An Abnormal ECG Beat Detection Approach for Long-term Monitoring of Heart Patients based on Hybrid Kernel Machine Ensemble

Peng Li, Kap Luk Chan, Sheng Fu, and S. M. Krishnan

Biomedical Engineering Research Center, Nanyang Technological University, Research Techno Plaza, 50 Nanyang Drive, XFrontiers Block, 6th Storey, 637553 Singapore
lipeng@pmail.ntu.edu.sg, {eklchan, msfu}@ntu.edu.sg

Abstract. In this paper, a novel hybrid kernel machine ensemble is proposed for abnormal ECG beat detection to facilitate long-term monitoring of heart patients. A binary SVM is trained using ECG beats from different patients to adapt to the reference values based on the general patient population. A one-class SVM is trained using only normal ECG beats from a specific patient to adapt to the specific reference value of the patient. Trained using different data sets, these two SVMs usually perform differently in classifying ECG beats of that specific patient. Therefore, integration of the two types of SVMs is expected to perform better than using either of them separately and that improving the generalization. Experimental results using MIT/BIH arrhythmia ECG database show good performance of our proposed ensemble and support its feasibility in practical clinical application.

1 Introduction

Electrocardiographic (ECG) signal is a recording of the cardiac activities, which is usually used by cardiologists to obtain information about the performance of heart function. Some typical ECG beats are illustrated in Figure 1. The analysis of heart beat cycles in ECG signal is very important for long-term monitoring of patients' heart conditions at patients' homes through a telemedicine network. However, it is very costly for the cardiologists to analyze the ECG recording beat by beat because the ECG recording may last for hours. Therefore, it is justified to develop a computer-assisted technique to examine and annotate the ECG recording, so to facilitate review by doctors. This computer annotation will assist doctors to select only the abnormal beats for further analysis.

Many algorithms have been applied to ECG beat cycle analysis, such as Kohonen self-organizing map [1, 2], learning vector quantization [2], multilayer perceptron [4], neural-fuzzy system [5] and Support Vector Machines (SVMs) [6, 7]. One of the major challenges faced by these ECG beat recognition algorithms is the large variation in the morphologies of ECG signals from patient to patient. The ranges of "normal beat" are different among the patients, which leads to the so-called poor generalization problem, i.e., an ECG detector finely tuned to the

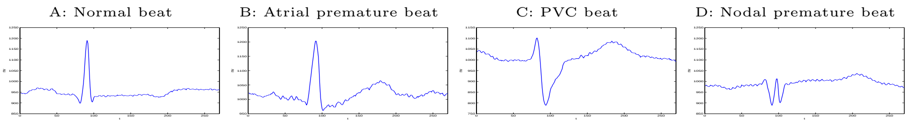

Fig. 1. Examples of ECG beats. A is a normal beat. The others are abnormal.

training data from a large group of people may perform poorly when classifying the ECG beats of an individual patient. Hu et.al. attempted to solve this problem using a mixture of expert approach [2]. A mixture of expert structure was formed by combining the knowledge of a global expert trained using ECG data from a large database and a local expert trained using 3 to 5 minutes of ECG signals from a specific patient. The mixture of expert system was then used to classify the ECG beats from the specific patient, and the classification performance was improved over the global expert. However, the major drawback of such an approach is that a local expert has to be constructed for each patient, and the ECG recording of each patient has to be annotated by a doctor in order to train the local expert, even with only 5 minutes of a patient's ECG recording. Such annotation process is very costly and discourages the practical application of this approach [2, 3].

In this paper, a hybrid kernel machine ensemble approach is proposed to address such generalization problem. One-class Support Vector Classifier (νSVC) [8] is a one-class classifier whose goal is to find a decision region to include patterns from one class - called targets, and exclude the patterns from the other classes - called outliers. It is a non-discriminative recognition-based model. A particular advantage of νSVC is that it can be trained using data from one class only. In the context of long-term monitoring of some heart patients, the normal ECG beats usually dominate the ECG recordings, that is, the number of abnormal ECG beats is far less than that of the normal ones such as for patients suspected to suffer or suffering from asymptomatic heart failure, congestive heart failure, cardiac dysfunction etc. Furthermore, there are many kinds of abnormal ECG beats corresponding to different cardiac diseases, such as atrial premature beats, ventricular escape beats, fusion of ventricular and normal beats, supraventricular premature beats and premature ventricular contraction (PVC) beats, etc. Some of the typical abnormal beats and a normal ECG beat are illustrated in Figure 1. These abnormal ECG beats appear different in morphology. On the other hand, the normal ECG beats usually appear similar to each other and show less variation, which implies that the concept "normal" is more compact compared to that of the concept "abnormal" and thus easier to be learned using few samples. Since normal ECG beats can be easily obtained from patients, a νSVC can then be trained using only the normal ECG beats from each specific patient. Incorporated with the local information of the patient, such a trained νSVC can be used as a specific reference adapted to that patient. On the other hand, binary Support Vector Machines ($2SVC$) is known to be a powerful discriminative model [9]. It can be trained using a large

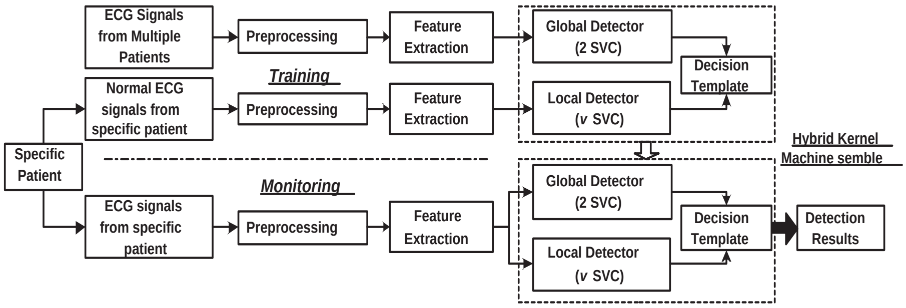

Fig. 2. Flowchart of the proposed framework for abnormal ECG beat detection.

database which consists of the ECG beats from a large group of people. Such a $2SVC$ incorporate the global information of the group of people, and thus it can be regarded as the reference values based on the general patient population. Due to different information learned by these two SVMs, they usually perform differently in classifying the ECG beats in the long-term ECG recording of the specific patient. Furthermore, νSVC is a non-discriminative recognition-based model and $2SVC$ is a discriminative model. Due to the complementary nature of such two types of SVMs, integration of the two types of kernel machines using an ensemble is expected to perform better than using either of them separately. Here Decision Template (DT) [10] is investigated as the fusion rule to integrate these two hybrid SVMs. Experimental results using MIT/BIH arrhythmia ECG database [11] show that our proposed patient-adaptable hybrid kernel machine ensemble outperforms both the local νSVC and the global $2SVC$. Compared to [2], such a hybrid kernel machine ensemble approach can relieve the doctors from annotating the ECG beats from each patient beat by beat and shows better or at least comparable performance to [2] and [3] in detecting the abnormal ECG beats of the specific patient without requiring annotated ECG from a specific patient, thus support its feasibility in practical clinical application.

2 Proposed Methodology

Figure 2 illustrates the flowchart of the proposed framework for ECG beat detection. The details are as follows.

2.1 Global Detector – Binary Support Vector Classifier

SVM is increasingly used in many medical applications and has been shown to achieve better performance than traditional classifiers [9]. SVM has good generalization ability by finding an optimal separating hyperplane which minimizes the classification errors made on the training set while maximize the "margin" between different classes. Given a two-class (labelled by $y_i = \pm 1$) training set

$X = \{x_i \in R^d | i = 1, 2, \cdots, N\}$ with N samples which are nonlinearly separable. The data are mapped to another feature space in which the data can be separated by an optimal separating hyperplane expressed as

$$f(x) = \sum_{i=1}^{N} y_i \beta_i K(x_i, x) + b \tag{1}$$

where b is a bias item, β_is $(i = 1, 2, \cdots, N)$ are the solution of a quadratic programming problem to find the maximum margin, $k(\cdot)$ is a kernel function, such as a Gaussian Radial Basis Function (RBF) kernel $k(x_i, x_j) = e^{\frac{-\|x_i - x_j\|^2}{\sigma}}$. There are only a few training samples whose β_is are non-zero, called the Support Vectors, which are either on or near to the separating hyperplane. The decision boundary, i.e. the separating hyperplane is along these support vectors, whose decision values $f(x)$ in (1) approach zero. Compared with the support vectors, the decision values of positive samples have larger positive values and those of negative samples have larger negative values. Therefore, the amplitude of the decision value can be regarded as the confidence of the SVM classifier. The larger the amplitude of $f(x)$, the more confident is the classification.

As a powerful discriminative model, $2SVC$ is very suitable as a global detector trained using ECG beats from a group of people. Recently, the application of SVM to ECG beat classification was investigated [6, 7]. These studies show that SVM can provide better classification performance than the other classifiers. In [7] only linear support vector machine was investigated. But the different classes of ECG beats may not be linearly separable. In [6] a multiple classifier network was introduced in which support vector machine was the basic classifier. But the selection of the hyperparameters was not investigated fully, which may limit the performance of SVMs. Therefore, there is still some scope to further improve the ECG beat classification using SVMs.

2.2 Local Detector – ν-Support Vector Classifier

ν–Support Vector Classifier (νSVC) is a kind of support vector machine [8] which can be used as an one-class classifier. Different from $2SVC$s, only data from one-class is used in the training of a νSVC, which makes it very suitable as a local detector to be trained using only "normal" ECG beats in the scenario of long-term monitoring of heart patients. νSVC is a recognition-based model rather than a discriminative model because it tries to estimate the support of the density of the target data [8]. Given a set of target data $X = \{x_i \in R^d | i = 1, 2, \cdots N\}$, the goal of νSVC is to find a decision function $f(x)$ such that most of the target data will have $f(x) \geq 0$ while most of the outliers will have $f(x) < 0$. The target data are mapped into a higher-dimensional space called feature space $\phi(x)$ in which the dot product can be computed using some kernel function, such as a RBF kernel. The mapped target data are separated from the origin corresponding the outliers with maximum margin using a hyperplane, which can be found by solving a quadratic programming problem [8]. The decision function corresponding to the hyperplane is

$$f(x) = \sum_{i=1}^{N} \beta_i K(x_i, x) + b \tag{2}$$

Similar to $2SVC$s, the amplitude of the decision function of νSVC is proportional to the confidence of the classification.

There are two hyperparameters to be tuned in νSVC and $2SVC$ using RBF kernel, the width parameter of the RBF kernel and the regularization parameter where the latter is used to control the tradeoff of the errors. The hyperparameters of $2SVC$ can be optimized using cross validation on the training set. The values of the hyperparameters are chosen so that the error of both target class and outlier class on the validation set is minimized. As for νSVC, no information about the "abnormal" class is available. Such problem can be solved by generating artificial outliers [12]. Given a set of target samples, some outlier samples are generated randomly with the assumption that the outliers are uniformly distributed in a hypercube or hypersphere. The union of targets and generated outliers is used as a validation set. Some targets will be excluded by the decision boundary to make a tighter boundary. Therefore, the νSVC can be trained properly even if there are some abnormal ECG beats in the training samples.

2.3 Learning An Ensemble to Integrate Two Hybrid SVMs

Ensemble is often understood as mixture of experts, classifier fusion and combination of multiple classifiers, etc [10]. It is a mechanism to combine a set of classifiers so that the resulted ensemble has superior classification performance over the individual classifiers in the ensemble. The necessary condition to the success of the ensemble is that the outputs of individual classifiers to the same inputs must be diverse [13]. The diversity of the two classifiers can be evaluated using Plain Disagreement Measure (P) [14]:

- Plain Disagreement Measure (P): $P = \frac{N_{dis}}{N_{all}}$, where N_{dis} is the number of samples that two classifiers disagree and N_{all} is the number of all samples in the validation set. P varies from 0 to 1. The larger the value of P, the higher the diversity. This measure was recommended for ensemble feature selection in [14].

The diversity of the classifiers can be obtained by using different training set, feature subset, classifiers and ensemble rules. Since the νSVC and $2SVC$ are trained using local information and global information respectively, the training sets of such two kernel machines can be considered diverse. Furthermore, the different nature of the two SVMs can help to increase the diversity further. Therefore, the ensemble of such two kernel machines is expected to improve the classification compared to either of the two SVMs.

Many fusion rules have been developed. In this study, Decision Template (DT) [10] was employed to integrate the two hybrid SVMs.

– Decision template: The decision template DT_j $(j = 1, 2)$ for class $y_j \in \{-1, +1\}$ is the average of the outputs of individual classifiers in the validation set to class y_j. The ensemble DT assigns an input x with the label given by the individual classifier whose Euclidean distance to the decision template DT_j is the smallest. The "normal" data from the specific patient and the generated artificial outliers can be used as a validation set so that the decision template of two SVMs can be learned.

Oracle (ORA) is the optimal case or an upper bound which an ensemble can reach rather than a real ensemble. It assigns a correct class label to the pattern iff at least one individual SVM produces a correct class label [10]. Here it is used for comparison purpose only.

3 Experiments and Results

3.1 Data Preprocessing and Feature Extraction

The ECG signals come from 44 recordings of the original MIT/BIH ECG arrhythmia database [11]. The original signal consists of two leads sampled at $360Hz$. The data from lead 1 were used in this study. The signal was first processed using two averaging filters to suppress noises [15]. Then the baseline shift of the ECG signal can be obtained using two consecutive median filters whose widths are 200ms and 600ms respectively. The baseline was subtracted from the original signal and the resulted signal was then baseline-corrected [3]. The R-peak of the ECG signal can be detected using the first derivative of the ECG signal as in [15]. A window of 180 samples in length was taken to each ECG beats such that the window covers most of the information from a particular cardiac cycle, as shown in Figure 1. The signal in each window was then down-sampled uniformly to form a feature vector of 38-dimensions. It has been shown that R-R interval is useful in recognition of some abnormal ECG beats [2]. Therefore, it was also included in this study by appending it to the 38-dimensional feature vector. The length of the feature vector to represent the ECG beat is then 39. In order to deal with the variation of the amplitudes of ECG signals among the different patients, the feature vectors were divided by the mean value of R peaks in the training data of each patient, so that the maximum value of the each ECG window was around 1. The normalized ECG feature vectors were then used in the classification.

3.2 Training and Test Procedure

MIT/BIH arrhythmia ECG database consists of 48 annotated recordings from 47 subjects and each recording is about 30 minutes in length. The labels in the annotation file made by expert cardiologists are used as the ground truth in training and evaluating the classifiers. The ECG beats annotated as "normal" (NOR) are taken as the target class in the current research. All of the other

Table 1. Results (average $\pm$ standard deviation) of abnormal ECG beat detection.

Classifiers	$2SVC$	νSVC	DT	ORA
BCR	0.774 $\pm$0.280	0.881 $\pm$0.146	0.912 $\pm$0.132	0.971 $\pm$0.048
SEN	0.808 $\pm$0.287	0.819 $\pm$0.210	0.876 $\pm$0.195	0.964 $\pm$0.075
SPE	0.819 $\pm$0.252	0.972 $\pm$0.051	0.967 $\pm$0.055	0.980 $\pm$0.044

beats are regarded as outlier class or "abnormal" class, including atrial premature beats, nodal premature beats, premature ventricular contraction beats and ventricular escape beats etc.

Four recordings (102, 104, 107, and 217) including paced beats of the MIT/BIH arrhythmia database are excluded from the study. The ECG beats from the other 44 recordings were split into two parts. The training set of the global $2SVC$ beat detector consists of 4000 normal beats and 4000 abnormal beats from 22 recordings, which is called TN_G. Five-fold cross validation on TN_G was used to select hyperparameters for $2SVC$. The ECG beats in each of the remaining 22 recordings were split into two subsets. The training sets of the local νSVC detectors were the normal beats from the first 5 minutes in each recording. Those ECG beats of the remaining 25 minutes in each recording were used as the test sets.

3.3 Evaluation Measures

Three measures are used to evaluate the performance of ECG beat classification, including sensitivity, specificity and balanced classification rate. *Sensitivity* (SEN) is the fraction of abnormal ECG beats that are correctly detected among all the abnormal ECG beats. *Specificity* (SPE) is the fraction of normal ECG beats that are correctly classified among all the normal ECG beats. *Balanced Classification Rate* (BCR) is the geometric mean of the SPE and SEN. $BCR = \sqrt{SEN \cdot SPE}$. Only when both SEN and SPE have large value can BCR has a large value. Therefore, the use of BCR can have a balanced performance in the evaluation of the classifiers which favors both higher SPE and higher SEN. It is very suitable to this study since the data sets are imbalanced.

3.4 Experimental Results

The classification results of using global $2SVC$, local νSVC and the ensembles are illustrated in Table 1, which are averaged over 22 test sets. The improved BCRs of the ensembles (DTs and ORAs) over the global $2SVC$ and local νSVCs in the 22 test sets are illustrated in Figure 3. The relations between the improved BCRs of ensembles (DTs) over the global and local SVMs and the plain disagreement measure are illustrated in Figure 4.

4 Discussions

It can be observed from Table 1 that the patient-adaptable kernel machine ensembles outperforms both the global $2SVC$ trained using large database TN_G

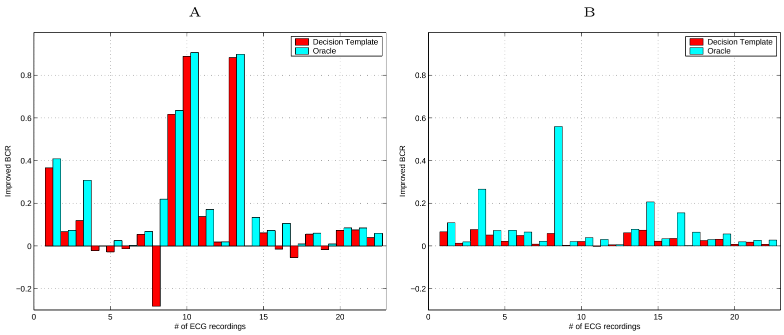

Fig. 3. The improved BCRs of the ensembles (DTs and ORAs) over the global $2SVC$ (A) and local νSVCs (B) of the 22 ECG recordings.

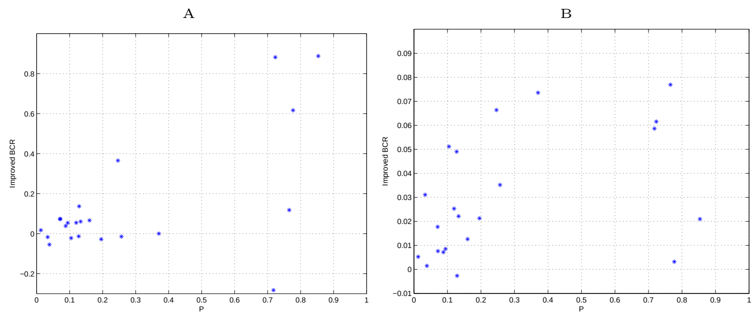

Fig. 4. The relation between the improved BCR of the ensembles (DTs) over the global $2SVC$ (A) and local νSVCs (B) and the plain disagreement measure.

after excluding the specific patients to be tested, and using the local νSVCs adapted to the specific patients. The improved average BCRs by the ensemble DT is 13.8% over the global $2SVC$ and 3.1% over the νSVCs. Furthermore, variance of the ensemble is less than either the global $2SVC$ or the local νSVC. This supports our claim that the hybrid kernel machine ensemble is superior than either the global $2SVC$ or the local νSVC in the abnormal ECG beat detection of the specific patients.

The local detector νSVCs outperform the global detector $2SVC$ in the abnormal ECG beat detection of the specific patients. The average BCR of the νSVC is 10.7% higher than that of the $2SVC$. It indicates that the local information is more important in the classification of the ECG beats from specific patient. Incorporation of such local information can help to deal with the difference between the distribution of training data and test data, hence help to improve the generalization.

The improved performance of the ensembles varies among the 22 test sets. The DT ensemble outperforms νSVCs in all of the 22 test sets in Figure 3

(B) and it outperforms the global $2SVC$ in 15 of the 22 test sets in Figure 3 (A). An exception is observed in the eighth recording. Here the BCR of DT is only 42.4% which is far less than 70.6% of $2SVC$, though it is still greater than 36.6% of νSVC. We observe that the SPE and SEN of νSVC in this recording were 100% and 13.4% respectively, which means that it correctly detected all the abnormal ECG beats but made a lot of false detection of the normal ECG beats. Since the performance of $2SVC$ in this recording is also not good, it may imply that there is a large overlap between the two class which prevents νSVC from discriminating the two class by modelling the "normal" class only. The low BCR of DT was then resulted from the low SEN of the local νSVC. Furthermore, there is a 6.0% gap in terms of BCR between the DT ensemble and the oracle which means there are still some space for further improvement of the DT ensemble.

It can be observed in Figure 4 that the plain disagreement measure is related to the improved BCRs. The larger the value of P, usually the more the improvement. However, there are only 22 test sets and the classification problem is different in each test set. Further testing is necessary to make the result more statistically reliable. This observation is roughly in agreement with [14] though it was used for ensemble feature selection purpose in [14].

Hu et.al. [2] concentrated on the classification between normal beats and ventricular ectopic beats using a mixture of two classifiers. The sensitivity and specificity achieved are 82.6% and 97.1%, which means it BCR is less than 90%. Philip et.al. [3] claimed to have achieved comparable performance to [2]. The BCR of our hybrid kernel machine ensemble is greater than 91% although only some "normal" ECG beats from each patient are used to train the local νSVCs. Therefore, our proposed method shows better or at least comparable performance compared to [2] and [3]. Another advantage of our method is that it can relieve the doctors from annotating the ECG beats one by one which is needed in [2].

5 Conclusion

In this paper, a new hybrid kernel machine ensemble is proposed to detect abnormal ECG beats for long-term monitoring of heart patients. A νSVC can be trained using only some "normal" ECG beats from a specific patient to obtain local information. A $2SVC$ can be trained using a large database which consists of ECG beats from many patients to obtain global information. Due to the different nature of two types of kernel machines, the ensemble of these two SVMs using decision template is able to outperform either of the two kernel machines in detecting the abnormal ECG beats from the specific patients. This approach can relieve the doctors from annotating the training ECG data beat by beat to train a local classifier and help improve the generalization. Experimental results using 44 ECG recordings of MIT/BIH arrhythmia database demonstrate the good performance of our proposed hybrid kernel machine ensemble and suggest its feasibility in practical clinical application.

6 Acknowledgements

The authors wish to acknowledge the support by Distributed Diagnosis and Home Healthcare project (D2H2) under Singapore-University of Washington Alliance (SUWA) Program in Bioengineering and Biomedical Engineering Research Center at Nanyang Technological University in Singapore. Also the first author would like to thank for the research scholarship from the School of Electrical and Electronic Engineering, Nanyang Technological University.

References

1. Vladutu, L., Papadimithou, S., Mavroudi, S., Bezerianos, A.: Ischemia detection using supervised learning for hierarchical neural networks based on kohonen-maps. Proceedings of the 23rd Ann. Inter. Con. of the IEEE EMBS. **2** (2001) 1688–1691
2. Hu Y. H., Palreddy S., Tompkins W. J.: A patient-adaptable ECG beat classifier using a mixture of experts approach. IEEE Trans. on Biomed. Eng. **44** (1997) 891–900
3. Philip de Chazal, O'Dwyer, M., Reilly, R.B.: Automatic classification of heartbeats using ECG morphology and heartbeat interval features. IEEE Trans. on Biomed. Eng. **51** (2004) 1196–1206
4. Guler I., Ubeyh E. D.: ECG beat classifier designed by combined neural network model. Pattern Recognition **38** (2005) 199–208
5. Engin M.: ECG beat classification using neuro-fuzzy network. Pattern Recognition Letters **25** (2004) 1715–1722
6. Osowski, S., Hoai, L.T., Markiewicz, T.: Support vector machine-based expert system for reliable heartbeat recognition. IEEE Trans. on Biomed. Eng. **51** (2004) 582–589
7. Millet-Roig, J., Ventura-Galiano, R., Chorro-Gasco, F.J., Cebrian, A.: Support Vector Machine for Arrhythmia Discrimination with Wavelet-Transform-Based Feature Selection. Computers in Cardiology (2000) 407–410
8. Scholkopf, B., Platt, J. C., Shawe-Taylor J., Smola, A. J., Williamson, R. C.: Estimating the support of a high-dimensional distribution. Neural Computation **13** (2001) 1443–1471
9. Cristianini, N., Shawe-Taylor, J. : An Introduction to Support Vector Machines. Cambridge University Press (2000)
10. Kuncheva, L., Bezdek, J., Duin, R.: Decision templates for multiple classifier fusion: an experimental comparison. Pattern Recognition. **34** (2001) 299–314
11. Massachusetts Institute of Technology: MIT-BIH Arrythmia ECG database. http://ecg.mit.edu/.
12. Tax, D. M. J., Duin R. P. W.: Uniform object generation for optimizing one-class classifiers. Journal of Machine Learning Research (2002) 155–173
13. Kuncheva, L. I., Whitaker, C. J.: Measures of diversity in classifier ensembles. Machine Learning**51** (2003) 181–207
14. Tsymbal A., Pechenizkiy M., Cunningham P.,: Diversity in search strategies for ensemble feature selection. Int. Journal on Information Fusion **6** (2004) 83–98
15. Christov, Ivaylo: Real time electrocardiogram QRS detection using combined adaptive threshold. BioMedical Engineering OnLine **3** (2004)