An Automatic Subject-Adaptable Heartbeat Classifier Based on Multiview Learning

Can Ye, *Member, IEEE*, B.V. K. Vijaya Kumar, *Fellow, IEEE*, and Miguel Tavares Coimbra, *Member, IEEE*

***Abstract*—In this paper, a novel subject-adaptable heartbeat classification model is presented, in order to address the significant interperson variations in ECG signals. A multiview learning approach is proposed to automate subject adaptation using a small amount of unlabeled personal data, without requiring manual labeling. The designed subject-customized models consist of two models, namely, *general classification model* and *specific classification model*. The general model is trained using similar subjects out of a population dataset, where a pattern matching based algorithm is developed to select the subjects that are "similar" to the particular test subject for model training. In contrast, the specific model is trained mainly on a small amount of high-confidence personal dataset, resulting from multiview-based learning. The learned general model represents the population knowledge, providing an interperson perspective for classification, while the specific model corresponds to the specific knowledge of the subject, offering an intraperson perspective for classification. The two models supplement each other and are combined to achieve improved personalized ECG analysis. The proposed methods have been validated on the MIT-BIH Arrhythmia Database, yielding an average classification accuracy of 99.4% for ventricular ectopic beat class and 98.3% for supraventricular ectopic beat class, which corresponds to a significant improvement over other published results.**

***Index Terms*—Automatic subject adaptation, ECG heartbeat classification, multiview learning, similar training data selection.**

I. Introduction

Electrocardiogram (ECG or EKG) records electrical activity of the heart over time. Computer-assisted ECG interpretation has proven to be very useful in timely detection and better management of cardiac disorders. Cardiac arrhythmias refer to a large group of conditions in which there is abnormal activity or behavior in heart [1]. Current ECG analysis algorithms typically adopt a beat-by-beat analysis strategy. Arrhythmia detection essentially determines the category of an ECG sequence, by recognizing the class labels of heartbeats within the given sequence. Therefore, automatic heartbeat classification is an essential step toward the arrhythmia detection. The automatic heartbeat classification of ECG signals is the focus of this study.

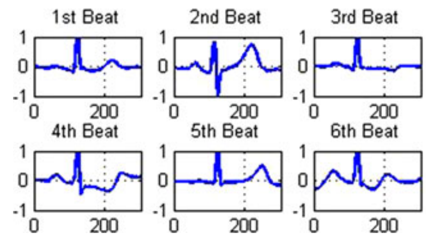

Fig. 1. Six normal beat patterns (on modified limb lead II) extracted from six subjects in MIT-BIH Arrhythmia Database (records 100, 103, 105, 112, 113, and 117).

One major challenge for reliable automatic ECG analysis is the significant variability in waveforms and characteristics of ECG signals among different patients and patient groups [2], termed as *interperson variations*. For instance, Fig. 1 shows six normal heartbeat patterns extracted from six different subjects. Such variations are due to the underlying unique physiological structure of each individual's heart. Most existing ECG analysis algorithms typically employ a *fixed mode*, in which a classifier is trained on a general population ECG database and then the classifier is applied to classify ECG signals from a new test subject without adaptation. Due to interperson variations, such fixed models that may perform well on a representative training database, could still fail in generalizing to an untrained subject, because the underlying distribution of the test subject might differ significantly from the distribution representing the training database. Motivated by this issue of interperson variations, we develop and investigate algorithms to construct subject-adaptable ECG heartbeat classification models, which are capable of adapting to the characteristics of a particular subject.

There have been a number of previous studies [2]–[7], which investigated the design of patient-specific heartbeat classifiers. These studies utilize the initial 5-min data from a particular subject as the *individual training data* for adaptation, in accordance with the ANSI/AAMI EC57 standard [8]. The major issue with these approaches is that they rely on an *expert-intervention mode*, which requires experts providing the ground truth labels of the individual training data, i.e., beat-by-beat annotations on the initial 5-min data. Llamedo and Martinez [6] presented a method to reduce the amount of such expert intervention, in which they proposed the clustering analysis on the individual training set and having physicians assist in annotating the centroids of clusters. Nevertheless, these algorithms typically

Manuscript received February 9, 2015; revised June 6, 2015 and July 21, 2015; accepted August 1, 2015. Date of publication August 13, 2015; date of current version December 6, 2016. This work was supported by the Portuguese Foundation for Science and Technology under Grant SFRH/BD/33519/2008.

C. Ye is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA, and also with the Department of Computer Science, Faculty of Science, University of Porto, Porto 4099-002, Portugal (e-mail: cany@ece.cmu.edu).

B. V. K. Vijaya Kumar is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: kumar@ece.cmu.edu).

M. T. Coimbra is with the Instituto de Telecomunicações, Faculdade de Ciências da Universidade do Porto, Porto, Portugal (e-mail: mcoimbra@dcc.fc.up.pt).

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/JBHI.2015.2468224

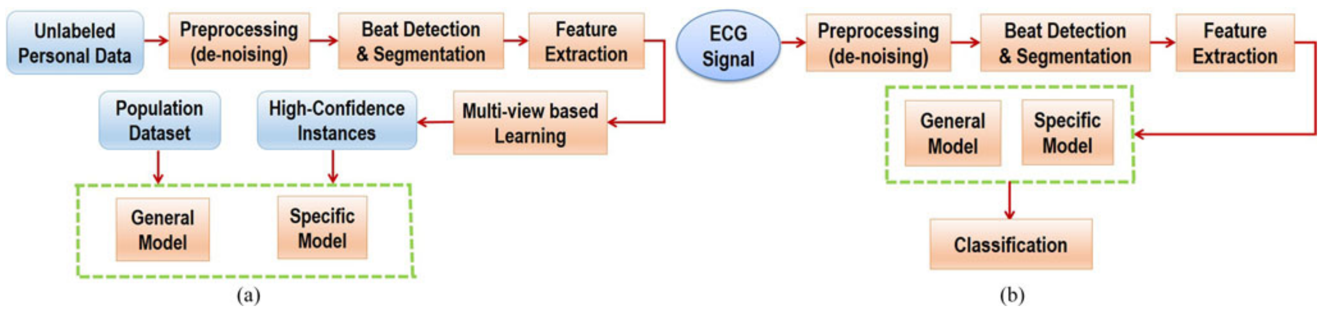

Fig. 2. Proposed algorithm of constructing subject-customized models for ECG heartbeat classification. (a) Model training framework. (b) Model testing framework.

require manual intervention in developing patient specific models (SMs). Such manual intervention is expensive, time-consuming, and sometimes unavailable, making it less practical and appealing for real-world applications, particularly considering the need of ambulatory health monitoring technologies for populations in areas with few medical professionals.

Despite using different features and classification models, these studies typically adopt a customized general classifier approach. A general classifier is first learnt based on a general population ECG database. When testing on a particular subject, the general model (GM) is adapted using the initial 5-min data from the test subject. There are two drawbacks for this approach. First, it employs the entire population database for training without discrimination. Due to the heterogeneous nature of the general population database (collected from diverse patient groups with different health conditions), it is likely that some patterns in the general training database are too dissimilar to be useful for classifying ECG of a particular subject. Sometimes, it can be even misleading to transfer the knowledge from dissimilar sources. Second, the size of individual training dataset is much smaller than the size of the general population training dataset. As a result of such imbalance in the training dataset, the learnt GMs are still biased to the patterns in the population training dataset. It is worth investigating the benefits of basing the decisions more strongly on the available personal evidence.

Motivated by the above, a novel subject-adaptable model is developed for ECG heartbeat classification, in order to address the interperson variations and achieve improved personalized ECG analysis. We aim at solving the following two problems in this study: 1) *how to perform the subject adaptation automatically, without any manual intervention?* 2) *how to design the subject-customized models based on a general population dataset and a small personal dataset?* Compared to the published studies, our innovations are threefold.

First, a multiview-based learning method is proposed for the automatic subject adaptation, termed as *automatic adaptation mode*. The proposed multiview learning method allows us to extract a set of high-confidence heartbeat instances from the original unlabeled individual dataset, representing reliable person-specific information which can be used for the development of person SMs. In contrast to the expert intervention mode, the construction of patient SMs is fully automatic, making it more appealing for real-world applications.

Second, a novel approach is presented for selecting suitable training data for a given subject. Instead of using the entire general population dataset, a subset is selected as suitable training data, which are similar to the given subject, based on a similarity measure between different subjects. These "similar" training subjects are utilized for learning the general classification models for the given subject, while the "dissimilar" training subjects are excluded from training. The learnt GM represents relevant population knowledge, which can be deemed as an interperson classification model.

Third, an SM is proposed in addition to the GM, aiming to emphasize the available individual evidence. The SM is trained mainly on the high-confidence individual dataset of the test subject, resulting from multiview-based personal learning. The SM is essentially an intraperson classification model, which can provide a better discrimination between individual normal and abnormal patterns. The SM and the GM are complementary to each other. We demonstrate that the combination of the two models contributes to the improved heartbeat classification performance.

The rest of the paper is organized as follows. Section II presents details of proposed methodologies, including the multiview learning method for automatic subject adaptation, and the framework of subject-customized models. The experimental results and a comparison with other studies are shown in Section III and conclusions are provided in Section IV.

II. Methodology

Fig. 2 presents the proposed architecture for automatically subject-adaptable models, which can be divided into training and testing stages. The preprocessing and feature extraction methodology is adopted from our previous study [9]. Baseline wander was corrected and the ECG signals were bandpass filtered to remove artifacts. The heartbeats were then segmented by extracting 0.83 s long signal surrounding each R peak. Each heartbeat is represented by a set of morphological and dynamic features, which contain wavelets, ICA, and RR interval features.

In the model training stage, the proposed multiview learning approach is employed to extract a subset of high-confidence instances from initial 5-min data of the subject. We refer to it as the automatic adaptation mode, since no manual labeling is required. The alternative is the expert intervention mode, where an expert is needed to label the individual dataset. The resulting individual dataset is combined with a general population dataset for training subject-customized models. The design of novel subject-customized models is the focus of this paper and will be discussed in detail in this section.

A. Automatic Subject Adaptation

1) Multiview-Based Semisupervised Learning: Semisupervised learning has emerged as an important research topic in machine learning during the past decade. In many automatic applications, it is likely that obtaining data is easy and cheap, but acquiring the reference labels for the data is expensive and time-consuming, or even impossible due to constraint of resources. The major problem addressed by semisupervised learning is how to exploit information contained in unlabeled data in conjunction with labeled data to boost learning performance, compared to only using labeled data [10].

Incorporating unlabeled data inevitably brings in risk for the classification task. One natural question arising is: *Does semisupervised learning always produce superior performance compared to using only labeled data?* The answer is "yes" in principle, as long as unlabeled data carries useful knowledge for the inference of test data [10]. In our problem, the small unlabeled individual dataset from that subject contains personal information, which may be beneficial in classifying ECG from the given subject. For instance, it is probable that the similarity of individual normal patterns preserves over time, given that they are generated by the same underlying cardiovascular system functioning in the normal state.

It is worth noting that our problem is slightly different from the most common semisupervised learning scenario, where there is far more unlabeled data available than labeled data. In our case, the unlabeled dataset we have, i.e., the initial 5-min data (only a few hundred beats) from that subject, is of a much smaller size compared to the available labeled dataset, which is the general population dataset precollected from other subjects and annotated by cardiologists. This difference motivates us to design our approach in a slightly different way than usual.

In a typical semisupervised learning approach, a classifier learnt using labeled data is applied to predict labels of unlabeled data. The high-confidence set of unlabeled data with the predicted labels, i.e., those instances with a higher confidence measure, is combined with labeled data to learn another classifier. The core idea of semisupervised learning is therefore to use the algorithm-generated labels to replace the manual reference labels for training. We should be cautious in making use of such algorithm-generated labels, since they may not be as reliable for use due to inherent randomness and uncertainty. In the heartbeat classification problem, a fixed classifier trained on the population database may fail in classifying ECG from an unobserved subject.

The randomness of a single view motivates the proposal of *multiview-based semisupervised learning*, or *multiview learning* [11]. The basic idea of multiview learning is to boost learning performance by integrating the diversity of multiview data. These views can be obtained from distinct sources or constructed from different perspectives. The success of multiview learning is ensured by exploiting two fundamental principles: *consensus principle* and *complementary principle* [12]. The complementary principle basically states that in a multiview setting, each view of data may contain some knowledge that the other views do not have and, hence, the combination of multiple views provides a more comprehensive description of the data. The randomness and uncertainty is hence significantly reduced, compared to using only a single perspective. The consensus principle means that the models from different views should agree on their answers about the same instance.

2) Static Model and Dynamic Model: Two types of models are proposed—the *static model* and the *dynamic model*, corresponding to two distinct perspectives, namely, the *standalone perspective* and the *holistic perspective*. Both types of models are solely trained on the annotated general population training dataset, and are applied to predict heartbeat instances from the individual training dataset of the subject.

The static model focuses on the characteristics of the given heartbeat itself, as proposed in our previous study [9]. Each heartbeat is represented by a set of morphological and temporal features, characterizing the morphology of and the temporal characteristics around the heartbeat signal. A support vector machine (SVM) classifier is trained for the classification of a heartbeat instance.

The dynamic model employs a time series analysis approach, jointly modeling the underlying heartbeat status sequence and the observed temporal interval sequence. The model is intended to incorporate information from neighboring heartbeats for inference of the class of the given heartbeat signal. The details of this idea are discussed in our study [13]. More importantly, the dynamic model integrates the background rhythm information (e.g., irregular heart rhythm pattern) presented in ECG signals, which contributes to an improved inference of the beat's status.

These two types of models are complementary. The static model concentrates on the characteristics of the given heartbeat alone, representing a "standalone" perspective; the dynamic model integrates the background information for the heartbeat classification, providing a "holistic" perspective. The two types of models can be perceived as two "experts" with different expertise and aim at exploring the consensus principle among the two models.

3) Automatic Subject Adaptation Framework: In long-term ambulatory monitoring, there are usually two channels of ECG signals [14], which are simultaneously collected by placing sensors at different locations of the body. These two signals can be perceived as two observations from distinct views on the underlying cardiac activity. Additional information can be incorporated by making use of both signals. In our previous study [9], the benefits of taking into account two channels of ECG signals for heartbeat classification are shown. Therefore, the use of both channels is explored and a static model is constructed for each channel of signal.

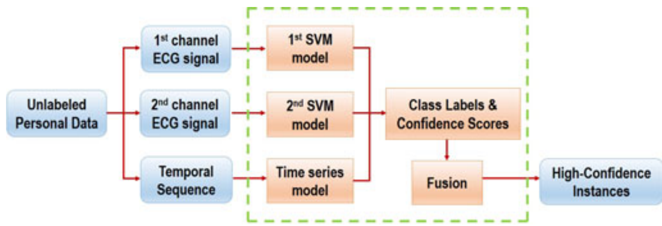

Fig. 3. Automatic adaptation mode: the framework for multiview-based semisupervised learning for automatic subject adaptation.

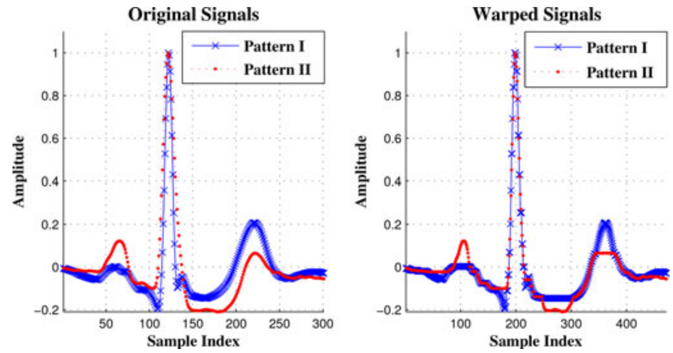

Fig. 4. DTW-based matching of the two signature patterns.

Fig. 3 illustrates the framework of the proposed multiview-based learning method for automatic subject adaptation. The proposed method explores the mutual agreement among the three models, including two static models and a dynamic model, corresponding to three distinct views. The three models are trained on the general population database, and are applied to classify the individual training dataset (i.e., initial 5-min) from the subject. Each model provides a class decision of the heartbeat instance, as well as a confidence level of that decision. All this information is fused to decide whether to preserve or discard the beat instance for the next-step model development. The final fusion is based on the consensus principle and confidence ratings among the three views. An instance is selected if all the three models make the same decision and the confidence level of each decision exceeds a certain threshold. In our experiments, the threshold was determined based on leave-one-out cross validation on the 22 training records (using 21 records for training and the remaining one record for testing, iterating through 22 records) when varying the threshold in the range of [0.4, 0.99]. The confidence level threshold was hereby empirically selected as 0.75. The resulting set of heartbeat instances are perceived as a high-confidence personal dataset, which is used in the subsequent development of individual SMs (along with their algorithm-generated labels).

The confidence level ϕ of the decision of model θ on the instance x is defined as the probability of the most probable class y^* (the winning class), i.e., $\theta(x) = P_\theta(y^*|x)$. The more model θ favors the winning class y^* over other classes, the more the classifier is confident of that decision on instance x. For a given heartbeat instance, each model generates the posterior probability estimates of all possible classes, and the confidence levels are estimated accordingly.

B. *Subject-Customized Models*

When it comes to the design of subject-customized models for a given subject, there is typically a general population dataset containing tens of thousands of heartbeats and all possible patterns, precollected from a number of subjects and annotated by cardiologists. On the other hand, there is only a small amount of high-confidence personal dataset, resulting from multiview-based learning on the subject's initial 5-min data. The proposed subject SMs consists of a GM and a SM, which represent general population knowledge and specific personal knowledge, respectively.

1) General Classification Model: The general classification model is trained mainly using the general population dataset. It is noted that the general population dataset is very heterogeneous with data typically collected from different population groups with diverse health conditions. Moreover, there may also exist variations in the electrode placement configuration. For instance, in the MIT-BIH Arrhythmia Database, the first lead signal is usually modified limb lead II, but occasionally is lead V5, which can lead to a quite different waveform morphology.

Due to the significant interperson variations and the electrode placement variations, some of the general population ECG signals are not suitable, or even misleading, for classifying ECG of a particular subject. Instead of directly using the entire population dataset for training, a training data selection method is proposed, which allows us to select a suitable training set for a given subject. A similar idea has proved to be useful in the multispectral endoscopy image classification [15], in which it is demonstrated that the learning performance can be significantly improved if a few datasets that are similar to a given test set are used for training.

The training subjects are selected, which are "similar" to the test subject, based on a similarity measure. The similarity between any two subjects is determined by calculating dynamic time warping (DTW) distance between their signature patterns. A signature pattern is derived for each training subject in the population database, as well as for the test subject, by averaging the first three consecutive normal heartbeats from the subject. Given a test subject, the similar and dissimilar training subjects are determined, by matching the signature pattern of the test subject against all signature patterns in the population dataset.

As mentioned, DTW matching is employed to measure the similarity between two signature patterns. DTW is a method proposed for matching two sequences that may vary in time or speed [16]. One well-known application of DTW is automatic speech recognition, in which it is used to handle different speaking speeds. In our problem, DTW matching is used to alleviate the difference caused by variations in heart rate, while capturing the actual difference in morphology. Fig. 4 shows the direct matching of the two patterns and the matching of their warped patterns (after DTW is performed).

The normalized DTW distance (normalized by the signal length) is used as the similarity metric between two signature

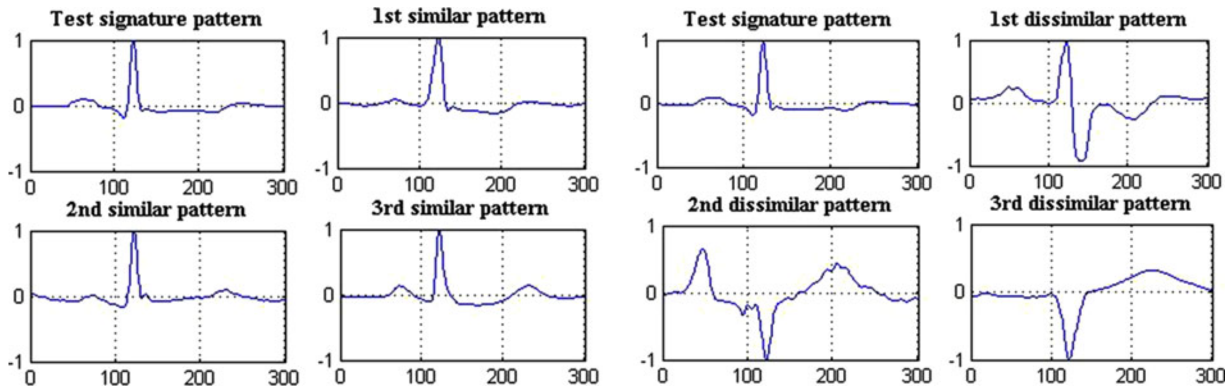

Fig. 5. Three similar patterns and three dissimilar patterns in the general population training dataset, for a given test subject (record 100). The three similar normal patterns are from records 223, 205, 101, and the three dissimilar normal patterns are from records 114, 230, 207.

patterns, and we classify a training subject as "similar" to the test subject if the normalized DTW distance between their signals does not exceed a certain threshold. In our experiments, the DTW threshold was determined by a similar method as used in selecting the confidence level threshold. The value of DTW threshold was empirically set as 0.015 (the performance was good when the threshold was in the range of $[0.012, 0.02]$). Fig. 5 shows three similar signature patterns and three dissimilar ones for a particular test signature pattern. Thus, for a given subject, the GM is trained based on the dataset of similar training subjects and the high-confidence personal dataset of the subject.

2) Specific Classification Model: The basic idea of the proposed SM is based on novelty detection. In many biomedical problems, the abnormal samples are underpresent in the database of available examples, due to the fact that the abnormal samples are difficult to capture, or expensive to obtain. The basic idea is to test for novelty against the representation of normality. If any test instance deviates significantly from the representation, it can then be assigned as an abnormal instance. Tarassenko *et al.* [17] applied a method based on novelty detection for the identification of masses (breast cancer) in mammograms.

In our problem, it is typical that the available instances are mostly normal, while abnormal samples are underpresent in the initial 5-min data of the subject. It motivates us to develop a representation of individual normality information, termed as the *one-class model*, which aims at providing an intraperson classification. One issue of the one-class model is that the classification performance is not robust enough and usually quite sensitive to the choices of representation and parameters, mainly caused by the absence of abnormal information. Intuitively, with little (if any) prior information of abnormal classes, it is difficult for classification model to determine how "far" from the normal class or which "part" of the feature space should be identified as abnormal.

To solve this issue, we consider making use of some prior knowledge of abnormal classes contained in the available abnormal samples of the training subjects. When learning the SM for a given subject, a mechanism is developed to automatically select a few representative abnormal samples from those similar training subjects, covering a range of common arrhythmia types—supraventricular ectopic beats (e.g., atrial premature contraction beats) and ventricular ectopic beats (e.g., premature ventricular contraction beats). The number of selected abnormal samples is typically in range of 50–150, depending on total number of abnormal samples contained in the corresponding similar training subjects. These samples can provide an informative representation of the abnormal class distribution in the test subject. The SM is hereby trained using a high-confidence personal dataset and a few representative abnormal samples, making it more robust in discrimination between normal and abnormal classes.

3) Model Fusion: The GM, learnt on the relevant population knowledge, provides an interperson perspective for classification. The SM, designed to focus on the subject specific knowledge, offers an intraperson perspective for classification. The two types of models are combined for the final heartbeat classification.

For training of both GM and SM, the basic framework described in [9] is adopted. The heartbeat signals are represented by a set of morphological and temporal features. An SVM with radial basis function kernel is used as the classification method. Both channels of signals are used. Hence, one GM and one SM are trained for each channel. The learnt models are applied respectively to classify the first lead and the second lead signal of the test heartbeat.

Given the signal $\mathbf{x}_l$ from lth lead ($l = 1, 2$), let $P_g^l(y = i|\mathbf{x}_l)$ be the posterior probability estimate from the lth GM, for the K-class problem. The probability estimates of the GMs are computed based on a Bayesian product approach

$$P_g(y = i \,|\, \{\mathbf{x}_1, \mathbf{x}_2\}) = \frac{\prod_{l=1}^{2} P_g^l(y = i \,|\, \mathbf{x}_l)}{\sum_{j=1}^{K} \prod_{l=1}^{2} P_g^l(y = j \,|\, \mathbf{x}_l)}. \quad (1)$$

Similarly, the probability estimates of the SMs are calculated as follows:

$$P_s(y = i \,|\, \{\mathbf{x}_1, \mathbf{x}_2\}) = \frac{\prod_{l=1}^{2} P_s^l(y = i \,|\, \mathbf{x}_l)}{\sum_{j=1}^{K} \prod_{l=1}^{2} P_s^l(y = j \,|\, \mathbf{x}_l)}. \quad (2)$$

TABLE I
RECORD-BY-RECORD PERFORMANCE[a] COMPARISON OF GENERAL MODEL, SPECIFIC MODEL AND FUSION OF MODELS

Rec.	*GM*[b]	*SM*[c]	*FM*[d]	Rec.	*GM*	*SM*	*FM*
100	99.79	99.16	99.63	212	99.87	99.75	99.96
103	99.89	99.34	99.89	213	89.85	85.55	87.51
105	98.98	97.92	98.98	214	99.84	99.04	98.88
111	99.77	99.33	99.61	219	97.24	99.04	99.21
113	99.81	99.67	99.67	221	99.85	99.25	100
117	99.84	99.92	99.92	222	92.25	89.98	90.45
121	99.94	99.81	99.94	228	99.41	99.82	99.77
123	99.85	99.76	99.76	231	99.84	100	100
200	96.31	96.03	96.45	232	21.41	94.34	80.88
202	83	96.95	97.75	233	98.13	96.27	96.48
210	97.5	97.82	97.82	234	97.82	97.2	97.82

[a] Overall classification accuracy in percentage, i.e., number of correctly classified beats * 100/total number of beats. [b] Performance of GM. [c] Performance of SM. [d] Performance of FM.

TABLE II
RECORD-BY-RECORD PERFORMANCE COMPARISON OF FIXED MODE, AUTOMATIC ADAPTATION MODE AND EXPERT INTERVENTION MODE

Rec.	*Fixed*[a]	*Auto.*[b]	*Expert*[c]	Rec.	*Fixed*	*Auto.*	*Expert*
100	99.82	99.63	100	212	60.14	99.96	100
103	99.9	99.89	99.88	213	88.46	87.51	96.48
105	99.14	98.98	99.21	214	36.42	98.88	99.84
111	94.77	99.61	99.77	219	95.64	99.21	99.38
113	99.67	99.67	99.87	221	99.26	100	100
117	25.49	99.92	99.92	222	90.91	90.45	91.16
121	99.94	99.94	99.94	228	99.32	99.77	99.77
123	99.67	99.76	99.76	231	99.11	100	100
200	96.34	96.45	98.11	232	24.36	80.88	97.98
202	83.8	97.75	97.91	233	96.39	96.48	99.53
210	97.66	97.82	98.05	234	96.14	97.82	97.82

[a]Performance of fixed mode. [b] Performance of automatic adaptation mode. [c] Performance of expert intervention mode.

TABLE III
CONFUSION MATRIX OF THE SUBJECT-CUSTOMIZED MODELS USING THE AUTOMATIC ADAPTATION MODE

		Predicted Label				
		n	s	v	f	q
Reference	N	36802	51	26	32	0
	S	597	983	28	0	0
	V	167	49	2460	10	0
	F	247	0	15	26	0
	Q	6	0	1	0	0

TABLE IV
CONFUSION MATRIX OF THE SUBJECT-CUSTOMIZED MODELS USING THE EXPERT INTERVENTION MODE

		Predicted Label				
		n	s	v	f	q
Reference	N	36881	9	16	5	0
	S	354	1230	24	0	0
	V	45	2	2607	32	0
	F	34	0	17	237	0
	Q	6	0	1	0	0

The final classwise probability estimates are obtained by combining outputs from the two categories of models

$$P_f(y = i \,|\, \{\mathbf{x}_1, \mathbf{x}_2\}) = \frac{P_g(y = i) P_s(y = i)}{\sum_{j=1}^{K} P_g(y = j) P_s(y = j)}. \quad (3)$$

The selected class k is determined as corresponding to the highest final probability estimate in M classes, given by

$$k = \underset{i}{\operatorname{argmax}}\, P_f(y = i \,|\, \{\mathbf{x}_1, \mathbf{x}_2\}). \quad (4)$$

III. RESULTS

A. Experimental Setup

The benchmark MIT-BIH Arrhythmia Database [18] is used for validation of the proposed algorithms. A subject-based evaluation strategy is employed. The dataset was divided into training and test datasets, each consisting of 22 recordings and corresponding to ECG signals from different subjects. The dataset division is consistent with prior investigations [3]–[6].

In the current setting, the set of 22 training recordings is regarded as the general population dataset for training. When it comes to test on a particular test recording, the initial 5-min unlabeled data from the subject is used for subject adaptation, as described in Section II-A. The GM and the SM for the subject are sequentially trained, as discussed in Section II-B. The trained models are applied separately to classify the remaining 25-min data of the subject. For each heartbeat signal, the outputs of the two models are combined to make the final decision.

B. GM Versus SM

In Table I, the overall results are presented, including the classification accuracy of the GM, the SM as well as the fusion of two models (FM), across the 22 test subjects. It is observed that the GM works slightly better than or as well as the SM for most records. However, the SM outperforms the GM on a few records (e.g., records 202, 232). This improvement occurs typically when there exists significant interindividual variations, which means that beat patterns of the test subject are quite different from the same class in the general training database, in terms of morphological and/or temporal characteristics. In other words, similar patterns have not been observed in the general training dataset. In these cases, giving an emphasis to the observed individual patterns seems to be beneficial. In general, the fusion of the two categories of models yields more robust classification performance than using only a single type of model.

C. Automatic Adaptation Mode Versus Expert Intervention Mode

We next study the performance improvement obtained by the proposed subject-customized models. The two possible modes are explored for training subject-customized models. The first mode is the expert intervention mode, in which the full 5-min initial data from the test subject, along with the expert-provided reference labels, are utilized for training the subject-customized

TABLE V
PERFORMANCE COMPARISON OF THE PROPOSED METHOD AND THE MAJOR REFERENCE WORKS

	Class "V"				Class "S"				
Method	*Acc*[a]	*Se*[b]	*Sp*[c]	*Ppr*[d]	*Acc*	*Se*	*Sp*	*Ppr*	Mode
Ye [14]	95.7	81.5	96.7	63.1	96.3	60.8	97.7	52.3	Fixed Mode
Hu [2]	94.8	78.9	96.8	75.8	N/A	N/A	N/A	N/A	Expert Intervention Mode
Chazal [3]	96.4	77.5	98.9	90.6	92.4	76.4	93.2	38.7	Expert Intervention Mode
Jiang [4]	98.8	94.3	99.4	95.8	97.5	74.9	98.8	78.8	Expert Intervention Mode
Tucker [5]	97.9	90.3	98.8	92.2	96.1	81.8	98.5	63.4	Expert Intervention Mode
Method I	99.7	97.1	99.9	98.5	99.1	76.5	99.9	99.1	Expert Intervention Mode
Method II	99.4	91.8	99.9	98.0	98.3	61.4	99.8	90.7	Automatic Adaptation Mode

[a]Accuracy of any given class = *(TP+TN) / (TP+TN+FP+FN)*, where *TP*, *TN*, *FP*, *FN* represents number of true positive, true negative, false positive, and false negative of the given class. [b] Sensitivity = *TP / (TP+FN)* [c] Specificity = *TN / (TN+FP)* [d] Positive prediction rate = *TP / (TP+FP)*.

models. It allows us to make a fair comparison with the related studies, which also adopted expert intervention mode. The second mode is the automatic intervention mode, in which the high-confidence individual dataset with the algorithm generated labels is used for model training. It makes it possible to construct the subject-customized models, without needing cumbersome manual labeling efforts. A comparison is made among the subject-customized models learnt in the two different modes and the fixed model built in the fixed mode, in which the model is trained on the population database and applied without adaptation.

Table II presents the performance comparison among the three modes, in terms of the overall classification accuracy on each of 22 test records. As we can observe, a significant performance improvement has been obtained for quite a few subjects where the general fixed model did not perform well enough or failed. It indicates that the proposed subject-customized models are effective in addressing interperson variations, leading to more robust heartbeat classification performance. We also notice that the subject-customized models constructed in the automatic adaptation mode achieve comparable performance to the ones built in the expert intervention mode. It suggests the usefulness of the proposed multiview-based semisupervised learning for subject adaptation, beneficial from building person-customized models without involving human efforts. The confusion matrices of two modes are presented in Tables III and IV, which contain overall performance in automatic adaptation mode and expert intervention mode, respectively.

The performance of proposed method is also compared in Table V against the major methods presented in the literature. As suggested by the ANSI/AAMI EC57 standard [8], we focus on evaluating the classification performance of the two main abnormal classes, i.e., the class "S" and the class "V," in terms of the class accuracy, sensitivity, specificity, and positive prediction rate. First, it can be seen that the proposed subject-customized model exhibits significant improvement in performance over earlier methods, while using the reference labels of initial 5-min data of the test subject. It is worth noting that using the reference labels renders the algorithms less practical for real-world applications. The multiview-based learning method allows us to fully automate the subject adaptation and the construction of the proposed subject-customized models, without resorting to cumbersome manual annotation. More importantly, the customized models with automatic adaptation still achieve outstanding classification performance compared to the earlier methods using reference labels.

In addition, the performance of automatic adaptation mode is also compared with the fixed mode. The performance improvement is remarkable (except for the sensitivity of class "S"), indicating the benefits of performing subject adaptation. The similar sensitivity of class "S" in two modes is possibly due to that the subset resulting from multiview learning typically contains only the individual baseline information. Our understanding is that the abnormal samples are relatively insufficient in the general training set, especially for class "S," making recognition of abnormal information less reliable during subject adaptation. Abnormal samples (if any) may not be captured and, hence, excluded from model adaptation due to low confidence levels. Further improvement on class "S" is expected when it is possible to access a more comprehensive general training dataset.

Finally, the algorithm runtime of automatic adaptation mode and expert intervention mode was studied. We performed the experiments in MATLAB(R2012b) on a laptop (configuration: CPU—Intel Core i5 2.40 GHz, RAM—4G). The models were trained on initial 5 min and tested on the remaining 25 min for each 30 min long record. Average runtime[1] of model training and testing are 22.9 and 3.4 s, respectively, in automatic adaptation mode, compared to 18.2 and 3.4 s in expert adaptation mode. It took more time in the automatic adaptation mode since there is an additional step (i.e., multiview learning) before model training. For the expert intervention mode, we assume the manual labels for initial 5 min are ready and can be directly used for model training. However, in a real-world setup, there will be additional time for an expert to provide manual labels for the initial 5-min data.

IV. CONCLUSION

In this study, a framework of subject-customized models is proposed for heartbeat classification, in order to address the

[1]The 30-min data recordings were processed in an offline batch mode. Besides, the preprocessing, segmentation, and feature extraction were performed beforehand and excluded from counting runtime. The SVM is implemented using LIBSVM package: http://www.csie.ntu.edu.tw/ cjlin/libsvm/

challenge of interperson variations in ECG signals. The proposed customized models consist of a GM and a SM, which are designed to focus on different perspectives and complement each other, so as to achieve more robust performance in heartbeat classification. Another important improvement over the earlier studies is the automatic construction of subject-customized models, instead of requiring manual effort, making the proposed algorithms appealing for real-world applications.

In the future, it may be beneficial to investigate more sophisticated fusion approach to merge the information of the two categories of models. Besides interperson variations, there exists the intraperson variations in ECG signals for the same subject in the long-term continuous monitoring, when the subject experiences the change in heart conditions or physical states. It is of interest to explore how to circumvent intraperson variations, in order to further improve personalized ECG signal analysis in ambulatory continuous monitoring.

REFERENCES

[1] K. Robert and E. C. Colleen, *Basis and Treatment of Cardiac Arrhythmias*, 1st ed. New York, NY, USA: Springer, 2006.

[2] Y. H. Hu, S. Palreddy, and W. J. Tompkins, “A patient-adaptable ECG beat classifier using a mixture of experts approach,” *IEEE Trans. Biomed. Eng.*, vol. 44, no. 9, pp. 891–900, Sep. 1997.

[3] P. De Chazal and R. B. 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.

[4] W. Jiang and S. G. Kong, “Block-based neural networks for personalized ECG signal classification,” *IEEE Trans. Neural Netw.*, vol. 18, no. 6, pp. 1750–1761, Nov. 2007.

[5] T. Ince, S. Kiranyaz, and M. Gabbouj, “A generic and robust systems for automated patient-specific classification of ECG signals,” *IEEE Trans. Biomed. Eng.*, vol. 56, no. 5, pp. 1415–1426, May 2009.

[6] M. Llamedo and J. P. Martinez, “An automatic patient-adapted ECG heartbeat classifier allowing expert assistance,” *IEEE Trans. Biomed. Eng.*, vol. 59, no. 8, pp. 2312–2320, Aug. 2012.

[7] C. Ye, B. V. K. Vijaya Kumar, and M. T. Coimbra, “Combining general multi-class and specific two-class classifiers for improved customized ECG heartbeat classification,” in *Proc. 21st Int. Conf. Pattern Recog.*, 2012, pp. 2428–2431.

[8] *Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*, ANSI/AAMI EC57, 1998.

[9] C. Ye, B. V. K. Vijaya Kumar, and M. T. Coimbra, “Heartbeat classification using morphological and dynamic features of ECG signals,” *IEEE Trans. Biomed. Eng.*, vol. 59, no. 10, pp. 2930–2941, Oct. 2012.

[10] O. Chapelle, B. Sholkopf, and A. Zien, *Semi-Supervised Learning*. New York, NY, USA: MIT Press, 2006.

[11] A. Blum and T. Mitchell, “Combining label and unlabeled data with co-training,” in *Proc. 11th Annu. Conf. Comput. Learn. Theory*, 1998, pp. 92–100.

[12] C. Xu, D. C. Tao, and C. Xu, *A Survey on Multi-View Learning*, 2013.

[13] C. Ye, “Advanced heartbeat classification models for reliable electrocardiogram analysis in ambulatory health monitoring,” Ph.D. dissertation, Dept. Elect. Comput. Eng., Carnegie Mellon Univ., Pittsburgh, PA, USA, 2013.

[14] P. J. Podrid and P. R. Kowey, *Cardiac Arrhythmia: Mechanisms, Diagnosis, and Management*. Baltimore, MD, USA: Williams & Wilkins, 2001.

[15] C. V. Dinh, M. Loog, R. Leitner, O. Rajadell, and R. P. W. Duin, “Training data selection for cancer detection in multispectral endoscopy images,” in *Proc. 21st Int. Conf. Pattern Recog.*, 2012, pp. 161–164.

[16] H. Sakoe and S. Chiba, “Dynamic programming algorithm optimization for spoken word recognition,” *IEEE Trans. Acoust., Speech Signal Process.*, vol. ASSP-26, no. 1, pp. 43–49, Feb. 1978.

[17] L. Tarassenko, P. Hayton, N. Cerneaz, and M. Brady, “Novelty detection for identification of masses in mammograms,” in *Proc. 4th Int. Conf. Artif. Neural Netw.*, 1995, pp. 442–447.

[18] MIT-BIH arrhythmia database. [Online]. Available: Physionet http://www.physionet.org/physiobank/database/mitdb/

[19] S. Dasgupta, M. L. Littman, and D. McAllester, “PAC generalization bounds for co-training,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2002, pp. 375–382.

Authors’ photographs and biographies not available at the time of publication.