A Mixed SVM-Based Hierarchical Learning Approach for Abnormal ECG Beat Recognition

Peng Li[1], Kap Luk Chan[1], Y. W. Chan[2]

[1] Biomedical Engineering Research Centre, Nanyang Technological University, Singapore 637553

[2] Department of Anaesthesia, Singapore General Hospital, 169608

Abstract—**A mixed Support Vector Machine (SVM)-based hierarchical learning approach is proposed to detect abnormal ECG beats. A global bi-class support vector classifier is first trained using ECG beats from different patients in a database. Then, a local novelty detector SVM is trained using only normal ECG beats from a specific patient. The fusion of the global and local classifiers can significantly improve the classification. Preliminary experimental results using MIT/BIH ECG database demonstrated good performance of our proposed approach.**

Keywords—**Support vector machines, ECG beat recognition, novelty detection, decision fusion**

I. Introduction

Electrocardiographic (ECG) signal is the recording of the cardiac activities, which provides reliable information about the performance of the heart. The analysis of heart beat cycles in ECG is very important for monitoring and diagnosis of patients' heart conditions in an intensive care unit or at patients' homes through a telemedicine network. Many algorithms have been applied to ECG beat cycle analysis, such as Kohonen self-organizing map, Learning Vector Quantization, multilayer perceptron, fuzzy and neural-fuzzy system, hybrid system and support vector machines etc. [1, 2, 3, 4]. However, the large variation in the morphologies of ECG signals from different patients is one of the challenges to the ECG beat recognition algorithms. That is, a finely tuned ECG detector to training set may perform badly to ECG beats of a different patient. To overcome this poor generalization problem, a mixture of expert approach was proposed in [5]. 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. When the mixture of expert system was used to classify the ECG beats from a specific patient, the classification performance was significantly improved compared to both the local and global expert. However, some drawbacks need to be overcome. First, it is very costly to develop a local expert for each patient, even with only 5 minutes of the patient' ECG record, because the ECG record has to be labeled by a medical expert in order to train the local expert. Second, the classifiers used as the global and local expert in the original work are Self-Organizing Map (SOM) and Learning Vector Quantization (LVQ). Both of them are unsupervised learning methods. It is well-known that the classification performance of the unsupervised learning methods is generally inferior to that of the supervised learning methods without using the labeling message. Since there are plenty of labeling information available in the large database such as MIT/BIH arrhythmia database [10], the classification performance may be greatly improved if powerful supervised learning methods are used. Third, only one method of combining of multiple classifiers is investigated in [5], the application of other fusion methods needs to be further investigated.

In this study, a mixed support vector machine-based hierarchical learning method is proposed to detect abnormal ECG beats from the normal ones. A support vector classifier is trained as the global beat detector and a ν-support vector classifier is trained as the local beat detector using only about 5 minutes of normal beats from a specific patient's ECG record. Then both classifiers are used to detect the ECG signals from the specific patient. Finally, a fusion rule is used to fuse the output of the two classifiers and gives the final decision. The classification of both global and local detectors can be improved using two ensemble learning methods, among which weighted averaging shows slightly better performance. The proposed method is tested using the MIT/BIH ECG arrhythmia database and the experimental results demonstrated the good performance of our proposed method.

II. Methodology

The flowchart of the proposed method is illustrated in figure 1. The details of the method are introduced as follows.

A. Support Vector Machines

SVM is increasingly used in many medical applications and has been shown to achieve better performance than traditional classifiers [6]. 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 (labeled by $y_i=\pm 1$) training set X with N samples; it may be difficult to linearly separate the two class in the original feature space directly. The data are mapped to another feature space where the data can be separated by an optimal separating hyperplane expressed as:

$$f(x) = w\phi(x) + \rho = \sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + \rho \qquad (1)$$

where $K(.)$ is a kernel function, ρ is a bias, α_i *(i=1, 2, ··, N) are* the solutions of a quadratic programming problem to find the maximum margin.

There are only a few training samples whose α_i s are non-zero. They are 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 magnitude of the decision value can also be regarded as the confidence of the SVM classifier. The larger the magnitude of *f(x)*, the more confident is the classification.

Recently, the application of SVM to ECG beat classification was investigated [2, 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 [2] a multiple classifier network was introduced in which support vector machine was the basic classifier. But the model selection was not investigated carefully, which may limit the performance of SVM. Therefore, there is still some scope to further improve the ECG beat classification using SVM.

B. ν-Support Vector Classifier

ν - Support Vector Classifier (ν-SVC) is a kind of SVM [8]. It is a novelty detection technique. The difference between ν-SVC and classical SVM is that only data from one class – called targets, are used in the training stage. Without the use of the data from the other class –outliers, the target data are mapped into the feature space corresponding to a kernel function and separated from the origin with maximum margin by a hyperplane

$$f(x) = w\phi(x) + \rho = \sum_{i=1}^{N} \alpha_i K(x_i, x) + \rho \qquad (2)$$

For a new point, the value of *f(x)* is *+1* if it belongs to the target. Otherwise the value of *f(x)* is *-1*. Compared to the binary class SVM (2-SVC), the classification performance of ν-SVC may be a little inferior due to the absence of the information from the outlier class in training, but it leads to a more compact classifier which not only needs less computation in training but also is robust when the training data set are seriously unbalanced. Therefore, it is more suitable to be used as a local classifier for ECG beat classification since no labeling of the data is needed.

C. Ensemble learning of 2-SVC and ν-SVC

There are many ensemble strategies for combing classifiers, such as majority voting, maximum rule, etc. [9].

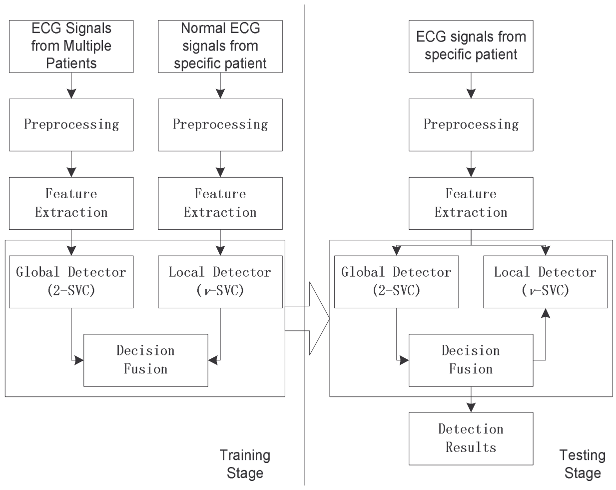

Fig.1. The flowchart of the proposed method.

Given *M* individual classifiers with decision value $f_i(x)$ and label y_i (±1), the goal is to find the final prediction y_{en}. We investigated 2 fusion rules in the experiment. One is max rule: $y_{en}=y_j$, where

$$\mathrm{j} = \arg\max_i |f_i(x)| \qquad (3)$$

The second one is the weighted average of the decision value of *M* individual classifiers:

$$y_{en} = \mathrm{sgn}(\sum_{i=1}^{M} w_i f_i(x)) \qquad (4)$$

where the weight w_i is based on the average classification accuracy rate of the *i-th* individual SVM classifier in the validation set. One advantage of using these two fusion rules is that they are easy to implement because no further training is required. While in [5], a set of extra validation data is needed to learn the gating function of the mixture of experts.

The 2-SVM and ν-SVC can be trained using ECG beats form a large database and normal ECG beats from a specific patient separately. The trained classifiers are then used to detect the beats from the same patient and the final decision is the ensemble of the two basic classifiers.

III. Experiments and Results

In order to test the performance of our proposed method, experiments were done using ECG signals from MIT/BIH ECG arrhythmia database. The details of the experiments are summarized below.

A. Data Preprocessing

The ECG signals come from 48 records of the original MIT/BIH ECG arrhythmia database. The original signal consists of two channels. The data from channel 1 were used in this study. The sampling frequency of the data was

360Hz. A rectangular beat window is formed so that a single ECG beat is contained in the window and the position of the R-peak in the QRS complex is located at the centre of window. The size of the window is chosen as 180 samples corresponding to 0.5 second, which is adequate to capture most of the information from a particular cardiac cycle. Prior to feature extraction, the ECG signals were preprocessed to remove noise and baseline shift due to power line interference, respiration and muscles tremors etc.

B. Feature Extractions

Many features have been proposed for the ECG beat recognition, such as QRS complex features, the statistical features, K-L expansion features, wavelet features etc. In this study, the ECG beat is down-sampled linearly to 30 samples/beat and a set of 30-D feature vectors are then formed. The original sample size can also be used. However, it will increase the computation complexity. We will show later that these features are not only easy to achieve and simple to calculated, but also very effective in the discrimination of abnormal beats from the normal ones. These feature vectors are normalized through dividing by the mean value of R peaks in each record of the database.

C. Training of the global and local experts

The ECG beats from the original 48 records of the MIT/BIH arrhythmia database are split into two parts. 1500 normal beats and 1500 abnormal beats from the first 24 records are used to train a global 2-SVC beat detector. Five-fold cross validation is used to select model parameters for 2-SVC. 5 ECG records were not used due to the lack of enough normal beats. The beats from each of the last 19 records are split into two sets, with the normal beats from the first 5 minutes of each record used to train a local ν-SVC detector and the remaining 25 minutes as the test set.

D. Evaluation Criteria and Experimental Results

Three criteria are used to evaluate the performance of ECG beat classification, including *sensitivity*, *specificity* and *classification rate*. Sensitivity is the fraction of abnormal ECG beats that are correctly detected among all the abnormal ECG beats. Specificity is the fraction of normal ECG beats that are correctly classified among all the normal ECG beats. Classification rate is the fraction of all correctly classified ECG beats - regardless of normal or abnormal among all the ECG beats. The experimental results of 19 records for the global 2-SVM classier, local ν-SVC classifier and the ensemble classifier are illustrated in table 1. In table 1, LD is the local beat detector – ν-SVC, GD is the global beat detector -2-SVC, MD1 is the max rule, and MD2 is the weighted averaging rule. Items in row Avg. are the average value of the 19 records.

IV. DISCUSSION

From Table 1, it is obvious that classification accuracy achieved using the proposed 2-SVM and ν-SVC based hierarchical learning approach is higher than both the global and local classifier. Both fusion strategies show good performance in improving the classification of the two basic classifiers, in which the weighted averaging performs slightly better than the max rule. Although the global SVM classifier can achieve more than 90% of the classification rate in training set, the average classification rate decreased to 84.7 % when it was used to test the ECG beats from different patients from the training set. The average classification rate of local ν-SVC classifier is higher than that of the global SVM classifier. This is due to its higher sensitivity – ability to recognize abnormal beats. Note that only 5 minutes of ECG records were used to train the local classifier. It shows that the patient-specific ECG beats are more informative which is important in improving the classification of the global SVM classifier.

The improvement of performance of the hierarchical learning approach is different among the 19 ECG records. In records 208, 213 and 220, the improvement in classification rate is significant, while it is less in some other records. If the performance of one classifier is very bad, the performance of the ensemble classifier will be influenced, but improvement can still be observed such as in record 212. This may be due to the fact that the decision boundaries of the two classifiers are different to each other and they can compensate each other to improve the classification. But the compensation of the two classifiers to each other is different for different training sets, which leads to different performance of the ensemble learning approach.

V. CONCLUSION

In this paper, a novel approach combining 2-class support vector machine and 1-class ν-support vector classifier is developed for ECG beat recognition. On one hand, 2-SVC is a powerful supervised learning tool, which is very suitable for acting as a global expert. On the other hand, ν-SVC is a robust and compact unsupervised learning machine, which can be trained using information from one class only and therefore is suitable to act as a local expert to relieve the medical experts from labeling the ECG beats for specific patients. The good performance of this approach is supported by experimental results on MIT/BIH arrhythmia ECG database and shows potential real-time application in intensive care units or telemedicine networks.

ACKNOWLEDGMENT

This study is supported by D2H2 project of Singapore-University of Washington Alliance.

REFERENCES

[1] S. Osowski and L.T. Hoai, "ECG beat recognition using fuzzy hybrid neural network," *IEEE Transactions on Biomedical Engineering*, Volume 48, No. 11, Pages:1265 – 1271, Nov. 2001.

[2] S. Osowski and L.T. Hoai, "Support vector machine-based expert system for reliable heartbeat recognition", *IEEE Transactions on Biomedical Engineering*, Volume: 51, No. 4, Pages: 582 – 589, April 2004.

[3] M.I. Owis, A.H. Abou-Zied, A.-B.M Youssef and Y.M. Kadah, "Study of features based on nonlinear dynamical modeling in ECG arrhythmia detection and classification," *IEEE Transactions on Biomedical Engineering*, Volume: 49, No. 7, Pages: 733 - 736, July, 2002.

[4] Z. Dokur and T. Olmez, "ECG beat classification by a novel hybrid neural network," *Comput Methods Programs Biomed.*, Vol. 66(2-3), pp.167-81, Sep, 2001.

[5] H.H. Yu, S. Palreddy and W.J. Tompkins, "A patient-adaptable ECG beat classifier using a mixture of experts approach", *IEEE Transactions on Biomedical Engineering*, Volume: 44, No.9, Pages: 891 – 900, Sept. 1997.

[6] N. Cristianini and J. Shawe-Taylor, "An Introduction to Support Vector Machines", Cambridge University Press, 2000.

[7] J. Millet-Roig and R. Ventura-Galiano, F. J. Chorro-Gasco and A. Cebrian, "Support Vector Machine for Arrhythmia Discrimination with Wavelet-Transform-Based Feature Selection", *Computers in Cardiology*, pp. 407 -410, 2000.

[8] B. Scholkopf, J. C. Platt, J. S. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the Support of a High-Dimensional Distribution", *Neural Computation*, Vol. 13, No. 7, pp. 1443-1471, July, 2001.

[9] J. Kittler, M. Hatef, R.P.W. Duin and J. Matas, "On Combining Classifiers", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 3, pp. 226-239, March, 1998.

[10] R. Mark and G. Moody, "MIT/BIH arrhythmia database directory," Massachusetts Inst. Technol. (MIT), 1988.

TABLE I

COMPARISON OF PERFORMANCE OF THE GLOBAL SVC, LOCAL ν-SVC AND HIERARCHICAL SYSTEM. ALL ITEMS ARE IN PERCENT (%).

Record	# of Test Data		Classification Rate				Specificity				Sensitivity			
(#)	Abn-ormal	Normal	LD	GD	MD1	MD2	LD	GD	MD1	MD2	LD	GD	MD1	MD2
201	410	1285	89.8	94.5	96.4	96.4	87.5	97.3	97.0	97.0	96.8	85.9	94.4	94.4
202	78	1706	29.8	89.6	43.0	43.2	26.6	91.1	40.4	40.6	100.0	56.4	100.0	100.0
203	487	2098	94.6	39.3	92.5	94.4	95.4	25.9	91.7	94.8	91.2	97.1	95.7	92.6
205	99	2123	84.8	99.5	97.6	97.6	84.1	99.9	97.5	97.5	100.0	91.9	100.0	100.0
208	1208	1321	94.8	94.1	95.9	95.9	91.4	96.7	94.5	94.5	98.5	91.2	97.4	97.4
209	413	2125	85.6	85.3	84.9	84.9	99.1	99.8	99.3	99.3	16.5	10.4	10.7	10.7
210	222	2011	92.9	81.5	93.6	93.5	92.4	83.1	93.1	93.0	97.3	66.7	97.3	97.3
212	1507	791	97.1	39.6	97.8	97.8	92.3	99.7	97.2	97.3	99.6	08.0	98.1	98.1
213	530	2211	90.6	92.3	95.9	95.9	90.7	100.0	98.5	98.6	90.4	60.0	84.9	84.5
215	161	2668	99.4	98.2	99.5	99.5	99.4	98.4	99.6	99.6	99.4	96.3	99.4	99.4
219	212	1710	78.5	71.8	83.6	82.2	76.1	68.8	81.9	80.4	97.6	96.2	96.7	96.7
220	113	1606	87.8	94.6	97.6	97.6	87.1	100.0	99.2	99.2	97.3	17.7	74.3	74.3
221	348	1699	96.3	99.3	98.5	98.5	95.5	99.2	98.2	98.2	100.0	99.7	100.0	100.0
222	539	1628	55.7	59.4	56.2	56.2	47.6	52.8	48.1	48.1	80.1	79.4	80.7	80.7
223	575	1623	74.7	88.0	83.7	83.5	68.0	93.7	80.7	80.3	93.7	72.2	92.0	92.3
228	362	1418	96.8	93.7	97.2	97.1	98.9	92.5	98.9	98.9	88.4	98.1	90.6	90.3
230	177	1874	95.9	98.3	98.0	98.0	95.5	98.2	97.8	97.8	100.0	100.0	100.0	100.0
233	758	1864	98.3	89.2	98.8	98.8	97.9	91.5	98.6	98.6	99.3	83.6	99.3	99.3
234	63	2236	95.1	97.8	97.9	97.9	95.9	100.0	99.0	99.0	66.7	22.2	57.1	57.1
Avg.	-	-	86.5	84.7	90.0	90.1	85.8	88.6	87.2	90.2	89.0	66.8	77.9	85.7