the form $X=>Y$ and is interpreted as "database *transactions* that satisfy X are likely to satisfy Y". X and Y are composed of items and are called itemsets. X is called the antecedent part or body of the rule and Y the consequent part or head of the rule. The most widely used framework for association rule evaluation is the support-confidence framework. The *support* of an association rule $X=>Y$ is the ratio of the transactions which contain the itemsets X and Y to the total number of transactions in the database. The confidence of an association rule $X=>Y$ is the ratio of the transactions that contain the itemsets X and Y to the transactions which contain the itemset X.

A recent field of data mining, classification using association rules [18], applies concepts used in association rule mining to the classification problem. Classification using association rules makes use of a special subset of association rules whose consequents are restricted to the classification class feature. These rules are called Class Association Rules (CARs) and, after their generation, they are used to form a classification model.

An effective algorithm that performs classification using association rules is the Classification Based on Associations (CBA) [18]. CBA first generates as candidate rules all the class association rules exceeding the given support and confidence thresholds using the apriori algorithm [25]. After the rule generation, CBA prunes the set of rules using the pessimistic error rate method [26]. More specifically if rule r's pessimistic error rate is higher than the pessimistic error rate of rule r^- (the latter is obtained by deleting one condition from the conditions of r), then rule r is pruned. This pruning procedure can cut down the generated number of rules substantially. In the testing phase, the above obtained rules are used for classification. When predicting the class label for an example, the best rule (i.e. the one having highest confidence) whose body is satisfied by the example is chosen for prediction.

C. Dataset and Implementation

In order to construct the dataset for training and testing our classification methodology, 11 hours of two-channel ECG recordings from the European Society of Cardiology (ESC) ST-T database [27] were used. More specifically the whole e0104 recording and the first hour of the e0103, e0105, e0108, e0113, e0114, e0147, e0159, e0162, and e0206 recordings were used. These 10 recordings were selected because their ischaemic ECG beats were characterized by significant waveform variability, which was observed by visual inspection. 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. This resulted in a dataset of 86,384 cardiac beats (half from every channel) annotated as normal, ischaemic or artefact. After removing the artefacts, the final dataset contained 76,989 cardiac beats, diagnosed as normal or ischaemic. From those, 2.5% of them, 1,936 beats (982 normal beats and 954 ischaemic) were used for rule mining (training) while the rest 75,053 (38,344 normal beats and 36,709 ischaemic) for testing the performance

Table 1: The discretization intervals (bins) for every one of the five features, produced from the equal depth binning algorithm

Bin	ST_Segment_Deviation	ST_Segment_Slope	ST_Segment_Area	T_Wave_Amplitude-T_Wave_Normal	Age
1	$(-\infty, -0.121)$	$(-\infty, 45.290)$	$(-\infty, -1.716)$	$(-\infty, -0.466)$	35
2	[-0.121, -0.079)	[45.290, 56.340)	[-1.716, -1.183)	[-0.466, -0.240)	46
3	[-0.079, -0.055)	[56.340, 65.050)	[-1.183, -0.780)	[-0.240, -0.051)	47
4	[-0.055, -0.031)	[65.050, 72.700)	[-0.780, -0.420)	[-0.051, -0.014)	55
5	[-0.031, -0.007)	[72.700, 80.860)	[-0.420, -0.076)	[-0.014, 0.005)	60
6	[-0.007, 0.016)	[80.860, 89.300)	[-0.076, 0.212)	[0.005, 0.026)	62
7	[0.016, 0.051)	[89.300, 98.280)	[0.212, 0.645)	[0.026, 0.053)	65
8	[0.051, 0.079)	[98.280, 107.70)	[0.645, 1.004)	[0.053, 0.126)	66
9	[0.079, 0.113)	[107.700, 118.390)	[1.004, 1.565)	[0.126, 0.369)	
10	$[0.113, +\infty)$	$[118.390, +\infty)$	$[1.565, +\infty)$	$[0.369, +\infty)$	

of the classification methodology. The training set was constructed by selecting iteratively the first beat out of a sequence of 40 ones.

The five features, described above, were extracted from every cardiac beat. All the five features were continuous valued, so discretization was applied in the second stage of the methodology using the equal depth binning algorithm. The number of bins *b* was set to 10, so each bin had approximately 194 samples (number of beats in the training set/number of bins). The equal depth binning technique could not be applied properly to the age feature. Having 10 patients, two of them with the same age, we created eight bins, each one containing samples with the same age. In Table 1 we can see the generated bins.

Finally, in the third stage of our methodology the rule mining and classification took place using the CBA algorithm. In brief, CBA extracted class association rules having as antecedent a subset (or all) of the five ECG features with its respected discrete value and as consequent the class of the ECG beat. After testing several values for the parameters of the algorithm, the optimum ones were obtained. The minimum support was set to 0.3% and the minimum confidence to 50%.

Results

The performance of the methodology was measured by means of two different measures: Sensitivity and Specificity. The obtained sensitivity and specificity is 87.91 and 88.02, respectively. The confusion matrix for the best performance of the proposed methodology is shown in Table 2. CBA in the testing phase generated 212 rules. 89 of them predicted ischaemic beats and the rest 123 predicted normal beats

Table 2: Confusion matrix for the best performance of the CBA algorithm.

	Classified as Ischaemic	Classified as Normal
Ischaemic	32271	4438
Normal	4594	33750

Below we can see some indicative rules extracted in the training phase of the methodology. These rules can assist the domain experts by providing them with previously unknown knowledge in quantitative electrocardiography.

Rule 1: IF ST Segment Area>=-0.076
AND ST Segment Area<0.212
AND ST Segment Slope>=98.28
AND ST Segment Slope<107.7
THEN Beat is Normal
(support=1.653%, confidence=96.875%)

Rule 2: IF ST Segment Area>-1.716
AND ST Segment Deviation>-0.121
THEN Beat is Ischaemic
(support=7.696%, confidence=95.973%)

Rule 3: IF ST Segment Area>=-1.716
AND ST Segment Area<-1.183
AND T Deviation-T Normal<-0.466
THEN Beat is Ischaemic
(support=3.564%, confidence 100.000%)

ISSN: 1727-1983 © 2005 IFMBE

Discussion

In this study, classification using association rules was applied for the identification of ischaemic beats in long duration ECGs. The methodology was implemented in a three stage schema: (a) ECG feature extraction, (b) feature discretization, (c) rule generation and beat classification. All stages of the proposed methodology were performed automatically. The equal depth binning algorithm was employed for the discretization of the continuous valued features and the CBA algorithm was tested for classification using association rules.

The proposed beat classification methodology compares favourably with previously reported ones (Table 3). All methods shown in Table 3 were tested using data from the ESC ST-T database, which is a standard reference database of ECG recordings from patients with myocardial ischaemia. However, some of the results reported in Table 3 refer to different subsets of the ESC ST-T database [7,9,10], thus their performance cannot be directly compared.

Table 3: Comparison of the performance of several methods for ischaemic beat detection evaluated using a subset of the ESC ST-T database.

Reference	Se[1] (%)	Sp[2] (%)	Acc[3] (%)
Papaloukas et al. 2002 [5]	70	63	
Papaloukas et al. 2002 [8]	90	90	
Stamkopoulos et al. 1998 [7]	75	79	
Maglaveras et al. 1998 [9]			56
Papadimitriou et al. 2001 [10]			73
Papadimitriou et al. 2001 [10]			77
Papadimitriou et al. 2001 [10]			80
Goletsis et al. 2003 [11]	90	89	
Goletsis et al. 2004 [12]	91	91	
Current work	**88**	**88**	**88**

[1]Se: Sensitivity
[2]Sp: Specificity
[3]Acc: Accuracy

It should be mentioned that most of the previously reported techniques are based on neural network approaches. Such methods exhibit a serious drawback compared with our association rule approach, due to their inability to provide explanations for their classification decisions. In contrast, due to the rule-based nature of the proposed methodology, it satisfies this important requirement, and it is able to provide for each beat, the reason (rule) leading to each decision.

A limitation of our methodology is the requirement of a representative training set in order to extract reliable rules. In addition, the utilization of association 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 association rule mining procedure should be considered as exploratory and hypothesis-generating.

Further improvement might focus on the utilization of more features extracted from the ECG describing better each beat and also the employment of other patient's clinical data. In addition, coherent information could be used in order to diagnose a beat based on the previous diagnosed beats. Moreover our system could be adapted to address other cardiac abnormalities, such us arrhythmias (arrhythmic beat detection).

Conclusions

We presented a novel methodology for the automated identification of ischaemic beats that employs association rules for classification. Our approach has comparable or better performance than other reported methods when tested with the ESC ST-T database. The main advantage of the proposed methodology is that it can provide interpretation for the decisions. The classification accuracy of the proposed methodology indicates that it could be part of an integrated system for myocardial ischaemia diagnosis. The overall ischaemic episode detection system should be assessed with several ECGs extracted from ambulatory recordings or continuous monitoring in the coronary care unit. The possibility of applying our methodology to clinical practice and evaluating its performance in real clinical conditions is of great interest.

Acknowledgments

This work was supported by the European Union as part of the project NOESIS: Platform for wide scale integration and visual representation of medical intelligence (IST-2002-507960).

References

[1] GOLDMAN M. J. (1982): 'Principles of clinical electrocardiography, 11th ed. Los Altos, CA: LANGE Medical Publications,.

[2] PITAS I., STRINTZIS M. G., GRIPPAS S., and XEROSTYLIDES C. (1983): 'Machine classification of ischemic electrocardiograms, Proc. of the IEEE Mediterranean Electrotechnical Conf. (MELECON), Athens, 1983.

[3] PAPALOUKAS C., FOTIADIS D. I., LIKAS A., and MICHALIS L. K. (2002): 'An expert system for ischemia detection based on parametric modeling and artificial neural networks', Proc. of Eur. Medical and Biological Eng. Conf., 2002, p. 742-743.

[4] SENHADJI L., CARRAULT G.,. BELLANGER J. J, and PASSARIELLO G. (1995): 'Comparing wavelet transforms for recognizing cardiac patterns', *IEEE Eng. Medicine and Biology* Magazine, **14**, pp. 167-173.

[5] PAPALOUKAS C., FOTIADIS D. I., LIAVAS A. P., LIKAS A., and MICHALIS L. K. (2001): 'A knowledge-based technique for automated detection of ischemic episodes in long duration electrocardiograms', *Med. Biol. Eng. Comput.*, **39**, pp. 105-112.

[6] PAPALOUKAS C.,. FOTIADIS D. I, LIKAS A.,. STROUMBIS C. S, and MICHALIS L. K. (2002): '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.*, **35**, pp. 27-34.

[7] STAMKOPOULOS T., DIAMANTARAS K., MAGLAVERAS N., and STRINTZIS M. (1998): 'ECG analysis using nonlinear PCA neural networks for ischemia detection', *IEEE Trans. on Signal Process.*, **46**, pp. 3058–3067.

[8] PAPALOUKAS C., FOTIADIS D. I., LIKAS A., and MICHALIS L. K. (2002): 'An ischemia detection method based on artificial neural networks', *Artif. Intell. Med.*, **24**, pp. 167-178.

[9] MAGLAVERAS N., STAMKOPOULOS T., PAPPAS C., and STRINTZIS M. (1998): 'ECG processing techniques based on neural networks and bidirectional associative memories', *J. Med. Eng. Technol.*, **22**, pp. 106-111.

[10] PAPADIMITRIOU S., MAVROUDI S, VLADUTU L., and BEZERIANOS A. (2001): 'Ischemia detection with a self-organizing map supplemented by supervised learning', *IEEE Trans. Neural Networks*, **12**, pp. 503-515.

[11] GOLETSIS Y., PAPALOUKAS C., FOTIADIS D. I., LIKAS A., and MICHALIS L. K. (2003): 'A multicriteria decision based approach for ischemia detection in long duration ECGs', Proc. IEEE EMBS 4th Int. Conf. Information Technology Applications in Biomedicine (ITAB 2003), 2003, p. 230–233.

[12] GOLETSIS Y., PAPALOUKAS C., FOTIADIS D. I., LIKAS A., and MICHALIS L. K. (2004): 'Automatic Ischemic beat classification using genetic algorithms and multicriteria decision analysis', *IEEE Trans. Biomedical Engineering*, **51**, pp. 1717- 1725.

[13] SILIPO R., TADDEI A., and MARCHESI C. (1994): 'Continuous monitoring and detection of ST-T changes in ischemic patients', Proceedings of the IEEE Computers in Cardiology, 1994, p. 225–228.

[14] SILIPO R., and MARCHESI C. (1998): 'Artificial neural networks for automatic ECG analysis', *IEEE Trans. Signal Process.*, **46**, pp. 1417-1425.

[15] VILA J., PRESEDO J., DELGADO M., BARRO S., RUIZ R, and PALACIOS F. (1997): 'SUTIL: Intelligent ischemia monitoring system', *Int. J. Med. Inform.*, **47**, pp. 193-214.

[16] JAGER F., MARK R. G., MOODY G. B., and DIVJAK S. (1992): 'Analysis of transient ST segment changes during ambulatory monitoring using the Karhunen-Loève transform', Proceedings of the IEEE Computers in Cardiology, 1992, p. 691-694.

[17] LEMIRE D., PHARAND C., RAJAONAH J.-C, DUBE B., AND LEBLANC A. R. (2000): 'Wavelet time entropy, T wave morphology and myocardial ischemia', *IEEE Trans. Biomed. Eng.*, **47**, pp. 967-970.

[18] LIU B., HSU W., and MA Y. (1998): 'Integrating classification and association rule mining', Proc. of the 4th International Conf. on Knowledge Discovery and Data Mining (KDD-98), New York, USA, 1998, p. 80-86.

[19] ANTONIE M., ZAOANE O. R., and COMAN A. (2003): 'Associative Classifiers for Medical Images', Lecture Notes in Artif. Intel. 2797, Mining Multimedia and Complex Data, Springer-Verlag, 2003, p. 68-83.

[20] DART T., CUI Y., CHATELLIER G., and DEGOULET P. (2003): 'Analysis of hospitalised patient flows using data-mining', *Stud. Health Technol. Inform.* **95**, pp. 263-268.

[21] EXARCHOS T. P., TZALLAS A. T., FOTIADIS D. I., KONITSIOTIS S. and GIANNOPOULOS S. (2005): 'A Data Mining Based Approach for the EEG Transient Event Detection and Classification', Proc. 18th IEEE International Symposium on Computer-Based Medical Systems, Dublin, Ireland, p. 35-40.

[22] DASKALOV K., DOTSINSKY I. A., and CHRISTOV I. I. (1998): 'Developments in ECG acquisition, preprocessing, parameter measurement, and recording', *IEEE Eng. Med. Biol.,* **17**, pp. 50-58.

[23] CATLETT J. (1991): 'On changing continuous attributes into ordered discrete attributes', Proc. of the Fifth European Working Session on Learning, Berlin: Springer-Verlag, p. 164-178.

[24] HAN J. and KAMBER M. (2001): 'Data Mining Concepts and Techniques',. Morgan Kaufmann, San Fransisco, CA.

[25] AGRAWAL R., and SRIKANT R. (1994): 'Fast algorithms for mining association rules', Proc. 20th Int. Conf. Very Large Data Bases, Santiago, Chile, 1994, p. 487-499.

[26] QUINLAN J. R. (1993): 'C4.5: Programs for Machine Learning'. Los Altos, California: Morgan Kaufmann.

[27] European Society of Cardiology (1991): European ST-T database directory. Pisa: S.T.A.R.