ECG Analysis Using Multiple Instance Learning for Myocardial Infarction Detection

Li Sun, Yanping Lu*, *Member, IEEE*, Kaitao Yang, and Shaozi Li

***Abstract*—This paper presents a useful technique for totally automatic detection of myocardial infarction from patients' ECGs. Due to the large number of heartbeats constituting an ECG and the high cost of having all the heartbeats manually labeled, supervised learning techniques have achieved limited success in ECG classification. In this paper, we first discuss the rationale for applying multiple instance learning (MIL) to automated ECG classification and then propose a new MIL strategy called latent topic MIL, by which ECGs are mapped into a topic space defined by a number of topics identified over all the unlabeled training heartbeats and support vector machine is directly applied to the ECG-level topic vectors. Our experimental results on real ECG datasets from the PTB diagnostic database demonstrate that, compared with existing MIL and supervised learning algorithms, the proposed algorithm is able to automatically detect ECGs with myocardial ischemia without labeling any heartbeats. Moreover, it improves classification quality in terms of both sensitivity and specificity.**

***Index Terms*—Classification, ECG analysis, multiple instance learning (MIL), myocardial infarction (MI).**

I. Introduction

An ECG is a graphic trace of the bioelectrical activity generated by the cardiac or heart muscle during a heartbeat. It provides cardiologists with useful information about the rhythm and functioning of the heart. Trained cardiologists can successfully discover different groups of cardiac abnormalities in ECG recordings. However, a 24-h ECG, possibly recording over 100 000 heartbeats, is usually too large to be examined in its entirety by a cardiologist. In fact, diagnosis of certain cardiac diseases may take several hours when done by visual inspection, and even so, some vital information may be missed due to the tedious manual procedure. Cardiologists therefore need automated tools to accurately analyze the vast amount of ECG data collected by monitoring devices. Automated classification of long-term ECG recordings is becoming a universal need in clinical applications.

Early attempts to tackle this issue were dominated by supervised learning-based approaches [1]–[8], which aim at adapting standard supervised methods. Some studies [1]–[3] extracted features related to ECG morphology, ECG waveforms, such as RR intervals, QRS complex, ST segment, etc., and heartbeat intervals, and employed linear discriminant analysis, decision trees, or rough-set-based methods for ECG classification, respectively. These rule-based methods would produce too many decision rules due to the very large number of heartbeats. In [4]–[6], based on extracted features such as Hermite coefficients, ECG waveforms, and the pre/post RR interval ratio, etc., neural network (NN)-based methods such as fuzzy NN and block-based NN were constructed for automatic ECG classification. In [7] and [8], support vector machine (SVM)-based methods were employed to categorize ECGs into distinct groups, based on features extracted from ECG signals such as higher order statistics, Hermite polynomials, ECG morphology, QRS complex duration, RR interval, etc.

These supervised learning methods are able to process ECG recordings without any additional input from experts, but they do not yield good quality results because they incorrectly consider the label of an ECG to be the true labels of its heartbeats. Indeed, an ECG is comprised of multiple heartbeats, a few of which may be positive and most of which are negative for a certain heart disease. These supervised learning-based techniques only take the inter-ECG difference (the difference between true labels of the whole ECGs) into account, but do not consider the intra-ECG difference (the difference between true labels of heartbeats within an ECG). To address this problem, in the past few years, there has been a growing interest in developing semisupervised learning (SSL) strategies [9]–[11] for ECG classification. Typically, these SSL approaches select a representative subset of heartbeats from the training set, get cardiologists to manually label them, and finally train classifiers on them.

In [9] and [10], based on ECG morphology and heartbeat interval features, experts validate the first few heartbeats of each ECG and NNs are trained on the annotated heartbeats to produce patient-adaptive classification systems. Another study [11] presents a string-matching method to construct a patient-adaptive cardiac profiling scheme for heartbeat classification. To a certain extent, these strategies boost heartbeat classification performance by incorporating expert knowledge in the form of heartbeat labels. Furthermore, they help to alleviate the time and money cost entailed in manually labeling large numbers of training heartbeats. However, because they merely select a

Manuscript received January 9, 2012; revised January 16, 2012 and May 14, 2012; accepted August 1, 2012. Date of publication August 23, 2012; date of current version November 22, 2012. This work was supported by the Natural Science Foundation of Fujian Province under Grant 2010J01345, Natural Science Foundation of China under Grant 61202143, the Fundamental Research Funds for the Central Universities under Grant 2011121052, the Research Fund for the Doctoral Program of Higher Education of China under Grant 201101211120024, and Shenzhen Key Laboratory for High Performance Data Mining with Shenzhen New Industry Development Fund under Grant CXB201005250021A. *Asterisk indicates corresponding author.*

L. Sun, K. Yang, and S. Li are with the Department of Cognitive Science, Xiamen University, Fujian 361005, China (e-mail; 278902213@qq.com; kaitao_yang@xmu.edu.cn; szlig@xmu.edu.cn).

*Y. Lu is with the Department of Cognitive Science, Xiamen University, Fujian, China (e-mail: yanpinglv@xmu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2012.2213597

fixed heartbeat length to construct the respective training sets, the collection of heartbeats is not good enough to represent the data distribution.

To solve this problem, active learning methods [12], [13] have been proposed for heartbeat classification by selecting heartbeats that are more representative of the statistical distribution of the data, in an iterative way. Basically, they first select some representative heartbeats to be labeled and added to the initial training set, then employ SVM to train on the labeled heartbeats. The procedure is iterated until the stopping criterion is reached.

Different SSL models differ in what heartbeats are selected for labeling and how classifiers are updated. Although semisupervised strategies can adaptively consider the intra-ECG difference as well as reduce the number of heartbeats to be labeled in the training set by about 90%, their performances largely depend on the quality and quantity of the heartbeats selected. And even this selection is costly for the cardiologist, in terms of time and money, because the size of the training set is usually pretty large, especially for long-term ECGs: typically over 100 000 beats for a single patient.

In this paper, we propose a multiple instance learning (MIL) strategy which can be successfully applied to build a totally automatic ECG classification procedure. We design a new representation of the ECG to construct a distance measure that considers both intra- and inter-ECG differences. Our method takes full advantage of the ECG labels, while not requiring any heartbeat labels. A series of experiments on a real ECG dataset show that compared with several mainstream algorithms based on supervised learning and MIL, the proposed algorithm achieves good performance on recognition of ECGs related to myocardial infarction (MI), in terms of both sensitivity and specificity.

The remainder of this paper is organized as follows. Section II briefly describes the background and explains the rationale behind using MIL for automated ECG classification. Section III presents our proposed latent topic multiple instance learning (LTMIL) method. Experimental results on real ECG datasets are given in Section IV. Section V draws conclusions and gives directions for future work.

II. Background

A. MI

MI is one of the heart diseases associated with very high death rates in recent years, so an automatic and effective classification tool is needed to recognize and locate MI in ECGs. There are many groups of MI, such as anterior, anterolateral, anteroseptal, inferior, inferolateral, and so on. Different groups of MI may have quite similar characteristics on ECGs, making it difficult for automatic tools to distinguish them correctly.

In later sections, we will compare the capacity of each algorithm to recognize MI on the PTB diagnostic ECG database [14]. In the current literature related to ECG classification, most authors have conducted experiments on the MIT ECG database [15]. One of our reasons for choosing the PTB rather than the MIT database is that the PTB database has abundant and well-classified ECG recordings related to MI, and our main task in this paper is to detect MI from ECGs. More importantly, in the PTB database, each ECG signal is represented from 15 angles, i.e., 15 leads (though we only use the conventional 12) and with a higher sampling rate, which is much better than the 2 leads and 360-Hz sampling rate of the ECGs in the MIT database.

Specifically, the PTB diagnostic ECG database contains 448 records from 290 individuals. A total of 369 of the ECG records are related to MI, while the rest correspond to healthy subjects. The subject group is composed of 209 males and 81 females, with ages ranging from 17 to 87. Each individual is characterized by one to five records. The electrical device automatically collects ECG signals from different angles, which are also called leads. Each record includes 15 simultaneously measured leads: the conventional 12 leads (i, ii, iii, avr, avl, avf, v1, v2, v3, v4, v5, v6) together with the three Frank lead ECGs (vx, vy, vz). Each ECG signal is digitized at 1000 samples/s, with 16-bit resolution over a range of $\pm$ 16.384 mV.

B. Rationale Behind Using MIL for Automatic Recognition of ECGs With MI

In contrast to standard supervised learning, where a classifier is typically presented with instances associated with class labels, MIL [16] provides a special learning framework in which a classifier is provided with a collection of bags (sets of unlabeled instances) related to class labels. In a binary MIL problem, a positive bag label indicates that at least one of its instances can be assigned a positive label. Instances in negative-labeled bags are definitely from the negative class. The number of instances in different bags can vary greatly. Therefore, by learning a model from a set of labeled bags each of which contains multiple unlabeled instances, the goal of MIL is to predict the class label for a new bag and even those of its instances.

In this paper, given a set of labeled ECGs, our task is to automatically predict whether an unseen ECG reflects MI or not. Basically, this task can be effectively represented as an MIL problem. On one hand, an original training set is usually composed of ECGs, each of which comprises hundreds or thousands of heartbeats. Due to the large total number of heartbeats in the training set, it is unlikely for experts to assign labels to all of them. Thus, the label of each heartbeat is uncertain even when the ECG label is reliable. On the other hand, due to the high death rate associated with MI, it is very necessary to consider an ECG as indicative of MI if at least one of its heartbeats is suspicious for MI. In the framework of MIL, an ECG and its heartbeats correspond to a bag and the instances contained in that bag.

III. MIL for MI Detection

The structure of our LTMIL system contains three main phases: preprocessing, feature extraction, and classification, as shown in Fig. 1. Our LTMIL system can not only differentiate ECGs with MI from normal and other abnormal ones, but also roughly predict which subgroup of MI a patient relates strongly with through the corresponding ECG.

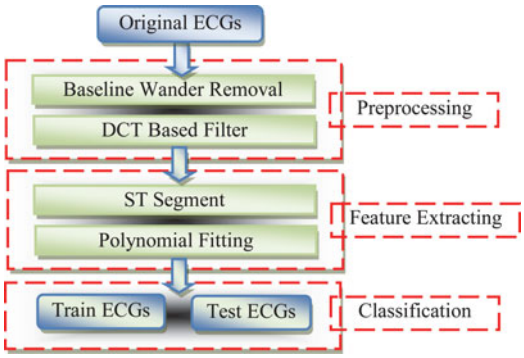

Fig. 1. Structure of the LTMIL system for automatic detection of MI from patients' ECGs.

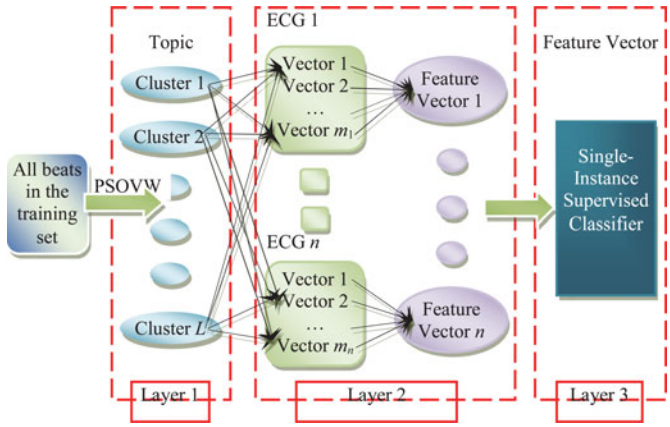

Fig. 2. Hierarchical framework of our LTMIL classification algorithm for automatic detection of MI.

A. Preprocessing

Since ECG signals are captured and recorded by skin electrodes, they are prone to contamination by various types of artifacts or noise. The artifacts usually fall into two types: low-frequency span (e.g., baseline wander) and high-frequency span (e.g., powerline inference, electromyography, and so on) [17]. These two types of spans are both difficult to remove because they usually overlap with real ECG signals in the frequency domain [17], [18]. Recently, Shin *et al.* have introduced an ideal filtering approach based on the discrete cosine transform (DCT), which is especially suitable for processing biological signals [19]. Compared with conventional filtering methods (e.g., FIR or IIR filters), this DCT-based filter has advantages because of its narrow transition band and infinite passband gain [19]. In this study, all ECG signals are processed by a DCT-based bandpass filter [19] whose passband is set to a frequency interval $[f_1\, f_2]$ so as to eliminate the influence of both low- and high-frequency artifacts. As suggested by previous work [20], f_1 should not exceed the cardiac fundamental frequency (CFF), while f_2 is chosen to be 45 Hz, which is slightly lower than the powerline frequency. CFF is calculated as follows: first, QRS complexes are extracted by a bandpass filter [19] whose passband is set to [5 Hz 15 Hz]; the QRS complexes are then transformed by the DCT into the frequency domain, in which the first dominant frequency is regarded as CFF.

B. Feature Extraction

Since MI is strongly reflected in the ST segments of ECGs [21], experiments on MI detection in this paper have been performed through segmentation and analysis of ST segments. First, we detect the characteristic points, such as R, S and T points, using the derivative-based algorithm [22]. After that, we uniformly sample 200 points from the ST segment of each one-lead heartbeat such that the ST segment of each one-lead heartbeat has 200 dimensions. Then, a five-order polynomial fitting is applied to extract features from ST segments. The fitted curves are very similar to the original ST segments, so the information in the ST segments, such as shape, width, height, curvature, etc., is mostly expressed by a few polynomial coefficients (n+1 coefficients for n-order fitting). We have thus reduced a large number of dimensions for each one-lead heartbeat to six polynomial coefficients. Since 12 ECG leads are used for the experiments in this paper, we obtain $6\times12 = 72$ polynomial coefficients for each heartbeat. Two ECG morphology features are also considered: the ratio of the average ST segment length to the average RR interval length for each patient, and the height-to-length ratio of the ST segment for each heartbeat. Thus, each heartbeat is projected into a 74-D space for classification.

C. Automated ECG Classification

The proposed LTMIL system is a three-level hierarchical model, in which each ECG and its heartbeats are represented as mixtures over a set of "topics." Each topic is in turn characterized by a distribution over heartbeats. Since we are trying to extend the techniques used in MIL to ECG, we need to define an analogy between their respective terms.

1) An ECG corresponds to a bag.
2) A heartbeat is equivalent to an instance.
3) A topic is a cluster over heartbeats.

For automated ECG classification, we first identify L topics from the training set of unlabeled heartbeats. We use the clustering method particle swarm optimizer for variable weighting (PSOVW) proposed by Lu *et al.* [23] because of the challenge of dealing with high-dimensional ECG data, but latent Dirichlet allocation with properly extracted ECG features could also be a valid approach. In this paper, each topic is modeled as a Gaussian distribution, which implies a distribution over heartbeats belonging to this topic. We then consider each ECG and its heartbeats as a mixture over the entire set of topics. Finally, a measure of the distance between ECGs can be constructed and any supervised learning-based technique can be applied to automated ECG classification. The framework of our LTMIL algorithm is illustrated in Fig. 2.

In the first layer, we identify L clusters over all the unlabeled beats in the training set, each of which will later be characterized by a Gaussian kernel. The intra-ECG difference, i.e., the heartbeat difference within an ECG, is therefore considered in these identified clusters. PSOVW is a variable-weighting k-means type algorithm and utilizes the swarm intelligence algorithm

particle swarm optimizer (PSO) [24] to find a set of optimal variable weights which transforms the Euclidean distance of high-dimensional data so that the associated cluster is reshaped into a dense hypersphere and can be effectively separated from other clusters. The reason for employing PSOVW instead of the conventional k-means algorithm in LTMIL is that since heartbeats are sparse in the space defined by the 74 features extracted from each beat with 12 leads, variable weighting helps reshape clusters over beats into dense regions and makes them easier to separate. Moreover, the clustering performance of PSOVW is stable due to its use of the PSO search strategy.

Each identified cluster is characterized by a Gaussian kernel, which is formulated as follows:

$$\mathrm{RBF}_l = \exp\left[-\frac{(x-\mu_l)^2}{2\sigma_l^2}\right] \tag{1}$$

where μ_l is the center of the lth cluster, determined by taking the mean value of heartbeats in the cluster, and $\sigma_l = \mu\frac{\sum_{i\neq l}\mathrm{dist}(\mu_i-\mu_l)}{l-1}$, where μ is the user-defined scale factor. Unlike traditional kernel parameter selection, our method for determining the standard deviation is to choose a different for each cluster and rely on the average distance between the basis function centers of cluster l and other clusters.

In the second layer, each heartbeat is in turn rerepresented as a mixture over latent topics. We map each heartbeat into a feature space defined by the L latent topics via a Gaussian kernel based similarity measure. As a result, each heartbeat is projected into an L-dimensional real vector, each component of which represents the generative possibility for a heartbeat given that topic. An ECG is then rerepresented by averaging its heartbeats' corresponding real vectors, weighted according to their respective contributions to recognition of this ECG. Since the abnormal heartbeats within an ECG usually contribute heavily to its classification, they are assigned large weights, while the normal beats are assigned small weights. The weights of heartbeats within an ECG are defined as follows.

Given the ith ECG X_i, whose m heartbeats have been rerepresented as L-dimensional real vectors $\{\mathrm{vector}_i{}^1, \ldots, \mathrm{vector}_i^m\}$, we treat this ECG as an undirected graph G, where each node corresponds to a heartbeat and any two different nodes have an edge only if the Gaussian distance between these two corresponding heartbeats is smaller than a threshold б. The Gaussian distance is defined as $\exp^{-x_i-x_j^2}$, $\forall x_i,\ x_j \in X$, and б is set as the average distance of heartbeats in this ECG.

Assume that $e_i^1, \ldots, e_i^m$ represents the number of edges connecting to each node in the graph G. The weight w_i^j of the jth heartbeat is set as $1/e_i^j$, inversely proportional to the number of the edges connecting it, which implies that within an ECG, the more edges a node is connected to, the less abnormal the corresponding beat is. Besides, the weights of m heartbeats within this ECG should be normalized by $w_i^j = \frac{1/e_i^j}{\sum_{j=1}^m 1/e_i^j}$, subject to $\sum_{j=1}^m w_i^j = 1$.

The ECG is then characterized by a mixture over topics, a weighted combination of the corresponding mixtures over topics for heartbeats, as shown in (2)

$$\mathrm{vector}_i = \sum_{j=1}^{m} w_i^j.\mathrm{vector}_i^j. \tag{2}$$

Therefore, the inter-ECG difference can be measured by the difference between mixtures over these L topics.

Finally, once labeled ECGs with varying numbers of heartbeats are transformed into real vectors with L dimensions, any supervised learning algorithm such as SVM, NN, KNN, random forest (RF), ensemble learning, etc., can be applied to the classification of the transformed ECGs. Since KNN has been demonstrated to be a simple but efficient classifier, we have opted to build KNN rather than SVM or NN into the LTMIL model. Since the performance of KNN is very sensitive to the size of the k nearest neighbors, we propose a simple ensemble learning method to enhance the robustness of KNN. We run KNN on the ECG datasets with different k in the range 3–31 with an interval of 2. Our approach to select an ensemble m out of all KNN classifiers is described in Algorithm 1.

Algorithm 1. KNN ensemble

KNNensemble

for k = 3 to 31
  The fuzzy set $S_k = \varphi$;
  For each test data object j,
    Among its k nearest neighbors, calculate the difference between the number of positively labeled neighbors P_j and the number of negatively labeled neighbors N_j, i.e. $|P_j - N_j|$;
    If $|P_j - N_j| < p(k)$, $S_k = \{j\} \cup S_k$.
end
Select m KNN classifiers with the minimal size of the fuzzy set.
Vote these m KNN classifiers.

The fuzzy set S_k implies the uncertainty of data objects being correctly classified by a specific KNN classifier. $p(k)$ is the threshold value for declaring the certainty of one data object being correctly classified by a given KNN classifier. We set $p(k) = 1+k/5$, which indicates that the larger the number of neighbors k, the more tolerance there is for the certainty. The pseudocode for the LTMIL algorithm is given in Algorithm 2.

IV. Experimental Results

To illustrate the performance of LTMIL, we conducted a series of experiments on the PTB diagnostic ECG database. The first group of experiments tests the ability of each algorithm to discriminate between ECGs with and without MI. Since MI has multiple, varied signatures on ECG waveforms, for the second group of experiments we chose ECGs with four main groups of MI and the healthy control (HC) ECGs. The series of two-class classifications on individual pairs of these five classes further investigates LTMIL's performance in predicting patients related to specific groups of MI.

In our experiments, we chose a number of algorithms for comparison, including four well-known MIL algorithms (expectation maximization-diverse density (EM-DD) [25], MI-KNN [26], radial basis function (RBF)-MIP [27], and CCE [28]), two typical supervised learning techniques (SVM and NN), and

Algorithm 2. Pseudo-code describing the LTMIL model.

```
LTMIL(B,Cluster,Classifier)
Input:
B: A set of n labeled bags{ X_1, X_2, …, X_n }; Cluster: the PSOVW clustering algorithm; Classifier: the standard classification algorithm;

Z←φ   % Z records all instances in the training set of n bags.
for each bag X_i□B   % Assume that mi is the number of instances in the bag X_i.
    for each instance x_i^j□X_i (j=1, …, mi)
        Z←Z□{x_i^j}

(cluster_1, …, cluster_L) ← Cluster(Z, L)  % Cluster Z into L clusters.
Φ_l← BuildRBF(cluster_l) (l=1,  , L)  % Build a Gaussian kernel for each cluster.

S←φ  % S records transformed bags and their related class labels.
for each bag X_i□B
    for each instance x_i^j□X_i
        vector_i,^j = ProjectRBF(Φ_1 ,…, Φ_L , x_i^j )
        (w_i^1…w_i^mi)=GetWeights (vector_i,^1 … vector_i^mi )
          vector_i,←Weighted( w_i^1...w_i^mi, vector_i,^1...vector_i^mi )
        ylabel_i← Getlabel(X_i)
        S←S□{< vector_i, , ylabel_i >}

C←TrainClassifier (S)

Output:
for each instance x_t^j (j=1,…, m) in the test bag X_t % m is the number of instances in the bag X_t.
    vector_t^j = ProjectRBF(Φ_1 ,…, Φ_L , x_t^j )
    (w_t^1,..,w_t^m) = GetWeights (vector_t^1..vector_t^m )
    vector_t ← Weighted ( w_t^1..w_t^m, vector_i,^1..vector_i^m )
    Label(X)← Predict (Vector_t, C)                    (3)
    Label(x_t^j)← Predict (vector_t^j, C)              (4)
```

three other SSL methods (SVM and NN with expert knowledge and active learning [12]). We implemented LTMIL mainly in MATLAB combined with C codes, such as PSOVW being called through the mexFunction interface. We utilized the MATLAB source codes of EM-DD, MI-KNN, RBF-MIP, and CCE obtained from Zhou's website [29]. For the supervised learning techniques, we used libSVM and NN in the toolbox of MATLAB 2010a to classify heartbeats by considering the label of an ECG as the labels of its heartbeats. For SSL methods such as SVM and NN with expert knowledge, cardiologists in the Department of Cardiovascular Medicine, First Affiliated Hospital of Xiamen University labeled the first 60 heartbeats of each ECG and SVM and NN were trained on annotated heartbeats to classify heartbeats. For the active learning algorithm provided by Pasolli and Melgani [12], we implemented it in MATLAB using libSVM instead of SVMlight to train the linear SVM at each iteration. The parameters were set to the values in the paper [12] and the cardiologists validated the query heartbeats. All the algorithms used the same preprocessing and features as our proposed LTMIL method.

A. Parameter Settings

In LTMIL, CCE, and RBF-MIP, there are two input parameters that need to be specified, i.e., the number of topics L required by clustering over instances and the scale factor μ. In the following experiments, we test them under different parameter configurations. For statistical purposes, all the experimental results were obtained by running tenfold cross validation 20 times. In detail, in each run with tenfold cross validation, each class of ECGs was randomly and averagely partitioned into ten parts, nine of which were used to train classifiers and the remaining one was employed to test the classifiers' performances. The average result is reported for each algorithm.

In the following experiments, we set the RBF kernel function used in SVM; the cost parameter C is set to e5, which indicates a high cost of misclassifying points; and the tolerance of termination criterion e is set to a small value e−5. In NNs, the network is created with one hidden layer of 50 neurons. Other parameters are set to default values.

B. Performance Evaluation

Since automated ECG classification for computer-aided diagnosis (CAD) is a tool to identify a heart disease from patients' ECGs, we use the statistical measures sensitivity and specificity to test the performance of binary classification tests. Here, consider a set of patients' records on a binary medical test conducted to identify a disease, where each patient taking the test either has or does not have the disease and the test output can be positive or negative, i.e., predicting that the patient has or does not have the disease. For binary classification tasks, sensitivity and specificity are then formulated in the following equations:

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{6}$$

where TP, FP, TN, and FN represent the number of sick patients correctly diagnosed as sick, healthy patients incorrectly diagnosed as sick, healthy patients correctly identified as healthy, and sick patients incorrectly identified as healthy, respectively. Theoretically, a perfect prediction achieves 100% sensitivity and 100% specificity. However, any actual classifier has to obtain a tradeoff between these two measures.

C. Experimental Results

1) Binary Classification Between MI ECGs and HC ECGs: In the first group of experiments for binary classification, we examine the performance of each algorithm to discriminate between ECGs with MI and HC ECGs. Specifically, 369 ECG records with MI and 79 HC ECG records were chosen from the PTB database for these experiments.

We first tested the performance of LTMIL with different parameter configurations. Considering its two parameters L and μ, we executed LTMIL under different parameter configurations. In detail, the number of clusters L ranges from 5 to 100 with an interval of 5 and the scale factor μ ranges from 0.1 to 0.9 with an interval of 0.2. For each parameter configuration, we ran LTMIL 20 times. The average sensitivity and specificity achieved by LTMIL with different parameter configurations are shown in Fig. 3.

From Fig. 3, we can see that for $\mu = 0.1$, the sensitivity of LTMIL falls dramatically from above 97% to approximately 85% on average as L increases from 5 to 100. On the other

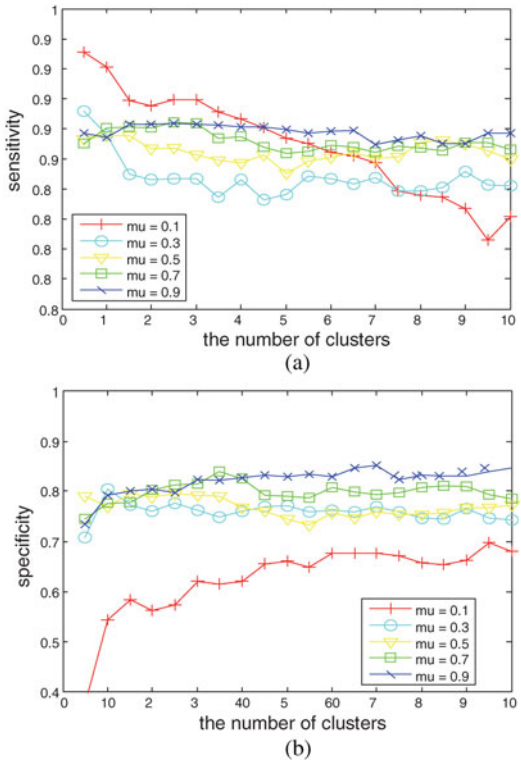

Fig. 3. Average sensitivity and specificity achieved by LTMIL change as the number of clusters L and the scale factor μ increase.

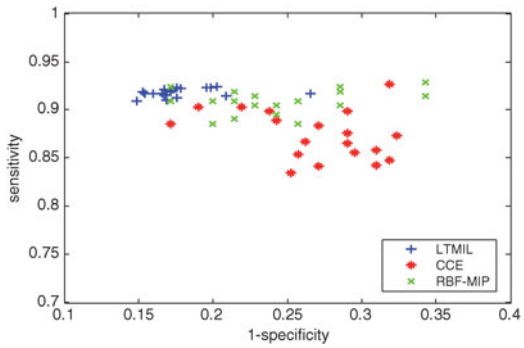

Fig. 4. ROC points obtained by compared algorithms with varying number of clusters.

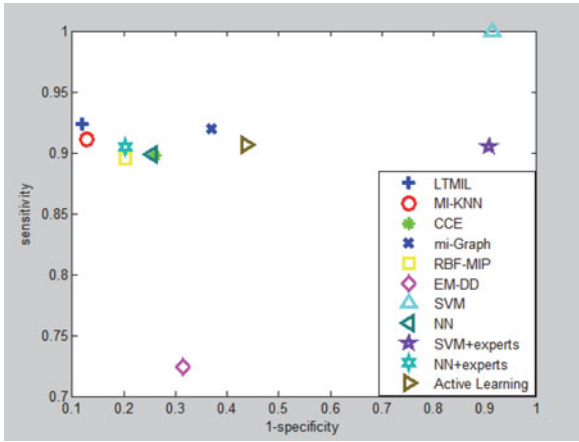

Fig. 5. ROC points achieved by each algorithm with the best.

hand, the specificity grows stably from below 40% to nearly 70%. When μ takes the values of 0.3 and 0.5, regardless of the value of L, the sensitivity of LTMIL fluctuates between 87% and over 90% and its specificity reaches 75%. From the two figures in Fig. 3, we can clearly see that for μ taking the value of 0.9, the sensitivity and specificity curves for LTMIL remain at a high but stable level when compared to μ taking other values. Therefore, we set μ to a value of 0.9 throughout the following experiments.

We also investigated the performance of the three hierarchical algorithms, LTMIL, CCE, and RBF-MIP, which have a common parameter, the number of clusters L. So, we tested the performances of LTMIL, CCE, and RBF-MIP with different L in the range of 5–100 with an interval of 5. The prediction results achieved by CCE, RBF-MIP ($\mu = 0.3$ [28]), and LTMIL ($\mu = 0.9$) are plotted in the receiver operating characteristic (ROC) space in Fig. 4. In this figure, the horizontal axis indicates the value of the specificity subtracted from one and the vertical axis indicates the value of the sensitivity in the ROC space. The best possible prediction algorithm will yield a point in the upper left corner of the ROC space, representing 100% sensitivity and 100% specificity.

From Fig. 4, we can clearly see that the points located by LTMIL are in the upper left portion of the ROC space, which means that, on average, LTMIL achieves better performance in terms of ROC than do CCE and RBF-MIP. The fact that the points identified by LTMIL are much denser than those obtained by CCE and RBF-MIP indicates that, of the three algorithms compared, LTMIL is the least sensitive to the parameter L. It should be noted that clustering in RBF-MIP is over labeled ECGs, while in LTMIL and CCE, clustering is over unlabeled heartbeats. In the following comparative studies, we set the number of clusters L to 30 for these three hierarchical algorithms.

For a clearer comparison, we assessed the sensitivity and specificity achieved by LTMIL, several typical well-known MIL algorithms (DD, MI-KNN, and mi-Graph), two standard supervised learning methods (SVM and NN), and three semisupervised techniques (SVM and NN with expert knowledge, and active learning). We ran each algorithm with a fixed parameter configuration 20 times. Fig. 5 plots the average result of each algorithm in the ROC space. In Table I, we also record the average sensitivity and specificity of each algorithm and examine the variance of the classification results. These experiments provide information on how well each algorithm is able to retrieve known classes for recognition of MI from patients' ECGs and how sensitive each algorithm is to the initial parameters. The three best results are marked in boldface for each performance measure.

A number of comments can be made by analyzing the ROC points plotted in Fig. 5 and the results recorded in Table I. First, LTMIL seems to perform better than the other algorithms tested on the real ECG dataset. In fact, it simultaneously achieves the third best sensitivity and the best specificity. Although SVM-based supervised and SSL techniques achieve the highest two sensitivities (100% and 99.3%, respectively), they obtain the

TABLE I
PERFORMANCE OF COMPARED MIL AND ML-BASED ALGORITHMS

Algorithms		Sensitivity	Specificity
Multiple Instance Learning	LTMIL	**92.3**±0.84	**88.1**±2.3
	MI-KNN	91.1±1.5	**87.3**±1.6
	CCE	89.8±1.7	74±6.8
	mi-Graph	92±3.9	63±4.4
	RBF-MIP	89.52±1.5	79.71±1.7
	EM-DD	72.4±3	68.5±16.9
Supervised Learning	SVM	**100**±0	8.64±0.16
	NN	89.9±1.2	74.5±1.9
Semi-supervised Learning	SVM+ experts	**99.3**±0.41	9.17±3.1
	NN+ experts	90.5±1.3	**79.81**±2.1
	Active Learning	90.7±1.5	56.3±3.4

TABLE II
COMPARATIVE PERFORMANCE OF LTMIL WITH KNN ENSEMBLE, SVM, NN, AND RF

Algorithms	Sensitivity	Specificity
LTMIL+KNN ensemble	92.3±0.84	**88.1**±2.3
LTMIL+SVM	**92.6**±0.65	82.4±4
LTMIL+NN	91.14±0.88	80.43±2.6
LTMIL+RF (random forest)	91.43±1.08	77.29±4.22

worst specificities, under 10%, which indicates that SVM barely identifies ECGs with MI. mi-Graph performs as well as LTMIL in terms of sensitivity (92%), but does poorly in terms of specificity (63%). LTMIL performs better than CCE on both sensitivity and specificity, although they both use the cluster space defined by clusters over instances. This is expected, as LTMIL employs more implicit strategies such as RBF kernels for representing clusters and the instance ensemble within a bag. Moreover, the performance of LTMIL is also stable, as it employs the PSO-based clustering technique rather than the k-means used in CCE, which is sensitive to initial cluster centroids.

Second, there is a clear trend in the performance of the MIL algorithms. MIL methods based on local classifiers such as LTMIL, RBF-MIP, and MI-KNN achieve better performance in terms of specificity and their sensitivity is also comparable with that of the MIL algorithms based on global classifiers, such as CCE and mi-Graph. Table I shows the performance of LTMIL combined with SVM, which achieves slightly higher sensitivity and much lower specificity than LTMIL combined with the KNN ensemble. Although sensitivity is a more important evaluation than specificity in CAD systems which do not want to miss patients with diseases based on the results of medical tests, we prefer LTMIL with the KNN ensemble because LTMIL with the KNN ensemble and with SVM are very close in terms of sensitivity.

Finally, the MIL and SSL algorithms basically perform better than supervised learning methods such as SVM and NN, except for EM-DD. This is because supervised learning methods take the class label of a bag to annotate the true class labels of its corresponding instances. This does not correspond to the actual situation in many real-life MIL problems, where positive bags contain instances not only from the positive class but also from the negative one. Especially, in this ECG classification problem, the fact that a patient suffers from MI does not mean each of the patient's heartbeats will reflect MI. The reason EM-DD achieves a poor performance in terms of both sensitivity and specificity is that it is especially designed for the molecular drug prediction problem. It is also very sensitive to initial points because of its local search strategy.

It is easy to understand that due to the expert knowledge incorporated in the form of labels of some heartbeats, SSL techniques improve the performance of the ECG classification to some extent. However, it can still be seen that LTMIL is superior to these three SSL methods. This is because SVM and NN are trained on the subset of the first 60 heartbeats of each ECG labeled by cardiologists, which is not adequate to represent the whole data distribution; and the performance of the active learning method relies on the quality of the set of heartbeats selected by clustering, which tends to discard small clusters carrying important diagnostic information.

To further investigate the reasons why LTMIL achieves better performance than other algorithms on the real ECG dataset, we also implemented the LTMIL model with different classifiers presented in MIL algorithms, namely SVM, NN, and RF. LTMIL is a preprocessed model where bags are mapped in a feature space onto single instance feature vectors; so, it is flexible and can be employed to combine with other supervised-learning classifiers. Here, LTMIL with each classifier was run for 20 trials. The average sensitivity and specificity yielded by the LTMIL method with each classifier are recorded in Table II. The results for LTMIL with the KNN ensemble are taken from Table I.

Table II gives the classification results of the LTMIL model with different classifiers on the ECG dataset. Obviously, the LTMIL model greatly improves the sensitivity and specificity of MI-KNN, CCE, mi-Graph, and RBF-MIP. From Tables I and II, we see that, on average, MI-KNN and RBF-MIP are less efficient than LTMIL+KNN because they identify clusters over bags based on Hausdorff distances, which yields inexplicable results because the instances in a bag vary considerably. LTMIL+SVM performs better than CCE and mi-Graph, although they employ the same classifier, SVM. As we explained previously, this is understandable, as LTMIL employs more implicit strategies such as RBF kernels representing clusters and the instance ensemble within a bag. In our experiments, we found that the high sensitivity yielded by the LTMIL model mainly benefits from the instance ensemble within a bag, while the high specificity is mainly derived from the RBF kernel representation of clusters over instances.

The results reported in Tables I and II suggest that the good performance of LTMIL is due to cluster identifications over instances, RBF kernels representing clusters, and the mechanism of the instance ensemble within a bag.

2) Binary Classification of Each Pair of Five Classes of ECGs: In order to further demonstrate the ability of LTMIL to predict patients most related to a particular group of MI, we then performed a two-class classification for each pair of five

TABLE III
PERFORMANCE COMPARISON ON SUBGROUPS OF ECGS

Accuracy	MI 1	MI 2	MI 3	MI 4	HC Sensitivity
MI 1	n/a	89.25±2.6	66.82±1.9	65.55±2.2	91.34±1.8
MI 2		n/a	91.78±2.8	89.33±1.2	95.86±1.9
MI 3			n/a	57.12±2.1	92.14±2.8
MI 4				n/a	91.14±2.6
HC Specificity	89.14±2.2	93.6±1.8	88.75±4.7	85.75±3.9	n/a

For classification conducted between any two out of four MI ECGs, the accuracy and corresponding standard deviation achieved by LTMIL are recorded. For classification carried on between any MI ECGs and HC ECGs, the sensitivity and specificity are used to measure the performance.

classes of MI ECGs. Depending on where infarction occurs in ECGs, MI has been divided into multiple groups in clinical diagnosis. In the PTB diagnostic ECG database, there are 14 groups of MI reflections in ECGs. Since nine of these groups of MI ECGs contain too few ECGs to be used to train classifiers, we chose the other five main groups of MI ECGs and the HC ECGs for our experiments. (Two of the MI groups, inferior and inferolateral, were ultimately merged on the advice of our physician advisors.) There are therefore four groups of MI used for experiments, i.e., anteroseptal (MI 1), inferior and inferolateral (MI 2), anterior (MI 3), and anterolateral (MI 4), containing 75, 141, 47, and 42 ECG records, respectively. The number of HC ECGs (HC) is 79. The evaluation results are recorded in Table III.

Table III illustrates that in addition to performing well on two-class distinction of MI ECGs from HC ECGs, LTMIL is also able to roughly distinguish different groups of MI ECGs. From the aforementioned table, it can be readily seen that LTMIL achieves over 91% sensitivity and over 85% specificity on distinction of MI ECGs of all kinds from HC ECGs. On average, it also yields approximately 90% accuracy in distinguishing the inferior and inferolateral (MI 2) group of ECGs from other groups of MI ECGs, although it achieves low classification accuracy (about 60%) in classification of other groups of MI ECGs. (This is because anteroseptal (MI 1), anterior (MI 3), and anterolateral (MI 4) have very similar reflections in heartbeats.) Moreover, LTMIL yields low variance (not more than three) in classification accuracy, sensitivity, and specificity because it employs the KNN ensemble.

V. CONCLUSION

The goal of this paper is to design an automatic system for detection of MI from patients' ECGs. This problem is challenging due to incomplete labeling of heartbeats, as experts cannot label a very large quantity of heartbeats. This paper addresses the difficulty with a LTMIL strategy, which produces a new representation of ECG to construct a distance measure that takes both inter-ECG and intra-ECG differences into consideration. Additionally, it takes full advantage of the ECG labels, while not requiring any heartbeat labels.

When tested on real ECG datasets from the PTB diagnostic database, our method achieves better performance than other algorithms based on MIL, and supervised and SSL. We ran the experiments with the help of cardiologists, who are very satisfied with the experimental results. Although the proposed method has not yet been demonstrated to actually improve clinical decision making by clinical evaluation, these preliminary results are highly encouraging and suggest that MIL can be used practically in the recognition of MI. It requires little expert labor for labeling ECGs and boosts performance by taking advantage of inter- and intra-ECG differences.

Of course, there is room for improvement. In all the experiments, we used identical input parameters; further tuning of these parameters may improve results. More work is needed to investigate automated parameter tuning. Based on preliminary experiments, it may also be possible to improve performance by first learning the optimal number of clusters and then building better classifiers.

REFERENCES

[1] P. Chazal, M. Dwyer, and R. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196–1206, Jul. 2004.
[2] F. M. Nolle, R. W. Bowser, F. K. Badura, J. M. Catlett, R. R. Guadapati, T. T. Hee, A. N. Mooss, and M. H. Sketch, Sr., "Evaluation of a frequency-domain algorithm to detect ventricular fibrillation in the surface ECG," in *Proc. Comput. Cardiol.*, 1988, pp. 337–340.
[3] S. Mitra, M. Mitra, and B. B. Chaudhuri, "A rough set based inference engine for ECG classification," *IEEE Trans. Instrum. Meas.*, vol. 55, no. 6, pp. 2198–2206, Dec. 2006.
[4] L. Shyu, Y. Wu, and W. Hu, "Using wavelet transform and fuzzy neural network for VPC detection from the holter ECG," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1269–1273, Jul. 2004.
[5] O. T. Inan, L. Giovangrandi, and G. T. A. Kovacs, "Robust neural-network based classification of premature ventricular contractions using wavelet transform and timing interval features," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2507–2515, Dec. 2006.
[6] W. Jiang and S. Kong, "Block-based neural networks for personalized ECG signal classification," *IEEE Trans. Neural Netw.*, vol. 18, no. 6, pp. 1750–1761, Nov. 2007.
[7] F. Melgani and Y. Bazi, "Classification of electrocardiogram signals with support vector machines and particle swarm optimization," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 5, pp. 667–677, Sep. 2008.
[8] A. Kampouraki, G. Manis, and C. Nikou, "Heartbeat time series classification with support vector machines," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 4, pp. 512–518, Jul. 2009.
[9] P. Chazal and R. Reilly, "A patient adapting heartbeat classifier using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2535–2543, Dec. 2006.
[10] T. Ince, S. Kiranyaz, and M. Gabbouj, "A generic and robust system for automated patient specific classification of ECG signals," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 5, pp. 1415–1426, May 2009.
[11] M. Faezipour, A. Saeed, S. C. Bulusu, M. Nourani, H. Minn, and L. Tamil, "A patient adaptive profiling scheme for ECG beat classification," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 5, pp. 1153–1165, Sep. 2010.
[12] E. Pasolli and F. Melgani, "Active learning methods for electrocardiographic signal classification," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 6, pp. 1405–1416, Nov. 2010.
[13] J. Wiens and J. Guttag, "Active learning applied to patient adaptive heartbeat classification," in *Proc. Neural Inf. Process. Syst.*, 2010, pp. 2442–2450.
[14] [Online]. Available: http://www.physionet.org/physiobanK/database/ptbdb
[15] R. Mark and G. Moody. MIT-BIH arrhythmia database directory. [Online]. Available: http://ecg.mit.edu/dbinfo.html
[16] T. G. Dietterich, R. H. Lathrop, and T. L. Perez, "Solving the multiple instance problem with axis parallel rectangles," *Artif. Intell.*, vol. 89, pp. 31–71, 1997.
[17] J. M. Lski and N. Henzel, "ECG baseline wander and powerline interference reduction using nonlinear filter bank," *Signal Process.*, vol. 85, no. 4, pp. 781–793, 2005.

[18] M. Velasco, B. Weng, and K. E. Barner, "ECG signal denoising and baseline wander correction based on the empirical mode decomposition," *Comput. Biol. Med.*, vol. 38, no. 1, pp. 1–13, 2008.
[19] H. S. Shin, C. Lee, and M. Lee, "Ideal filtering approach on DCT domain for biomedical signals: Index blocked DCT filtering method (IB-DCTFM)," *J. Med. Syst.*, vol. 34, no. 4, pp. 741–753, 2010.
[20] J. S. Paul, M. R. S. Reddy, and V. J. Kumar, "Data processing of stress ECGs using discrete cosine transform," *Comput. Biol. Med.*, vol. 28, no. 6, pp. 639–658, 1998.
[21] R. S. Wright, J. L. Anderson, C. D. Adams, C. R. Bridges, D. E. Casey, Jr., S. M. Ettinger, F. M. Fesmire, T. G. Ganiats, H. Jneid, A. M. Lincoff, E. D. Peterson, G. J. Philippides, P. Theroux, N. K. Wenger, J. P. Zidar, E. M. Antman, R. M. Califf, W. E. Chavey, J. S. Hochman, and T. N. Levin, "2011 ACCF/AHA focused update of the Guidelines for the Management of Patients with Unstable Angina/Non–ST-Elevation Myocardial Infarction (updating the 2007 guideline): A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines developed in collaboration with the American College of Emergency Physicians, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons," *J. Amer. Coll. Cardiol.*, vol. 57, no. 19, pp. 1920–1959, 2011.
[22] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. 32, no. 3, pp. 230–236, Mar. 1985.
[23] Y. Lu, S. Wang, S. Li, and C. Zhou, "Particle swarm optimizer for variable weighting in clustering high-dimensional data," *Mach. Learn.*, vol. 82, pp. 43–70, 2011.
[24] R. C. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proc. 6th Int. Symp. Micromach. Human Sci.*, Nagoya, Japan, 1995, pp. 39–43.
[25] Q. Zhang and S. Goldman, "Em-DD: An improved multiple instance learning technique," in *Proc. Neural Inf. Process. Syst.*, 2002, pp. 1073–1080.
[26] J. Wang and J. D. Zucker, "Solving the multiple instance problem: A lazy learning approach," in *Proc. 17th Int. Conf. Mach. Learn.*, 2000, pp. 1119–1125.
[27] Z. Zhou and M. Zhang, "Solving multi-instance problems with classifier ensemble based on constructive clustering," *Knowl. Inf. Syst.*, vol. 11, no. 2, pp. 155–170, 2007.
[28] M.-L. Zhang and Z.-H. Zhou, "Adapting RBF neural networks to multi-instance learning," *Neural Process. Lett.*, vol. 23, no. 1, pp. 1–26, 2006.
[29] [Online]. Available: http://cs.nju.edu.cn/zhouzh

Li Sun received the B.E. degree in computer science from Fujian Normal University, Fujian, China, in 2010. From 2010 to 2012, he as a master student had studied in the area of machine learning applied to ECG analysis in Xiamen University, China. He is currently a Ph.D. candidate in machine learning and computer vision at the College of Science and Engineering, University of Glasgow, Glasgow, U.K.

Yanping Lu (M'11) received the Ph.D. degree in data mining from the University of Sherbrooke, Sherbrooke, QC, Canada, and the second Ph.D. degree in machine learning from Xiamen University, Fujian, China, in 2009.

Since October 2009, she has been an Assistant Professor in the Department of Cognitive Science, Xiamen University. Her current research activity is related to machine learning, data mining and applications in ECG analysis, image classification, object tracking, etc.

Kaitao Yang received the B.E. degree in electrical engineering from the Minzu University of China, Beijing, China, in 2009, and the M.S. degree in signal processing from Xiamen University, Fujian, China, in 2012. From September 2012, he will start working toward the Ph.D. degree in biomedical signal processing in the Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands.

Shaozi Li received the doctoral degree in computer vision form the National University of Defense Technology in 2009.

He is a Professor at Xiamen University, Fujian, China, where he is currently the Director of the Department of Cognitive Science. His research interests include pattern recognition, machine learning techniques applied to computer vision, and biomedical signal/images. He has coauthored more than 70 publications in these areas.