

Chapter V

A Concept Learning-Based Patient-Adaptable Abnormal ECG Beat Detector for Long-Term Monitoring of Heart Patients

Peng Li, Nanyang Technological University, Singapore

Kap Luk Chan, Nanyang Technological University, Singapore

Sheng Fu, Nanyang Technological University, Singapore

Shankar M. Krishnan, Nanyang Technological University, Singapore

Abstract

In this chapter, a new concept learning-based approach is presented for abnormal ECG beat detection to facilitate long-term monitoring of heart patients. The novelty in our approach is the use of complementary concept—“normal” for the learning task. The concept “normal” can be learned by a v -support vector classifier (v -SVC) using only normal ECG beats from a

specific patient to relieve the doctors from annotating the training data beat by beat to train a classifier. The learned model can then be used to detect abnormal beats in the long-term ECG recording of the same patient. We have compared with other methods, including multilayer feedforward neural networks, binary support vector machines, and so forth. Experimental results on MIT/BIH arrhythmia ECG database demonstrate that such a patient-adaptable concept learning model outperforms these classifiers even though they are trained using tens of thousands of ECG beats from a large group of patients.

Introduction

Electrocardiogram (ECG) is a recording of the electric potential variation due to the cardiac activities, which is often used by doctors to obtain reliable information about the performance of the heart function. The analysis of heart beat cycles in ECG signal is very important for long-term monitoring of heart patients. However, it is very costly for the medical expert to analyze the ECG recording beat by beat since the ECG records may last for hours. Therefore, it is justified to develop a computer-assisted technique to examine and annotate the ECG recording to facilitate review by medical experts. This computer annotation will assist doctors to select only the abnormal beats for further analysis.

Annotation of ECG recording requires the detection of various types of heart beats. This is a pattern recognition task. Very often, a classifier is to be trained to recognize different types of beats. The training set of the classifier, such as a multilayer neural network, is usually a large database which consists of the ECG beats from a large pool of patients. However, these classifiers suffer from the problem of poor generalization because there are usually some variations in the “normal” range among human beings. Even doctors may experience difficulty in assessing abnormal ECG beats if only considering the reference values based on the general patient population. There is a need to incorporate local information of a specific patient to improve the recognition of abnormal ECG beats and thus help to improve the generalization.

In this chapter, we proposed to approach this problem based on kernel concept learning. The benefit of concept learning is that only the information from one class is needed for training a classifier (a class density estimator to be precise). A concept learning model, called one-class support vector machine (SVM), is investigated to learn the concept “normal” and then used to detect the normal beats in ECG recording; hence, the abnormal beats are revealed as the complementary of the normal beats. Such model can relieve the doctors from

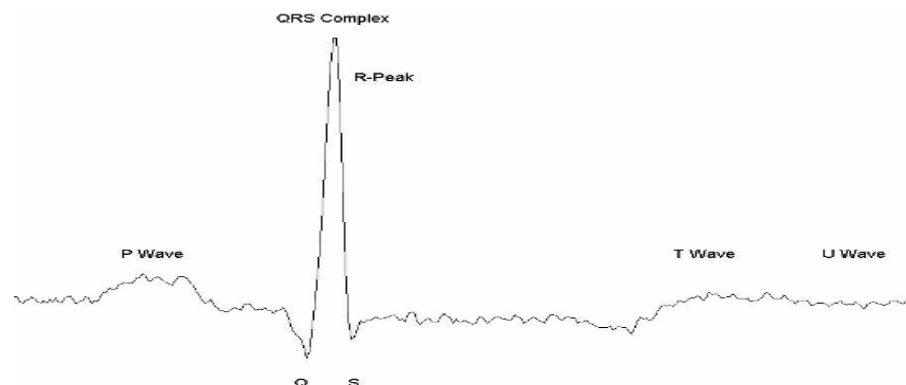
annotating the ECG beats one by one for training a binary classifier such as neural networks and binary support vector machines. The proposed method is compared to the generally used methods such as multilayer perceptron, using MIT/BIH arrhythmia ECG database (Massachusetts Institute of Technology, 1997).

Background

Electrocardiogram (ECG) is a recording of the heart's electrical currents obtained with the electrocardiograph, an instrument designed for recording the electrical currents that traverse the heart and initiate its contraction (Yanowitz, 2003). A typical ECG signal is illustrated in Figure 1 where different segments of the ECG signal characterize different cardiac activities. For example, P wave indicates the sequential activation (depolarization) of the left and right atria, ST-T wave signifies ventricular repolarization, QRS complex reveals the left and right ventricular depolarization, and so forth (Yanowitz, 2003). The main objectives of ECG monitoring include:

1. Detecting arrhythmias that occur intermittently or during certain physical activities.
2. Evaluating symptoms (such as chest pain, dizziness, or fainting) of possible heart disease.

Figure 1. A typical ECG beat



3. Detecting poor blood flow to heart muscle (ischemia), which may indicate coronary artery disease.
4. Monitoring the effectiveness of treatment (such as medication or a pacemaker or automatic defibrillator) for irregular heart rhythms.

Since ECG signal provides reliable information about the performance of the heart, the analysis of heart beat cycles in ECG signal is very important for long-term monitoring and diagnosis of patients' heart conditions in an intensive care unit or at patients' homes through a telemedicine network. However, it is very costly for the doctor to analyze the ECG records beat by beat since the ECG records may last for hours. Therefore, it is meaningful to develop some computer-assisted techniques to examine the ECG records and select only the abnormal beats for further analysis.

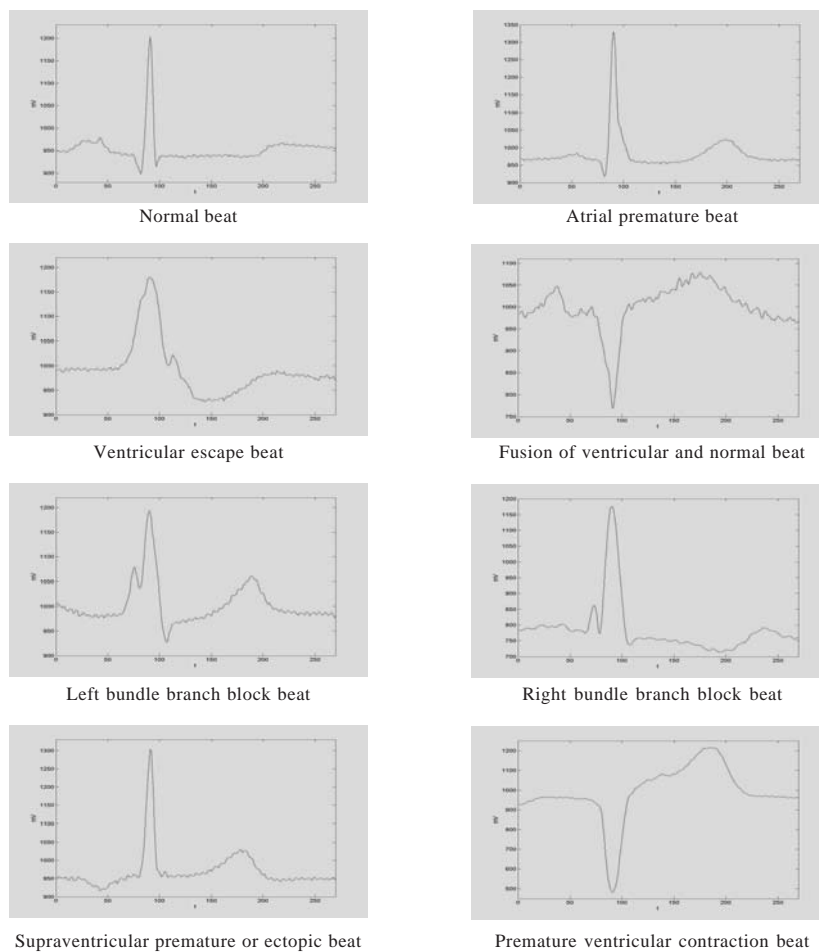
The problem at hand can be considered as a machine learning or data mining problem. Many algorithms have been applied to ECG beat cycle analysis, among which, neural networks are the most generally used model. Kohonen self-organizing maps were investigated in ECG beat recognition (Baig et al., 2001; Hu et al., 1997). Learning vector quantization was employed in Hu et al. (1997) and Baig et al. (2001). Stamkopoulos et al. (1998) proposed to detect ischemia using non-linear PCA neural networks. Guler and Ubeyh (2005) developed a combined multilayer perceptron neural network model for ECG beat classification. A survey of ECG pattern recognition based on nonlinear transformations and neural networks can be found in Maglaveras et al. (1998). There were also some attempts to incorporate fuzzy logic into the neural networks for ECG analysis (Engin, 2004; Osowski & Linh, 2001; Shyu et al., 2004). Recently, support vector machines, a new method emerging from the neural network research community, were introduced for ECG signal recognition (Millet-Roig et al., 2000; Osowski et al., 2004; Strauss et al., 2001).

One of the challenges faced by these ECG beat recognition algorithms is the large variation in the morphologies of ECG signals from different patients. The range of "normal beats" is different among the patients, which causes a so-called poor generalization problem; that is, a finely tuned ECG detector to the training data from a group of people may perform badly to the ECG beats of a new patient. Hu et al. (1997) attempted to solve this problem using a mixture of expert approaches. Such a mixture of expert structures 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 systems was used to classify the ECG beats from the specific patient, the classification performance was improved compared to that based on 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 records of each patient have 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 an annotation process is very costly and discourages the practical application of this approach. Another problem lies in the unbalanced data problem. In the scenario of long-term monitoring of heart patients, normal ECG beats usually dominate the ECG records. It takes a long time to collect sufficient and balanced normal and abnormal ECG data to construct a good local expert; otherwise, the local expert may suffer from the unbalanced data problem (Japkowicz & Stephen, 2002).

In this chapter, a kernel concept learning-based approach is proposed to solve such a generalization problem. One-class support vector classifier (ν -SVC) (Scholkopf et al., 2001) is a concept learning model whose goal is to find a decision boundary to include patterns from one class (called targets) and exclude the patterns from the other classes (called outliers). A particular benefit of ν -SVC is that it can be trained using only the data from one class. In the scenario of long-term monitoring of heart patients, the normal ECG beats usually dominate the ECG records; that is, the number of abnormal ECG beats is far less than that of the normal ones. 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, left bundle branch block beats, right bundle branch block beats, supraventricular premature or ectopic beats, premature ventricular contraction beats, and so on. Some of the typical abnormal beats and a normal ECG beat are illustrated in Figure 2. On one hand, these abnormal ECG beats appear differently 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 learn using few samples. Since normal ECG beats can be easily obtained from the patients, a ν -SVC can be trained using only the normal ECG beats from each specific patient to learn the concept "normal." The trained ν -SVC can then be used to detect from ECG records of the same patient and find "normal" ECG beats and thus detect the complementary abnormal beats. Such a kernel concept learning model can relieve the doctors from annotating the ECG beats from each patient beat by beat. Our experimental results using MIT/BIH arrhythmia ECG database (Massachusetts Institute of Technology, 1997) show that the patient-adaptable ν -SVC, constructed using only hundreds of normal ECG beats from a specific patient, outperforms all the other classifiers trained using tens of thousands of data from a group of patients in detecting the abnormal ECG beats of the specific patient. This suggests that our approach has good potential for practical clinical application.

Figure 2. Eight samples of normal and abnormal ECG beats using data from MIT/BIH arrhythmia database (Massachusetts Institute of Technology, 1997)

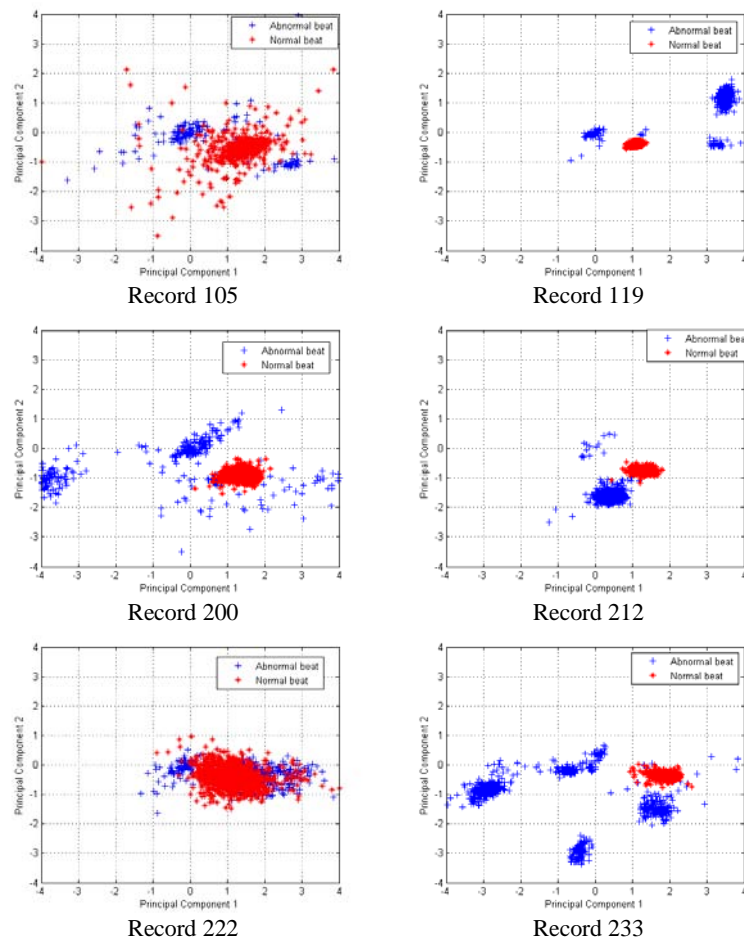


Generalization Problem in ECG Beat Detection

A fundamental assumption in the field of pattern recognition is that the distribution of the training samples is the same as that of the test samples (Duda et al., 2001). However, such an assumption may not hold in practical applications. The abnormal ECG beat detection problem is one of such examples. Figure 3 illustrates the distribution of the first two principal components of the original 39

dimensional feature vectors of ECG beats obtained by using Karhunen-Loeve transform (principal component analysis—PCA) from 4 records of MIT/BIH arrhythmia ECG database, where the red asterisks indicate normal ECG beats, and the blue crosses are abnormal ones. Although some discriminative information may be lost using PCA, it can be observed that the distribution of “normal” ECG beats are different in each patient. We may even plot such data distribution of the whole database, but the difference of the ECG beats among patients still exists. This is the difference between the population and a specific patient. Although an ECG detector can be finely trained using the ECG beats from a large database which consists of the records of different patients, it may perform badly in detecting ECG beats of other patients not in the database. This is the so-called generalization problem.

Figure 3. Scatterplot of ECG data of six patients in MIT/BIH arrhythmia database showing the first two principal components of PCA projection



The solution of such generalization problem lies in the incorporation of local information of a specific patient to the ECG detector. Since the distribution of the training samples is not the same as that of the test samples, some information about the test samples has to be added to train the ECG detector properly. It is infeasible to ask the doctors to annotate the ECG beats directly from the specific patient to be used for training the classifier due to the high cost of the process (de Chazal et al., 2004; Hu et al., 1997), and also the unbalanced data problem usually exists in the training data. Therefore, a concept learning-based approach is proposed to construct a patient-adaptable abnormal ECG detector.

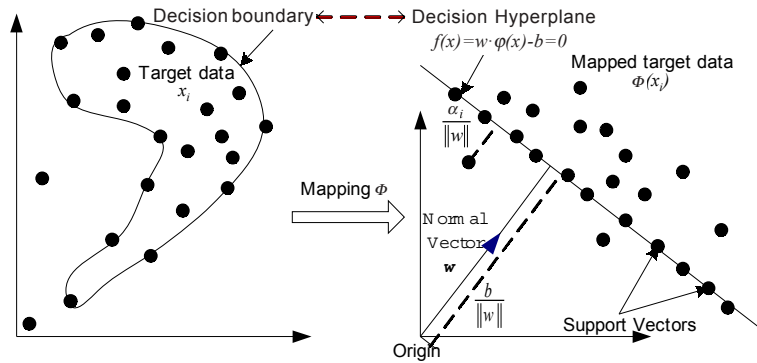
Concept Learning

Concept learning is also called one-class classification or novelty detection (Tax, 2001). The goal of concept learning is to find a descriptive model for a set of data. Different from classical binary classification, in concept learning, only data from one class (called the target) are used in the training stage, while no information is used about the other class (called the outliers). The philosophy behind concept learning is in agreement with the way human beings learn a concept. Suppose we expect to teach a child the concept of “tiger.” We need to give him or her some examples of tigers and do not need to give the examples of non-“tiger,” such as horse, elephant, or chicken. That is to say, people can learn a concept using only the examples of the target. Of course, the information about non-target or outliers is helpful to improve the discrimination between the target and non-target classes. However, using the examples from only the target class is sufficient to learn the concept of the target and recognize whether a new example belongs to the target or not.

One-Class Support Vector Classifier

One-class support vector classifier (ν -SVC) is a kind of support vector machine (Scholkopf et al., 2001) which can be used as a concept learning tool. Given a set of target data $X = \{\mathbf{x}_i \in R^d \mid i = 1, 2, \dots, N\}$, the goal of ν -SVC 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 $f(x) < 0$. It might be difficult to find such a decision function directly in the original space; therefore, the target data are mapped into a higher dimensional space called feature space $\phi(x)$ (illustrated in Figure 4) in which the dot product can be computed using some kernel function

Figure 4. Illustration of the kernel mapping of *v*-SVC



$$k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \tag{1}$$

such as a Gaussian radial basis function (RBF) kernel

$$k(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{\sigma}} \tag{2}$$

where $\sigma > 0$ is the width parameter of the Gaussian.

The mapped target data are separated from the origin with maximum margin using a hyperplane where the origin corresponds to the outliers. The hyperplane can be found by solving the following problem

$$\min_{w, \alpha_i, b} \frac{\|w\|^2}{2} + \frac{1}{vN} \sum_{i=1}^N \alpha_i - b \tag{3}$$

subject to

$$w \cdot \phi(x_i) - b + \alpha_i \geq 0, \alpha_i \geq 0, i = 1, 2, \dots, N \tag{4}$$

where α_i is a slack variable, and $v \in (0, 1]$ is a regularization parameter to control the effect of outliers and allows for target samples falling outside of the decision boundary (see Figure 4). The decision function corresponding to the hyperplane is

$$f(x) = w \cdot \phi(x) - b \quad (5)$$

where w is a weight vector and b is a bias item, similar to those of the neural networks.

The function in (3) is called the objective function and in (4) are called inequality constraints. (3) and (4) form a constrained optimization problem, which is usually dealt with by introducing Lagrange multipliers $\beta_i \geq 0$, $\gamma_i \geq 0$, and a Lagrangian (Amari, 1998) for this can be written as

$$L(w, \alpha_i, b, \beta_i, \gamma_i) = \frac{\|w\|^2}{2} + \frac{1}{vN} \sum_{i=1}^N \alpha_i - b - \sum_{i=1}^N \gamma_i \alpha_i - \sum_{i=1}^N \beta_i (w \cdot \phi(x_i) - b + \alpha_i) \quad (6)$$

Setting the partial derivatives of Lagrangian with respect to w , α_i , and b to zero, the new constraints are

$$\sum_{i=1}^N \beta_i = 1 \quad (7)$$

$$\frac{1}{vN} - \beta_i - \gamma_i = 0 \quad (8)$$

$$w = \sum_{i=1}^N \beta_i \phi(x_i) \quad (9)$$

Substituting (7-9) to the Lagrangian (6) and using kernel function (1), the dual problem is

$$\max_{\beta} -\frac{1}{2} \sum_{i,j=1}^N \beta_i \beta_j k(x_i, x_j) \quad (10)$$

subject to

$$\sum_{i=1}^N \beta_i = 1 \quad (11)$$

$$0 \leq \beta_i = \frac{1}{vN} - \gamma_i \leq \frac{1}{vN} \quad (12)$$

This is a quadratic programming problem which can be solved using some standard algorithms such as sequential minimization optimization (Scholkopf et al., 2001).

The decision function in (5) can be reformulated as follows using (9) and kernel function (1)

$$f(x) = \sum_{i=1}^N \beta_i (\phi(x_i) \cdot \phi(x)) - b = \sum_{i=1}^N \beta_i k(x_i, x) - b \quad (13)$$

At the optimal point, a set of conditions have to be satisfied, known as the Karush-Kuhn-Tucker (KKT) optimal conditions (Kuhn & Tucker, 1951). Exploiting the KKT conditions, the following three cases of β_i s can be the result:

1. $\beta_i = \frac{1}{vN}$: $f(x_i) < 0$, the target data x_i is outside the decision hyperplane (incorrectly classified). Such a x_i is called bounded support vector (BSV).
2. $0 < \beta_i < \frac{1}{vN}$: $f(x_i) = 0$, the target data x_i is on the decision hyperplane. Such a x_i is called unbounded support vector (USV).
3. $\beta_i = 0$: $f(x_i) > 0$, the target data x_i is inside the decision hyperplane (correctly classified).

All of the target data whose $\beta_i > 0$ are called support vectors (SVs), including both BSVs and USVs. It can be observed from (13) that only support vectors contribute to the decision function. The number of support vectors is usually far less than that of the target samples. These support vectors can be regarded as a sparse representation or compact template of the whole target dataset. Given

a new pattern, it is compared with the support vectors in the decision function (13). If the new pattern is from the targets, the decision function has a large positive value. If the new pattern is from the outlier class, it is different from the support vectors, and the decision function has a large negative value. The larger the value of the decision function, the more confident the decision. From these, we can see the great similarity between the support vector machines and neural networks.

At the optimal point, the constraints in (4) become equalities. Both β_i and γ_i are in $(0, \frac{1}{vN})$. The bias b can be recovered using one of such a USV β_k corresponding to the target sample x_k

$$b = w \cdot \phi(x_k) = \sum_{i=1}^N \beta_i k(x_i, x_k) \quad (14)$$

Another kernel concept learning model is support vector data description (SVDD) (Tax & Duin, 1999). It is proved that when a RBF kernel (2) is used, SVDD is equivalent to ν -SVC (Scholkopf et al., 2001; Tax, 2001). Therefore, only ν -SVC is investigated in the current study.

Model Selection

Model selection for concept learning with ν -SVC is still an open problem because only the data from one class is used in the training stage and no outlier is available. Some attempts have been made to select the model parameters of concept learning with ν -SVC using only the information from the target data. For example, Tax proposed the use of consistency of the classifiers to select model parameters based on the error on the target class only (Tax & Muller, 2004). But this may be biased since no information of the outliers is used. Another solution is generating artificial outliers, uniformly distributed in a hypersphere or hypercube, which are used to estimate the outlier error (Tax & Duin, 2002). The latter is investigated in the current problem.

There are two hyperparameters to be tuned in ν -SVC using RBF kernel, the width parameter σ in (2) and the regularization parameter ν in (3). Let N_{SV} be the number of support vectors. Since the constraint in (11) is $\sum_{i=1}^N \beta_i = 1$, and the upper bound of β_i in (12) is $\frac{1}{vN}$, the following equation can be written

$$N_{SV} \frac{1}{vN} \leq 1 \tag{15}$$

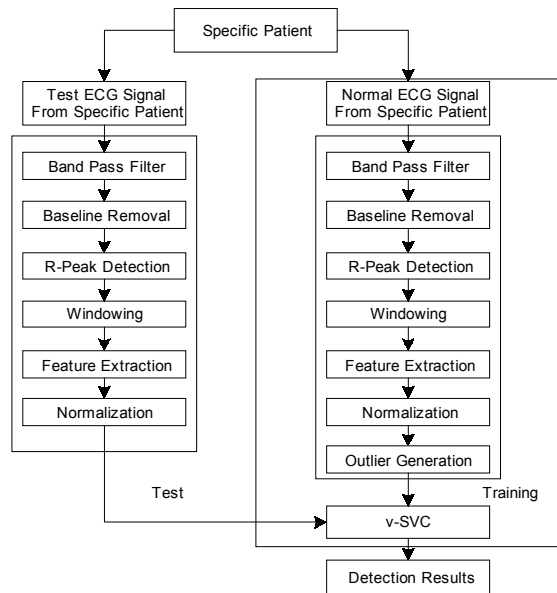
thus

$$v \geq \frac{N_{SV}}{N} \tag{16}$$

Obviously, v is the upper bound of the fraction of support vectors among all the target data. This observation can help to select the value of v . The larger the value of v , the more target samples will be excluded from the decision boundary, and this leads to larger error of the target class. Therefore, the value of v cannot be very large. Here it is fixed as 0.01, which means the training error of the target class cannot be larger than 1%.

The value of σ can be selected using an artificial validation set (Tax & Duin, 2002). Given a set of target samples, some outlier samples are generated randomly with the assumption that the outliers are uniformly distributed around the target class. The union of targets and generated outliers is used as a validation

Figure 5. Flowchart of the proposed framework for abnormal ECG beat detection



set. The value of σ is chosen so that the error of both target class and outlier class on the validation set is minimized.

The selection of σ depends on the distribution of the generated outliers. From Figure 3, it is observed that the normal ECG data are usually in the center of the feature space, and the abnormal ones are distributed around the normal data. Therefore, the assumption of uniformly distributed outliers approximately holds, which is demonstrated in the experimental results.

Proposed Framework

Figure 5 illustrates the flowchart of the proposed framework for ECG beat detection. The details are as follows.

ECG Signal Preprocessing

The ECG signal is usually coupled with noise and baseline shift due to power line interference, respiration and muscle tremors, and so forth, which have to be removed in favor of further analysis. The ECG signal is first processed using two averaging filters (Christov, 2004):

1. *Electromyogram noise suppression*: averages samples in an interval of 28 ms.
2. *Power line interference suppression*: averages samples in a period of the interference frequency of the power line.

The baseline of the ECG signal can be obtained using two consecutive median filters to the ECG signal after noise suppression whose widths are 200ms and 600ms, respectively (de Chazal et al., 2004). The baseline is subtracted from the original signal, and the resulted signal is then baseline-corrected.

After noise suppression and baseline correction, the R-peak of the ECG signal can be detected using the first derivative of the ECG signal (Christov, 2004), which is used in the following process.

Feature Extraction

Many features have been proposed for ECG beat recognition, such as time domain representation (Hu et al., 1997), heartbeat interval features (de Chazal

et al., 2004), Hermite functions (Gopalakrishnan et al., 2004), autoregressive modeling-based representation (Zhang et al., 2004), and wavelet transform-based representation (Engin, 2004; Millet-Roig et al., 2000; Shyu et al., 2004). Here we investigate the use of raw amplitude of the time domain ECG signals after noise suppression and baseline shift removal as feature vectors to represent the ECG beats. After the R-peak is found, the ECG signal in a window of 550 ms is taken as an ECG beat. The lengths of the signal before and after the R-peak in each beat are 140 ms and 410 ms, respectively, such that the window covers most of the characterization of the ECG beat. The signal in each window is then down-sampled uniformly to form a feature vector of 38-dimensions. It has been shown that R-R interval (the interval between two consecutive R-peaks) (de Chazal et al., 2004; Hu et al., 1997) is useful in recognition of some abnormal ECG beats. Therefore, it is also used 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.

There are some variations in the amplitude ranges of ECG signals among human beings, which imply that a normalization procedure is necessary to the ECG feature vectors. The feature vectors are then divided by the mean value of R peaks in the training data of each patient, such that the maximum amplitude in each ECG beat window is around 1. The normalized ECG feature vectors are then used.

Learning the Concept Normal for Abnormal ECG Beat Detection

In order to detect abnormal beats from the long-term ECG records of a patient, an abnormal ECG beat detector can be constructed using ν -SVC based on a short period of the normal ECG beats from the same patient such that the model can learn the concept of “normal” beat of the patient. A new ECG beat can be classified to the “normal” class or non-“normal” class by the trained ν -SVC. The abnormal ECG beats are thus annotated automatically for the doctors for further review.

Experimental Results and Discussions

An experiment is conducted using MIT/BIH arrhythmia ECG database to demonstrate the feasibility of the proposed patient-adaptable concept learning approach. The details of the experiment are as follows.

Data Preparation

MIT/BIH arrhythmia ECG database consists of 48 annotated records from 47 patients, and each record is about 30 minutes in length. The labels in the annotation file made by cardiologists are used as the ground truth in training and evaluating the classifiers. The ECG beats labeled as “normal” (NOR) are taken as the target class whose number in the database is more than 70,000. All of the other beats are regarded as outlier class or “abnormal” class, including atrial premature beats, nodal premature beats, ventricular escape beats, fusion of ventricular and normal beats, and so forth, which totals around 30,000.

Table 1. Classification performance of the classifiers in each recording of the test set

Record #	Number of Beats			v-SVC				LSVC			
	Abnormal	Normal	Total	TP	FN	FP	TN	TP	FN	FP	TN
100	30	1862	1892	24	6	75	1787	28	2	440	1422
103	8	1731	1739	8	0	484	1247	6	2	17	1714
105	141	2098	2239	141	0	261	1837	130	11	1362	736
113	5	1488	1493	5	0	20	1468	5	0	47	1441
117	4	1275	1279	4	0	442	833	3	1	267	1008
119	448	1294	1742	448	0	52	1242	333	115	37	1257
121	12	1548	1560	11	1	202	1346	12	0	824	724
123	3	1259	1262	3	0	60	1199	3	0	2	1257
200	891	1432	2323	854	37	25	1407	828	63	69	1363
202	78	1707	1785	78	0	1399	308	75	3	807	900
210	223	2011	2234	218	5	127	1884	206	17	355	1656
212	1507	792	2299	1506	1	47	745	1396	111	64	728
213	530	2212	2742	430	100	144	2068	237	293	3	2209
215	161	2669	2830	160	1	4	2665	95	66	6	2663
219	212	1711	1923	205	7	129	1582	186	26	79	1632
221	348	1700	2048	348	0	70	1630	345	3	176	1524
222	539	1629	2168	429	110	752	877	207	332	649	980
228	362	1419	1781	314	48	0	1419	329	33	83	1336
230	177	1875	2052	177	0	18	1857	174	3	21	1854
231	1370	302	1672	1370	0	24	278	770	600	0	302
233	758	1865	2623	752	6	39	1826	683	75	115	1750
234	63	2237	2300	62	1	42	2195	39	24	9	2228
Total	7870	36116	43986	7547	323	4416	31700	6090	1780	5432	30684

Four records (102, 104, 107, and 217), including paced beats, are excluded from the study in compliance with the standards recommended for reporting performance results of cardiac rhythms by the Association for the Advancement of Medical Instrumentation (AAMI) (Mark & Wallen, 1987). Among the remaining 44 records, 22 records are used to train classical binary classifiers, including SVMs and other neural networks for comparison with the concept learning method. They are records 101, 106, 108, 109, 111, 112, 114, 115, 116, 118, 122, 124, 201, 203, 205, 207, 208, 209, 214, 220, 223, and 230. This dataset is called DB1. The other 22 records are split into two parts. The normal ECG beats in the first one sixth of each of the 22 records (Table 1) (each about 300 beats) are used as the training set to construct the ν -SVCs. These data form DB2. The remaining five sixth of each of the 22 records is used as a test set to evaluate the performance of the ν -SVCs and the binary classifiers trained on the other 22 ECG records. Such a dataset is called DB3. The 22 records in Table 1 are selected so that most of the ECG beats are normal, which is appropriate to simulate the real ECG records captured in a long-term monitoring scenario.

The original signals in the MIT/BIH arrhythmia database are two-leads, sampled at 360 Hz. The ECG signal of Lead 1 is used in this study. The ECG signals are processed following the procedure described in the previous section. Each ECG beat is represented by a 39-D feature vector.

Evaluating Criteria

Table 2 illustrates the full classification matrix for calculating the evaluating criteria, including true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The criteria used to evaluate the performance of ECG beat classification include 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.

$$SEN = \frac{TP}{TP+FN} \quad (17)$$

Table 2. Full classification matrix for calculating the evaluation measure

Ground Truth	Classification Results	
	Abnormal	Normal
Abnormal	True Positive (TP)	False Negative (FN)
Normal	False Positive (FP)	Truth Negative (TN)

Specificity (SPE) is the fraction of normal ECG beats that are correctly classified among all the normal ECG beats.

$$SPE = \frac{TN}{TN+FP} \quad (18)$$

The generally used average classification rate (ACR) is the fraction of all correctly classified ECG beats, regardless of normal or abnormal among all the ECG beats.

$$ACR = \frac{TP+TN}{TP+TN+FP+FN} \quad (19)$$

As mentioned in the previous section, the “normal” class dominates the test set. Commonly used average classification rate is not valid in such an imbalanced dataset. For example, if a classifier is trained to classify all the test data as normal beats, it has SPE = 100% and SEN = 0%, but the ACR is still high because the number of abnormal beats is too small. Therefore, another evaluation measure is used in this study, balanced classification rate (BCR). BCR is the average value of SEN and SPE.

$$BCR = \frac{SEN+SPE}{2} \quad (20)$$

Only when both SEN and SPE have large values can BCR have a large value. Therefore, the use of BCR can have a balanced performance in the evaluation of the classifiers which favor both lower false positives and false negatives. This measure is more suitable for the current study than ACR.

Training Global Binary Classifiers for Comparison

A set of commonly used binary classifiers are trained using 22 ECG records outside the test records to compare with the proposed concept learning method. The training set DB1 consists of 31,069 “normal” beats and 19,661 “abnormal” beats. There are some classifiers that have problems with training using such a large database; therefore, a smaller subset of these training sets is formed by randomly selecting 4,000 “normal” beats and 4,000 “abnormal” beats, called DB11. A Matlab toolbox called PRTOOLS (Duin et al., 2004) is used in our

experiment to construct most of these classifiers otherwise mentioned. The classifiers include:

1. *Linear Bayes normal classifier (LDC)*: The linear classifier between the two classes by assuming normal densities with equal covariance matrices.
2. *Quadratic Bayes normal classifier (QDC)*: The quadratic classifier between the two classes by assuming normal densities.
3. *Nearest mean classifier (NMC)*: The nearest mean classifier between the two classes. The test pattern is classified so the class whose mean value is closer to the test pattern in the feature space.
4. *Backpropagation feedforward neural network classifier (BNN)*: A feedforward neural network classifier with m units in a hidden layer, the training is stopped after l epochs. The hyperparameters m and l are optimized using 5-fold cross validation on the training set. The optimal values in DB1 are $m = 44$ and $l = 2000$, respectively.
5. *Binary support vector machines*: Including a linear SVM (LSVC) and a SVM with RBF kernel (RSVC). LIBSVM is used in this study (Chang & Lin, 2001). The hyperparameters are similar to ν -SVC, which are optimized using 5-fold cross validation on the training set. SVMs have problem to deal with large training dataset (Cristianini & Shawe-Taylor, 2000). So the compact subset DB11 (8,000 ECG beats) is used for training SVMs. The hyperparameters are optimized using 5-fold cross validation on the training set DB11. The optimal values are regularization parameter $C = 4096$ for LSVC and $C = 16$, $\sigma = 0.7$ for RSVC.

The classification results of using these binary classifiers trained on the large dataset and the ν -SVC trained using only “normal” ECG beats from the specific patient are illustrated in Table 3. The results are averaged over 22 test ECG records. In accordance with AAMI recommendations to present the results

Table 3. Comparison of abnormal ECG beat detection using ν -SVC trained with only normal data from the specific patient and the binary classifiers trained with large database excluding the specific patient

Classifier	ν -SVC	LSVC	RSVC	LDC	QDC	NMC	BNN
BCR	0.917	0.840	0.814	0.833	0.773	0.730	0.838
SPE	0.958	0.823	0.819	0.822	0.696	0.494	0.766
SEN	0.876	0.856	0.808	0.844	0.850	0.965	0.909
ACR	0.883	0.831	0.803	0.828	0.845	0.883	0.851

tape-by-tape, the classification results of ν -SVC and the best binary classifier LSVC of each record are shown in Table 1.

Discussions

From Table 3, it can be observed that the patient-adaptable concept learning model, ν -SVC trained on DB2, outperforms all the binary classifiers trained using large database DB1 excluding the specific patient to be tested. The best binary classifier is LSVC and BNN whose BCRs are about 84%. The BCR of ν -SVC is about 92%, which is much greater than all of the binary classifiers. It indicates that the local information is very important in the classification of the ECG beats from a specific patient. The incorporation of such local information can help deal with the gap between the distribution of training dataset and test dataset, which helps to improve the generalization.

Another observation is that it seems that the linear classifier shows better performance than those of the non-linear classifier. RBF SVM is usually superior to Linear SVM in the classification. However, in Table 3, the linear SVM outperforms RBF SVM. This is caused by the difference of training dataset and test dataset. RBF SVM can produce a more flexible decision boundary than linear SVM. Since it is tuned to fit to the training set, it cannot generalize well in the test set due to the difference between the two datasets. On the other hand, the linear SVM shows more robustness although it produces larger error in the training set. Therefore, the proposed concept learning method is suitable to solve the problem at hand.

The performance of ν -SVC varies among the 22 test sets. The ν -SVC model performs well in most of these test sets. Only in some of them, it does not perform well. For example, the BCR in record 222 is only about 67%. Figure 3 illustrates the data distribution of this record. It shows that the “normal” ECG beats and “abnormal” beats overlap greatly in these two records. In this case, even the optimal Bayesian classifier may produce a large error rate. More efficient features are needed to further improve the classification between the two classes, which is our future work.

The classification results in Table 2 and Table 3 show that the features used are quite efficient in discriminating abnormal ECG beats from those of the normal ones. Compared to other features, such as heartbeat interval features (de Chazal et al., 2004), Hermite functions (Gopalakrishnan et al., 2004), autoregressive modeling-based representation (Zhang et al., 2004), and wavelet transform-based representation (Engin, 2004; Millet-Roig et al., 2000; Shyu et al., 2004), the currently used features are simpler to implement because only R-peak of each

ECG beat needs to be detected to extract these features and no further detection such as QRS detection or other transform is necessary anymore. Therefore, its computation complexity is far less than those of the other features.

Hu et al. (1997) concentrated on the classification of 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 its BCR is about 90%. De Chazal et al. (2004) have claimed that they achieved comparable performance to the method of Hu's using a linear discriminant classifier. Our concept learning-based method achieved a balanced classification rate of 92% although only some "normal" ECG beats from each patient are used to train the ν -SVC model. Furthermore, the data records including the test data and the training data of ν -SVC are seriously unbalanced. Hu et al. (1997) and de Chazal et al. (2004)'s methods have problems in training a good classifier in such cases. Therefore, our proposed method shows better or at least comparable performance compared to Hu et al. (1997) and de Chazal et al. (2004). Another advantage of our method is that it can relieve the doctors from annotating the ECG beats one by one as needed in Hu et al. (1997), and it is easier to be constructed to adapt to the specific patient.

Conclusion

In this chapter, a new concept learning-based approach is proposed to detect abnormal ECG beats for long-term monitoring of heart patients. A kernel concept learning model, ν -SVC, can be trained to extract the concept "normal" using only some "normal" ECG beats from a patient, which can then be used to detect "abnormal" ECG beats in long-term ECG records of the same patient. Such an approach can relieve doctors from annotating the training ECG data beat by beat and also addresses the generalization problem in ECG signal classification among patients. Experiments were conducted using 44 ECG records of MIT/BIH arrhythmia database. The experimental results demonstrate the good performance of our proposed concept learning method and suggest its potential for practical clinical application.

Acknowledgment

The authors wish to acknowledge the support by Distributed Diagnosis and Home Healthcare project (D2H2) under Singapore-University of Washington

Alliance (SUWA) Program and Biomedical Engineering Research Center at Nanyang Technological University in Singapore. Also, the first author would like to express gratitude for the research scholarship from the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore.

References

- Amari, S. I. (1998). Natural gradient works efficiently in learning. *Neural Computation*, 10(2), 251-276.
- Baig, M. H., Rasool, A., & Bhatti, M. I. (2001, October). Classification of electrocardiogram using SOM, LVQ and beat detection methods in localization of cardiac arrhythmias. *Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2, 1684-1687.
- Chang, C. C., & Lin, C. J. (2001). *LIBSVM: A library for support vector machines*. Retrieved October 7, 2005, from <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- Christov, I. (2004). Real time electrocardiogram QRS detection using combined adaptive threshold. *BioMedical Engineering Online*, 3. Retrieved October 7, 2005, from <http://www.biomedical-engineering-online.com/content/3/1/28>
- Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines*. New York: Cambridge University Press.
- de Chazal, P., O'Dwyer, M., & Reilly, R. B. (2004). Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, 51, 1196-1206.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (2nd ed.). New York: John Wiley & Sons.
- Duin, R. P. W., Juszczak, P., Paclik, P., Pekalska E., de Ridder, D., & Tax, D. M. J. (2004). *PRTools4, A Matlab Toolbox for pattern recognition*. Delft University of Technology. Retrieved October 7, 2005, from <http://www.prtools.org/>
- Engin, M. (2004). ECG beat classification using neuro-fuzzy network. *Pattern Recognition Letters*, 25, 1715-1722.
- Gopalakrishnan, R., Acharya, S., & Mugler, D. H. (2004, April). Automated diagnosis of ischemic heart disease using dilated discrete Hermite functions. *Proceedings of the IEEE 30th Annual Northeast Bioengineering Conference* (pp. 97-98).

- Guler, I., & Ubeyh, E. D. (2005). ECG beat classifier designed by combined neural network model. *Pattern Recognition*, 38, 199-208.
- Hu, Y. H., Palreddy, S., & Tompkins, W.J. (1997). A patient-adaptable ECG beat classifier using a mixture of experts approach. *IEEE Transactions on Biomedical Engineering*, 44, 891-900.
- Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis*, 6, 429-450.
- Kuhn, H. W., & Tucker A. W. (1951). Nonlinear programming. *Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probabilistics*, 481-492
- Maglaveras, N., Stamkopoulos, T., Diamantaras, K., Pappas, C., & Strintzis. M. (1998). ECG pattern recognition and classification using non-linear transformations and neural networks: A review. *International Journal of Medical Informatics*, 52, 191-208.
- Mark, R., & Wallen, R. (1987). *AAMI-recommended practice: Testing and reporting performance results of ventricular arrhythmia detection algorithms*. Association for the Advancement of Medical Instrumentation, Arrhythmia Monitoring Subcommittee. AAMI ECAR-1987.
- Massachusetts Institute of Technology. (1997). *MIT-BIH Arrhythmia database*. Retrieved October 7, 2005, from <http://ecg.mit.edu/>
- Millet-Roig, J., Ventura-Galiano, R., Chorro-Gasco, F.J., & Cebrian, A. (2000). Support vector machine for arrhythmia discrimination with wavelet-transform-based feature selection. *Computers in Cardiology*, 28, 407-410.
- Oowski, S., Hoai, L. T., & Markiewicz, T. (2004). Support vector machine-based expert system for reliable heartbeat recognition. *IEEE Transactions on Biomedical Engineering*, 51, 582-589.
- Oowski, S., & Linh, T. H. (2001). ECG beat recognition using fuzzy hybrid neural network. *IEEE Transactions on Biomedical Engineering*, 48, 1265-1271.
- Scholkopf, B., Platt, J. C., Shawe-Taylor J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, 13, 1443-1471.
- Shyu, L. Y., Wu, Y. H., & Hu, W. (2004). Using wavelet transform and fuzzy neural network for VPC detection from the holter ECG. *IEEE Transactions on Biomedical Engineering*, 51, 1269-1273.
- Stamkopoulos, T., Diamantaras, K., Maglaveras, N., & Strintzis, M. (1998). ECG analysis using nonlinear PCA neural networks for ischemia detection. *IEEE Transactions on Signal Processing*, 46, 3058-3067.