Optimal Length-Constrained Segmentation and Subject-Adaptive Learning for Real-time Arrhythmia Detection

Kha Vo, Eryk Dutkiewicz

Global Big Data Technologies Centre, School of Electrical and Data Engineering
University of Technology Sydney, Australia
anhkha.vo@student.uts.edu.au,eryk.dutkiewicz@uts.edu.au

ABSTRACT

An algorithm of data segmentation with length constraints for each segment is presented and applied in the context of arrhythmia detection. The additivity property of the cost function for each segment yields the induction proof of the exact global optimal solution. The experiments were conducted on the MIT-BIH arrhythmia dataset with the heartbeat categories recommended by the ANSI/AAMI EC57:1998 standard. The heartbeat classification task is enhanced by an adaptive learning scheme. Incremental support vector machine is used to integrate a small number of expert-annotated samples specific to the subject into the existing classifier previously learned from the dataset. The proposed segmentation scheme obtains the sensitivity of 99.89% and the positive predictivity of 99.83%. The classification sensitivities of ventricular and supraventricular detection are significantly boosted from 85.9% and 83.5% (subject-unadaptive) to 97.7% and 93.2% (subject-adaptive), respectively. Similarly the predictivities increase from 94.8% to 99.3% (ventricular), and from 67.7% to 88.0% (supraventricular) when plugging in the adaptive learning method. The signal processing framework is conducted in a simulated real-time model. As compared to the previously reported studies we achieve a competitive performance in terms of all assessment measures.

KEYWORDS

Data segmentation, heartbeat classification, adaptive learning, arrhythmia detection

1 INTRODUCTION

Correctly detecting heartbeat abnormalities is of great importance to provide promptly therapy and prevent life-threatening problems for patients with cardiac diseases. The detection of arrhythmias, or simply known as abnormal heart rhythms, is a crucial task in electrocardiogram (ECG) monitoring. Arrhythmias are presented in two major types. The first includes long-term rhythmic arrhythmias, i.e., bradycardia, tachycardia. The other is represented by transient morphological arrhythmias, i.e., ventricular ectopic beats (VEB), supraventricular ectopic beats (SVEB), as recommended by the Association for the Advancement of Medical Instrumentation (AAMI) [1]. This study focuses on the latter type and proposes a unified framework of ECG segmentation and classification between VEBs, SVEBs and non-VEBs, non-SVEBs.

Researchers have put their efforts to improve these following four major areas: preprocessing techniques (Bayesian [2], wavelet transform [3, 4], digital filters [5, 6]), feature extraction methods (RR-intervals [7], QRS complexes [8], ECG morphology [5, 6]), heartbeat segmentation algorithms (based on QRS complexes, R-peaks [9, 10], or P-waves, T-waves [11, 12]), and learning algorithms (linear discriminant analysis (LDA) [5, 6, 13, 14], support vector machine (SVM) [15, 16], artificial neural networks (ANN) [17, 18], reservoir computing with logistic regression (RCLR) [19]). Among those areas, this paper proposes methods on the segmentation and learning stages. For the other two, we select appropriate and inexpensive ECG preprocessing techniques and ECG features for the sake of real-time performance.

With regards to the segmentation problem, the methods mentioned in the literature are all based on the ECG morphology (i.e., the PQRST-wave features). One setback for these methods is that we must specify a specific ECG lead to process, since those mentioned morphologies are different among ECG leads. This paper aims to propose a general method not dependent on any special ECG feature but only the provided data itself, hence can be applicable to any ECG lead, or more generally, to any data type. One benefit of this approach is the boosted accuracy resulting from combination of the individual segmentation results from all possible ECG leads. The only information that can be exploited are the length constraints of each heartbeat segment (i.e, normal human resting heart rate varies between 0.6 and 1.0 seconds). We are inspired by the fact that normal heartbeats of a subject measured at a specific experimental setting share a similar pattern to a certain extent. We formulate that problem into an optimization problem with segment length constraints, then solve it using a novel dynamic programming algorithm. Our approach is developed from the method of curve fitting by line segments [20]. However the mentioned work did not impose length constraints on segments. Therefore it cannot be applicable to real-time ECG systems as the time needed for searching for the optimal solution far exceeds the allowable time for real-time processing.

Another major difficulty imposed on heartbeat classification is the cardiac specificity of individual subjects. The trained classifier has no prior knowledge of the validation subject's pattern, hence heavily affects the system's accuracy. Previous studies also targeted this problem, i.e., adaptive LDA [6, 14], generalized regression neural networks [21], or switching Kalman filter [22]. For the real-time mode, however, one would consider several issues that have not been satisfactorily resolved in the mentioned studies, i.e., updating the classifier in a restricted amount of time, or discarding undesirable samples to avoid data accumulation. In this paper we propose a real-time subject-adaptive learning scheme using the incremental support vector machine algorithm [23].

In summary, our main contributions are:

· Propose a data segmentation algorithm with segment-length constraints. The exact global optimal solution is guaranteed by a dynamic program. The algorithm can perform in real-time, does not require the input number of segments, and is applicable to any ECG lead or any data type.

· Propose a real-time learning method to have the heartbeat classifier get adapted to a new subject. A small number of expert-annotated samples will enhance the existing SVM classifier using the Incremental SVM toolbox.

The rest of this paper is organized as follows: Section 2 presents the optimal segmentation method. The system's framework including data preprocessing, dataset description, feature extraction and adaptive learning subsections are described in Section 3. Section 4 presents the experimental settings and performance assessment results. Finally the conclusions are given in Section 5.

2 OPTIMAL LENGTH-CONSTRAINT DATASEGMENTATION

The 1-D data signal of ordered individual data points d_i is denoted by $\boldsymbol{d}_{N_1:N_2} = \{d_i, N_1 \leq i \leq N_2\}$. We define a data segment $\boldsymbol{s}_i$ as any subset of $\boldsymbol{d}_{N_1:N_2}$ which has consecutive data points d_i. A set of M non-overlapping contiguous segments constructing the data $\boldsymbol{d}_{N_1:N_2}$ is called a *segmentation* of $\boldsymbol{d}_{N_1:N_2}$, denoted as

$$S = \{\boldsymbol{s}_i | \bigcup_{i=1}^{M} \boldsymbol{s}_i = \boldsymbol{d}_{N_1:N_2}, \boldsymbol{s}_i \cap_{i \neq i^*} \boldsymbol{s}_i = \varnothing\}. \quad (1)$$

The set of all possible segmentation S of $\boldsymbol{d}_{N_1:N_2}$ is denoted as $\boldsymbol{S}_{N_1:N_2}$. Given a full signal $\boldsymbol{d}$ of N points (i.e., $\boldsymbol{d} = \boldsymbol{d}_{1:N}$) and the length constraints on each data segment to be segmented (i.e., $l_{\min} \leq |\boldsymbol{s}_i| \leq l_{\max}$), the segmentation problem of $\boldsymbol{d}$ can be formulated to the following constrained optimization problem

$$\underset{S \in \boldsymbol{S}_{1:N}}{\text{minimize}} \quad W(S) = \sum_{\boldsymbol{s}_i \in S} ||r_{\boldsymbol{\mu}}(\boldsymbol{s}_i) - \boldsymbol{\mu}||_2^2 \quad (2)$$
$$\text{subject to} \quad l_{\min} \leq |\boldsymbol{s}_i| \leq l_{\max},$$

where $r_{\boldsymbol{\mu}}(\boldsymbol{x})$ is the interpolated vector of $\boldsymbol{x}$ with length equaling to that of $\boldsymbol{\mu}$. Here $|\boldsymbol{x}|$ denotes the length of $\boldsymbol{x}$, $||\boldsymbol{x}||_2^2$ denotes the ℓ_2-norm of $\boldsymbol{x}$ in R^N, and $\boldsymbol{\mu}$ represents as the template model for $\boldsymbol{s}_i$ and only dependent on the change

point locations. In this paper's context, $\boldsymbol{\mu}$ is obtained by averaging the selected typical normal heartbeat samples from various learning subjects.

The number of all possible partitions S exponentially increases with respect to the number of data points N, leading to the intractability of the optimization problem (2) when N is large. On a different perspective, we can consider W as the sum of M additive individual cost function F as

$$W = \sum_{i=1}^{M} F(\boldsymbol{s}_i), \quad (3)$$

where $F(\boldsymbol{x}) = ||r_{\boldsymbol{\mu}}(\boldsymbol{x}) - \boldsymbol{\mu}||_2^2$. Using this denotation it is easily observed that the original objective function W in problem (2) possesses the property of *additivity*. By this important property, we are able to propose the dynamic programming method to find the exact global optimal of the problem (2) given the length constraints of each block.

Lemma 1. Given two disjoint segments A and B, i.e., $A \cap B = \varnothing$, then $W(A \cup B) = W(A) + W(B)$.

Proof. The proof implies in the additivity property of the objective function W as: $W(A \cup B) = \sum_{\boldsymbol{s}_i \in A \cup B} F(\boldsymbol{s}_i) = \sum_{\boldsymbol{s}_i \in A} F(\boldsymbol{s}_i) + \sum_{\boldsymbol{s}_i \in B} F(\boldsymbol{s}_i) - \sum_{\boldsymbol{s}_i \in A \cap B} F(\boldsymbol{s}_i) = W(A) + W(B)$.

Theorem 1. Let $S^* = \{\boldsymbol{s}_i^*, 1 \leq i \leq M^*\}$ denoted the global optimum solution of problem (2), and $S^{*,m} = \{\boldsymbol{s}_i | \boldsymbol{s}_i \in S^*, 1 \leq i \leq m \leq M^*\}$ denoted the set of the first m segments of S^* that covers the first n points of the original signal (i.e., $\boldsymbol{s}_i = \boldsymbol{d}_{1:n}$). Then $S^{*,m}$ is the global optimal solution of the following optimization problem

$$\underset{S \in \boldsymbol{S}_{1:n}}{\text{minimize}} \quad W(S) = \sum_{\boldsymbol{s}_i \in S} ||r_{\boldsymbol{\mu}}(\boldsymbol{s}_i) - \boldsymbol{\mu}||_2^2 \quad (4)$$
$$\text{subject to} \quad l_{\min} \leq |\boldsymbol{s}_i| \leq l_{\max}.$$

Proof. Suppose there exists an optimal solution of problem (4) other than $S^{*,m}$, denoted as S^t. This leads to $W(S^t) < W(S^{*,m})$. Let $\bar{S} = S^* \setminus S^{*,m}$. We have $\min_{S \in \boldsymbol{S}_{1:N}} W(S) = W(S^*) = W(S^{*,m} \cup \bar{S}) = W(S^{*,m}) + W(\bar{S}) > W(S^t) + W(\bar{S}) = W(S^t \cup \bar{S})$.

Since $S^t \cup \bar{S} \in \boldsymbol{S}_{1:N}$ and $S^t \cap \bar{S} = \varnothing$, $S^t \cup \bar{S}$ becomes the optimal solution of problem (2) rather than S^*. This contradiction proves that S^t never exists, therefore $S^{*,m}$ is the optimal solution of problem (4).

Theorem 2. The Algorithm 1 (procedure WarmUp-Phase) guarantees to find the exact optimal m-segment segmentation, $S^{*,m}$, of n first data samples $\boldsymbol{d}_{1:n}$, for some $n \in [2l_{\min}, 2l_{\max}]$.

Proof. We exploit Theorem 1 in the sense that we must initially search for the first two segments of the optimal solution, $S^{*,2}$ (corresponding to the procedure WarmUp-Phase). Taken the segment length constraints into account, we see that the last index of the first segment, denoted as $\text{Idx}(\boldsymbol{s}_1)$, lies somewhere between the interval $[l_{\min}, l_{\min} + \Delta b]$ (where $\Delta l = l_{\max} - l_{\min}$). $\text{Idx}(\boldsymbol{s}_2)$ therefore falls in $[\text{Idx}(\boldsymbol{s}_1) + l_{\min}, \text{Idx}(\boldsymbol{s}_1) + l_{\min} + \Delta b]$. As a result $\text{Idx}(\boldsymbol{s}_2)$ can take values from $[\min\{\text{Idx}(\boldsymbol{s}_1)\} + l_{\min}, \max\{\text{Idx}(\boldsymbol{s}_1)\} + l_{\min} + \Delta b] = [2l_{\min}, 2l_{\max}]$. The procedure WarmUpPhase will find the

2

Algorithm 1 Optimal Segmentation Algorithm
    Input: 1-D data $\boldsymbol{d}_{1:N}$, segment constraints $l_{\min}$, $l_{\max}$.
        Output: Optimal segmentation S^*.
1: **procedure** WarmUpPhase
2:    $i \leftarrow 2l_{\min} - 1$
3:    **while** $i < 2l_{\max}$ **do**
4:       $i \leftarrow i + 1$
5:       $W^*(i) = \min\{F(\boldsymbol{d}_{1:j}) + F(\boldsymbol{d}_{j+1:i})\}_{j=\max\{l_{\min}, i-l_{\max}\}}^{\min\{i-l_{\min}, l_{\max}\}}$
6:       Save value of j where $W^*(i)$ occurs
7:       **if** $i \geq kl_{\min}$ ($k \geq 3$ and possibly largest) **then**
8:          $W^{\mathrm{t}} = \min\{W^*(j) + F(\boldsymbol{d}_{j+1:i})\}_{j=(k-1)l_{\min}}^{i-l_{\min}}$
9:          $W^*(i) \leftarrow \min\{W^*(i), W^{\mathrm{t}}\}$
10:      **end if**
11: **end while**
12: **return** a

Algorithm 2 Optimal Segmentation Algorithm
    Input: 1-D data $\boldsymbol{d}_{1:N}$, segment constraints $l_{\min}$, $l_{\max}$.
        Output: Optimal segmentation S^*.
1: **procedure** MainPhase
2:    $i \leftarrow 2l_{\max} + 1$
3:    **while** $i \leq N$ and $i - l_{\min} > 2l_{\min}$ **do**
4:       $W^*(i) \leftarrow \min\{W^*(j) + F(\boldsymbol{d}_{j+1:i})\}_{j=\max\{i-l_{\max}, 2l_{\min}\}}^{i-l_{\min}}$
5:       $i \leftarrow i + 1$
6:    **end while**
7: **return** a

Figure 1: The diagram of our proposed ECG processing framework.

exact value of $\mathrm{Idx}(\boldsymbol{s}_1)$ for each different value of $\mathrm{Idx}(\boldsymbol{s}_2)$. As explained, the search for all possible values of $\mathrm{Idx}(\boldsymbol{s}_2)$ must be conducted in the interval $[2l_{\min}, 2l_{\max}]$, as indexed by the iterative variable i in Algorithm 1. For each given i, we search for the smallest segmentation cost value $W^*(i)$ which is the sum of the first two segments, $\boldsymbol{s}_1 = \boldsymbol{d}_{1:j}$ and $\boldsymbol{s}_2 = \boldsymbol{d}_{j+1:i}$. The second iterative variable j indexes all possible values of $\mathrm{Idx}(\boldsymbol{s}_1)$. The smallest value j can take is $\max\{l_{\min}, i - l_{\max}\}$ to not violate the length constraint of $\boldsymbol{s}_2$, and the largest value of j is $\min\{i - l_{\min}, l_{\max}\}$ to not violate the length constraint of $\boldsymbol{s}_1$.

The above search for the optimal cost of the first two segments inevitably covers the possibilities of more than two segments in the search range. If $2l_{\max} > kl_{\min}$ (where $k \geq 3$), extra segmentation will occur as there can be $3, \ldots, k$ segments for the given iteration i. For $k = 3$, we need to select between the optimal two-segment cost, $W^*(i)$, and the optimal three-segment cost, W^{t}. The resulting optimal number of segments can be 2 or 3, and we denote this value as m^*. For $k > 3$, this recursive method always compare between the m^*-segment and the $(m^* + 1)$-segment schemes. The algorithm 1 therefore obtains the exact optimal segmentation of $\boldsymbol{d}_{1:n}$ for some $n \in [2l_{\min}, 2l_{\max}]$.

Theorem 3. The Algorithm 2 (procedure MainPhase) guarantees to find the exact optimal solution, S^*, of the input data $\boldsymbol{d}_{1:N}$.

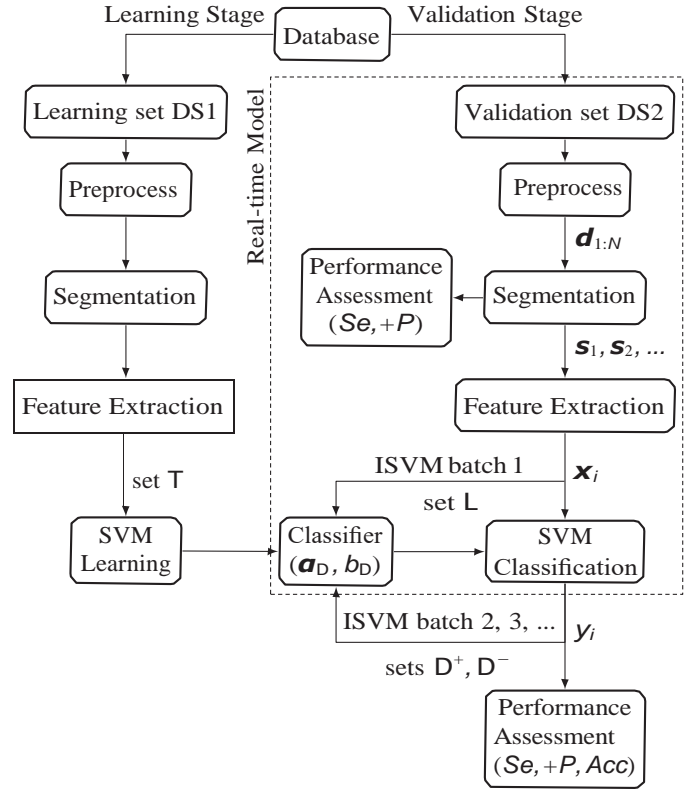**Proof.** Suppose $S^{*,m}$ is obtained, then by Theorem 1 we can always obtain $S^{*,m+1}$ by conducting the search for the minimum including the value of $W(S^{*,m-1})$ $+ F(\boldsymbol{d}_{\mathrm{Ind}(\boldsymbol{s}_{m-1})+1:\mathrm{Ind}(\boldsymbol{s}_m^*)})$, since $\boldsymbol{d}_{\mathrm{Ind}(\boldsymbol{s}_{m-1}^*)+1:\mathrm{Ind}(\boldsymbol{s}^*_m)} = \boldsymbol{s}_m^*$. Line 4 in Algorithm 2 performs this thorough search and guarantees that $\boldsymbol{s}_{\tilde{m}}^*$ will be included in the search pool (which by definition will generate the optimal cost). Recall that M^* is denoted as the number of segments in S^*, Algorithm 2 will successfully find S^* when i indexes the last sample (d_N), given that S^{*,M^*-1} has already achieved. Similar process happens for S^{*,M^*-2} (i.e., when $i = \mathrm{Ind}(\boldsymbol{s}^*_{M^*-1})$ we will achieve S^{*,M^*-1}, given that S^{*,M^*-2} is obtained). Finally, by Theorem 2 we already have the optimal for the first m segments, hence Theorem 3 holds.

The proposed constrained segmentation algorithm works in time $O(n\Delta b)$. Comparing to the other methods [20, 24] run in $O(n^2)$, our algorithm has the considerably lower complexity. The benefit of this point is discussed in more detail in Section 4.

3 FRAMEWORK
3.1 ECG Preprocessing
Various ECG denoising methods had been proposed depending intrinsically on the final objective of each study. Studies focusing on ECG feature extraction or segmentation require

3

a more sophisticated preprocessing technique (i.e., multiadaptive bionic wavelet transform [4], adaptive neural network [25]) as one would preserve the signal morphology. A specific ECG lead must be specified for these methods. On the other hand research aiming on pattern classification generally deal with noise removal (i.e., Bayesian [2], wavelet transform [3], digital filters [5, 6]) and are more prone to ECG distortion. In our work we adopt two selected techniques for the sake of real-time computational simplicity while efficiently removing artifacts. First, the raw signal is decomposed using Daubechies wavelet (db6) [3] with the 1st, 2nd level (corresponding to the high frequency band) and the 9th level (corresponding baseline wander) discarded. The reconstructed signal then is passed through an adaptive tunable notch finite-impulse response filter [26] to suppress power line interference and muscle contraction noise.

3.2 Heartbeat Categories

The experiments were conducted on the Massachusetts Institute of Technology and Beth Israel Hospital arrhythmia database (MITBIH) [27], consisting of 109492 labeled heartbeats from 46 patients. Recommended by the ANSI/AAMI EC57:1998 standard [1], fifteen beat types of this database are grouped into 5 classes: normal beat (N), ventricular ectopic beat (V), supraventricular ectopic beat (S), fusion of normal and ventricular (F), and unknown (Q), as presented in Table 1. Two classification schemes are to be conducted: scheme A for class V (or VEB) against non-VEB, and scheme B for class S (or SVEB) against non-SVEB. The learning and validation set configuration is described in Section 4.1.

Table 1: Five Hearbeat Categories Recommended by the AAMI Standard and the Corresponding Number of Samples Used in the Validation Stage. The class labels (1, -1) of two classification schemes are given in the last two columns.

AAMI Class	MITBIH Annotation	Total beats	Validation beats	Scheme A	Scheme B
N	N, L, R, e, j	90126	44259	-1	-1
V	V, E, !	7235	3221	1	-1
S	A, a, J, S, x	2779	1837	-1	1
F	F	803	388	-1	-1
Q	f, Q, /	15	7	-1	-1

3.3 Feature Extraction

There were numerous studies proposing different schemes to extract ECG features from segmented data [5–8, 13, 16]. Assigning all points of a segmented heartbeat, even after down-sampling, causes detrimental effects to the classification process for several reasons. One of them is the high dimensionality of feature vectors which decreases the real-time adaptive learning process. Another drawback would be the redundancy of trivial information, since only special morphology features substantially determine the class

Table 2: List of ECG Features for Classification Stage.

Study	Feature Description
[5]	· RR-intervals (4 features) · Heart-beat intervals A (3 features), · Morphology 1A (9 features)
[13]	· Discrete Wavelet Transform (7 features) (as derived from Section II.C.3.a in [13])
[16]	· Principle Component Analysis (12 features) · Linear Discriminant Analysis (12 features) · Independent Component Analysis (12 features)

of a heartbeat. Moreover there are hidden essential features related to the segments' neighborhood (i.e., abrupt inconsistencies in successive beats), or to the frequency domain [13, 16].

In this study, we select a combined set of features with the aim to avoid those mentioned problems, as shown in Table 2. The set includes 29 features of RR-interval and interpolated ECG morphology [5], 7 features of discrete wavelet transform coefficients [13], and 36 features obtained from 3 different dimension-reduction techniques [16]. The selected features form a 59-dimension feature vector for each heartbeat. It should be noted that although the segmentation decisions are combined on multiple separate ECG leads (see Section 4.2), the selected features only require lead A of MITBIH. This will generate significantly lower dimension feature vectors for our real-time performance.

3.4 Adaptive Learning using Incremental Support Vector Machine

Various learning algorithms for heartbeat classification were suggested with promising results as mentioned in the literature. In our study SVM [28] is employed as the baseline method for the exploitation of the incremental SVM (ISVM) toolbox [23]. ISVM is an analytical method which can integrate a small portion of newly-classified samples into the existing SVM solution without re-training all the sample pool from scratch. That is done by analytically adding each new sample at a time to the solution while retaining the Karush-Kuhn-Tucker conditions on all previously learned data. ISVM is therefore tailored with our original purpose to tackle the ECG subject-adapting problem and efficient in the real-time context. Furthermore ISVM can also *unlearn* samples from the solution which is essential for the real-time speed adjustment in this study. We will briefly summarize both SVM and ISVM as *black boxes* (i.e., with just inputs and outputs without regards to the internal structure) for notational convenience. This method was reported to provide promising result promising results in the context of subject-adaptive brain-computer interface (BCI) [29–32].

The learning dataset $T = \{\overline{x}_i, \overline{y}_i\}_{i=1}^{N_T}$ consisting of N_1 labeled samples from various subjects, is given for the learning
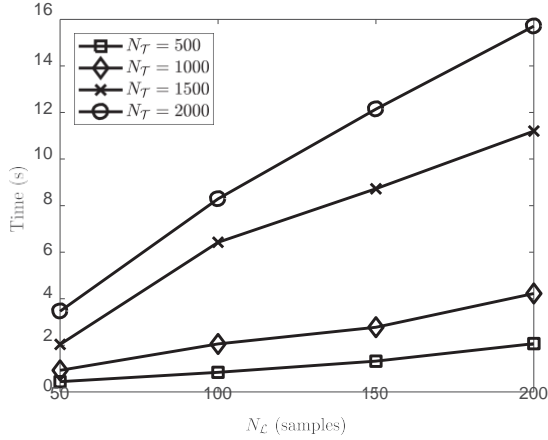
4

Figure 3: ISVM learning time benchmark result, performed on an Intel Core i5-6300HQ, 8GB RAM, Windows 10 64 bit.

Algorithm 3 Real-time Adaptive Learning Routine for the Proposed ECG Framework

---

1: Input: Initial learning set T, expert-annotated set L, validation segmented samples $\mathsf{S} = \{\boldsymbol{s}_i\}$, validation samples per batch k_1, discard samples per batch k_2.
2: Output: Corresponding real-time class label y_i of $\boldsymbol{s}_i$
3: Initialization: $k \leftarrow k_1, (\boldsymbol{a}_D, b_D) \leftarrow \text{ISVM}^+ \left(\text{SVM}(\mathsf{T}), \mathsf{L} \right)$
4: **while** STOP $= 0$ **do**
5:   $y_i = f(\boldsymbol{s}_i, \boldsymbol{a}_D, b_D)$
6:   **if** $i = k$ **then**
7:    $\mathsf{D}^- \leftarrow k_2$ random samples of T
8:    $(\boldsymbol{a}_D, b_D) \leftarrow \text{ISVM}^- \left((\boldsymbol{a}_D, b_D), \mathsf{D}^- \right)$
9:    $\mathsf{D}^+ \leftarrow \{\boldsymbol{s}_j, y_j\}_{j=k-k_2+1}^{k}$
10:    $(\boldsymbol{a}_D, b_D) \leftarrow \text{ISVM}_+ \left((\boldsymbol{a}_D, b_D), \mathsf{D}+ \right)$
11:    $\mathsf{T} \leftarrow \mathsf{T} \setminus \mathsf{D}^-$
12:    $k \leftarrow k + k_1$
13:   **end if**
14:   $i \leftarrow i + 1$
15: **end while**

---

stage. In our context, each sample $\overline{\boldsymbol{x}}_i$ corresponds to one 59-dimensional heartbeat segment, and $\overline{y}_i$ is the corresponding label taken the value of 1 or -1. SVM learns from these samples and builds a classifier denoted (in dual form) by $(\boldsymbol{a}_T, b_T)$, or in other words

$$(\boldsymbol{a}_T, b_T) = \text{SVM}(\mathsf{T}). \tag{5}$$

Then, a set $\mathsf{L} = \{\boldsymbol{x}_i, \hat{y}_i\}_{i=1}^{N_L}$ with N_L manually-annotated labels $\hat{y}_i$ (where $N_L \ll N_T$), is used construct the subject-specific classifier based on $(\boldsymbol{a}_T, b_T)$. ISVM only deals with N_L new samples to construct $(\boldsymbol{a}_{T \cup L}, b_{T \cup L})$, the SVM solution of the combined set $\mathsf{T} \cup \mathsf{L}$, as

$$(\boldsymbol{a}_{T \cup L}, b_{T \cup L}) = \text{SVM}(\mathsf{T} \cup \mathsf{L}) = \text{ISVM}^+ \left((\boldsymbol{a}_T, b_T), \mathsf{L} \right). \tag{6}$$

ISVM is also able to unlearn a set T^- (where $\mathsf{T}^- \subset \mathsf{T}$) from T as

$$(\boldsymbol{a}_{T \setminus T^-}, b_{T \setminus T^-}) = \text{SVM}(\mathsf{T} \setminus \mathsf{T}^-) = \text{ISVM}^- \left((\boldsymbol{a}_T, b_T), \mathsf{T}_- \right) \tag{7}$$

The class decision of any new unclassified sample $\boldsymbol{x}_i$ classified by any trained set D is given by

$$y_i = f(\boldsymbol{x}_i, \boldsymbol{a}_D, b_D) = \text{sign} \left(\sum_{\alpha_j, \boldsymbol{x}_j, y_j \in \mathsf{D}} \alpha_j y_j \boldsymbol{x}_j^T \boldsymbol{x}_i + b_D \right). \tag{8}$$

The algorithm for ISVM batch learning is summarized in Algorithm 3. In the validation stage, the first 100 expert-labeled beats are used as ISVM reinforced set L for each subject. Then after each k_1 classified beats, the last k_2 ones are fed into ISVM again with k_2 arbitrary samples of T discarded from the learning pool. This is referred to as *batch learning*. For the sake of real-time performance $k_1 = 200$ and $k_2 = 100$ are chosen. The classification accuracy is calculated based on the label of every individual beat of the whole real-time process. The real-time mechanism is simulated by a virtual data pumping model built by MATLAB Simulink.

4 EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Dataset Configuration

The MITBIH is divided into the learning and validation sets (namely DS1 and DS2 respectively) as in [5]. Ten subjects who possess the highest percentages of abnormal beats in DS1 (as named in MITBIH: 106, 116, 119, 201, 203, 207, 208, 209, 215, 223) are designated for the learning stage (hereinafter referred to as *learning subjects*). All subjects in DS2 are validated and hereinafter referred to as *validation subjects*.

Only 500 random beats from both classes in DS1 for each classification scheme are selected for the initial SVM classifier construction. The reason for this setting is that there will be a maximum of $N_T = 500$ samples/class $\times$ 2 classes $=$ 1000 training samples. Our benchmarking result, as shown in Fig. 3, points out that it took approximately 2 seconds to integrate $N_L = 100$ samples into the 1000-sample pool. For the real-time requirements, in these 2 seconds there will be approximately 2-3 beats pending for classification after the ISVM updating process. This is the threshold of our system's processing capacity.

4.2 Segmentation Performance Assessment

The separate segmenting results from two leads are combined if a lead A segment has above 80% of their data points in common with a lead B segment within the same time frame, or vice versa. In that case both segments from two leads are designated as a successfully-detected segment. Two standard measures to evaluate the segmentation stage are the sensitivity (Se) and the positive predictivity ($+P$), calculated by

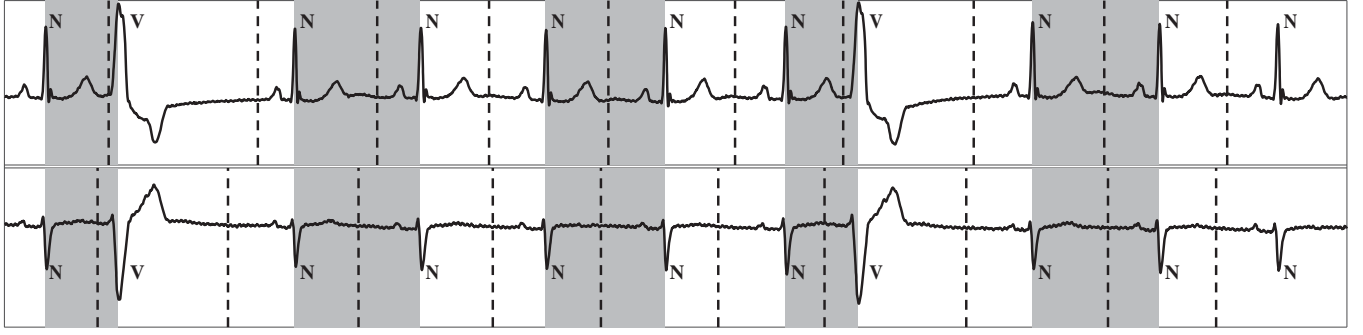$$Se = \frac{T^+}{T^+ + F^-}, \tag{9}$$

5

Figure 2: An example result of the proposed optimal segmentation method performed simultaneously on two ECG leads of MITBIH (the first 10 seconds of subject 119). The dashed vertical lines represent the segmenting change points of our method. The toggle of white-grey background indicates the RR-intervals suggested by MITBIH. Each RR-interval is associated with an annotation letter indicating the AAMI beat class (see Table 1). All heartbeats in this figure are successfully segmented as each segmented interval of our method (bounded by two consecutive vertical dashed lines) contains exactly one annotation.

Table 3: Comparison of Segmentation and Classification Methods using the MITBIH database.

(a) Comparison of Segmentation Methods

Study	Method's Brief Description	Multi-lead	T^+	F^+	F^-	$Se(\%)$	$+P(\%)$
[12]	Wavelet-based QRS detection		109208	153	220	99.80	99.86
[10]	Filter banks' frequency analysis	../	90535	406	374	99.59	99.56
[33]	Analytic QRS detection	../	N/A	N/A	N/A	99.22	99.73
[34]	Analytic QRS detection		N/A	N/A	N/A	99.80	99.80
[35]	Wavelet-based QRS and T-wave detection	../	109635	135	184	99.83	99.88
[36]	Curve-length R-peak detection		109342	154	177	99.86	99.84
[37]	Inexpensive nonlinear filter		109494	353	616	99.43	99.67
[38]	Digital filters and signal energy analysis		109357	97	107	99.90	99.91
Ours (lead A)	Real-time length-constrained optimization		100334	187	210	99.81	99.79
Ours (lead A and B)	Real-time length-constrained optimization	../	100449	113	169	**99.89**	**99.83**

(b) Comparison of Classification Methods

Study	Year	Description	Subject Adaptive	Real-time Mode	VEB (Scheme A)			SVEB (Scheme B)		
					Se	+P	Acc	Se	+P	Acc
[5]	2004	LDA			77.7	81.9	97.4	75.9	38.5	94.6
[6]	2006	Reinforced LDA	../		94.3	96.2	99.4	87.7	47.0	95.9
[14]	2012	LDA with expectation clustering	../		93.0	97.0	N/A	92.0	90.0	N/A
[19]	2014	Resevoir computing		../	87.7	100.0	99.8	84.5	24.4	97.4
[22]	2015	Kalman switching filter	../		97.3	99.9	N/A	N/A	N/A	N/A
[21]	2017	Regression neural network	../	../	88.0	92.6	98.9	85.5	92.2	99.3
Ours		Reinforced ISVM learning	../	../	**97.7**	**99.3**	**99.8**	**93.2**	**88.0**	**99.3**

$$+P = \frac{T_+}{T^+ + F^+}, \qquad (10)$$

where T^+, F^-, and F^+ are the number of correctly segmented heartbeats (true positive), actual heartbeats that are not detected (false positive), and segmented heartbeats that do not correspond to any actual heartbeat (false positive) respectively. The result of our proposed segmentation method and the related studies is shown in Table 3(a). As presented there,

our segmentation method can provide multi-lead processing and the result is competitive to the state-of-the-arts.

4.3 Classification Performance Assessment

With regards to the performance assessment of the classification stage, the three most important measures are the *sensitivity* (*Se*), *positive predictivity* (*+P*), and *accuracy*

6

Table 4: Beat-by-beat Result of Our Method.

Classified label		Non-adaptive				Adaptive			
		VEB		SVEB		VEB		SVEB	
		1	−1	1	−1	1	−1	1	−1
True	1	2766	455	1534	303	3148	73	1712	125
label	−1	153	46338	732	47143	21	46470	233	47642
Se(%)		85.9		83.5		97.7		93.2	
$+P$(%)		94.8		67.7		99.3		88.0	
Acc(%)		98.8		97.9		99.8		99.3	

(Acc) of ventricular detection (scheme A) and supraventricular detection (scheme B), calculated by

$$Se = \frac{\text{number of class 1 samples correctly classified}}{\text{number of class 1 samples classified}}, \quad (11)$$

$$+P = \frac{\text{number of class 1 samples correctly classified}}{\text{number of samples classified into class 1}}, \quad (12)$$

$$Acc = \frac{\text{number of samples correctly classified}}{\text{number of samples classified}}, \quad (13)$$

with the class labeling strategy as shown in Table. 1. The measures in Eq.(11), Eq.(12) and Eq.(13) are interpreted in a simple way and identical to the same measures used in most related studies. The beat-by-beat confusion matrix result and the comparison with typical selected researches are presented in Table 4 and Table 3(b) respectively. Thanks to ISVM fortification there are more 382 VEB beats (from 2766 to 3148) and 178 SVEB beats (from 1534 to 1712) successfully detected. As compared to the previous studies we obtained a promising result in all three assessment measures.

5 CONCLUSION

We consider the ECG segmentation process as a global optimization problem rather than the local detection problem of specific heartbeat features. With the proposed dynamic programming algorithm written to solve this problem, the achievement of the optimal solution is guaranteed. This method also offers a wide range of real-time applications for other data types alongside ECG.

With regards to the classification stage, we are able to establish the subject-adapting capability by use of the ISVM toolbox. As a result significant boosts in sensitivity, predictivity and accuracy of ventricular and supraventricular detection are obtained. The overall performance assessment when combining both the proposed segmentation and classification methods is highly competitive to the previous studies.

REFERENCES

[1] "Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms," *Association for the Advancement of Medical Instrumentation*, 1998.

[2] R. Sameni, M. B. Shamsollahi, C. Jutten, and G. D. Clifford, "A nonlinear Bayesian filtering framework for ECG denoising," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 12, pp. 2172–85, 2007.

[3] B. N. Singh and A. K. Tiwari, "Optimal selection of wavelet basis function applied to ECG signal denoising," *Digital Signal Processing*, vol. 16, no. 3, pp. 275–287, 2006.

[4] O. Sayadi and M. B. Shamsollahi, "Multiadaptive Bionic Wavelet Transform: Application to ECG Denoising and Baseline Wandering Reduction," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, p. 041274, 2007.

[5] C. Philip de, M. O. Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 7, pp. 1196–1206, 2004.

[6] P. de Chazal and R. B. Reilly, "A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12 Pt 1, pp. 2535–43, 2006.

[7] R. G. Kumar and Y. S. Kumaraswamy, "Investigation and classification of ECG beat using Input Output Additional Weighted Feed Forward Neural Network," in *2013 International Conference on Signal Processing , Image Processing Pattern Recognition*, pp. 200–205, Feb 2013.

[8] M. Korürek and B. Doğan, "ECG beat classification using particle swarm optimization and radial basis function neural network," *Expert Systems with Applications*, vol. 37, no. 12, pp. 7563–7569, 2010.

[9] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, pp. 230–236, March 1985.

[10] V. X. Afonso, W. J. Tompkins, T. Q. Nguyen, and L. Shen, "ECG beat detection using filter banks," *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 2, pp. 192–202, 1999.

[11] P. Laguna, R. Jané, and P. Caminal, "Automatic Detection of Wave Boundaries in Multilead ECG Signals: Validation with the CSE Database," *Computers and Biomedical Research*, vol. 27, no. 1, pp. 45–60, 1994.

[12] J. P. Martinez, R. Almeida, S. Olmos, A. P. Rocha, and P. Laguna, "A Wavelet-Based ECG Delineator: Evaluation on Standard Databases," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 4, pp. 570–581, 2004.

[13] M. Llamedo and J. P. Martínez, "Heartbeat Classification Using Feature Selection Driven by Database Generalization Criteria," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 616–625, 2011.

[14] M. Llamedo and J. P. Martinez, "An Automatic Patient-Adapted ECG Heartbeat Classifier Allowing Expert Assistance," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 8, pp. 2312–2320, 2012.

[15] C. Ye, B. V. K. V. Kumar, and M. T. Coimbra, "Combining general multi-class and specific two-class classifiers for improved customized ECG heartbeat classification," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pp. 2428–2431, Nov 2012.

[16] R. Martis, U. R. Acharya, and C. Lim, "ECG beat classification using PCA, LDA, ICA and discrete wavelet transform," *Biomedical Signal Processing and Control*, vol. 8, no. 5, pp. 437–448, 2013.

7

[17] T. Mar, S. Zaunseder, J. P. Martínez, M. Llamedo, and R. Poll, "Optimization of ECG Classification by Means of Feature Selection," *IEEE Transactions on Biomedical Engineering*, vol. 58, pp. 2168–2177, Aug 2011.

[18] E. D. Ubeyli, "Combining recurrent neural networks with eigenvector methods for classification of ECG beats," *Digital Signal Processing*, vol. 19, no. 2, pp. 320–329, 2009.

[19] M. Escalona-Moran, M. Soriano, I. Fischer, and C. R Mirasso, "Electrocardiogram Classification Using Reservoir Computing With Logistic Regression," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 3, pp. 892–898, 2014.

[20] R. Bellman, "On the Approximation of Curves by Line Segments Using Dynamic Programming," *Commun. ACM*, vol. 4, pp. 284–, June 1961.

[21] P. Li, Y. Wang, J. He, L. Wang, Y. Tian, T. s. Zhou, T. Li, and J. s. Li, "High-Performance Personalized Heartbeat Classification Model for Long-Term ECG Signal," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 78–86, 2017.

[22] J. Oster, J. Behar, O. Sayadi, S. Nemati, A. E. W. Johnson, and G. D. Clifford, "Semisupervised ECG Ventricular Beat Classification With Novelty Detection Based on Switching Kalman Filters," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 9, pp. 2125–2134, 2015.

[23] G. Cauwenberghs and T. Poggio, "Incremental and decremental support vector machine learning," in *Proceedings of the 13th International Conference on Neural Information Processing Systems (NIPS'00)*, (3008808), pp. 388–394, MIT Press, 2000.

[24] B. Jackson, J. D. Scargle, D. Barnes, S. Arabhi, A. Alt, P. Gioumousis, E. Gwin, P. Sangtrakulcharoen, L. Tan, and T. T. Tsai, "An algorithm for optimal partitioning of data on an interval," *IEEE Signal Processing Letters*, vol. 12, pp. 105–108, Feb 2005.

[25] Q. Xue, Y. H. Hu, and W. J. Tompkins, "Neural-network-based adaptive matched filtering for QRS detection," *IEEE Transactions on Biomedical Engineering*, vol. 39, no. 4, pp. 317–329, 1992.

[26] A. R. Verma and Y. Singh, "Adaptive Tunable Notch Filter for ECG Signal Enhancement," *Procedia Computer Science*, vol. 57, no. Supplement C, pp. 332–337, 2015.

[27] R. Mark and G. Moody, "MIT-BIH Arrhythmia Database," 1997. [Online]. Available: http://ecg.mit.edu/dbinfo.html.

[28] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A Training Algorithm for Optimal Margin Classifiers," in *Proceedings of the 5th Annual Workshop on Computational Learning Theory (COLT '92)* (D. Haussler, ed.), pp. 144–152, ACM Press, New York, NY, USA, 1992.

[29] K. Vo, T. Pham, D. N. Nguyen, H. H. Kha, and E. Dutkiewicz, "Subject-Independent ERP-Based Brain-Computer Interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 4, pp. 719–728, 2018.

[30] K. Vo, D. N. Nguyen, H. H. Kha, and E. Dutkiewicz, "Real-time analysis on ensemble SVM scores to reduce P300-Speller intensification time," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4383–4386, 2017.

[31] K. Vo, D. Nguyen, H. H. Kha, and E. Dutkiewicz, "Subject-Independent P300 BCI using Ensemble Classifier, Dynamic Stopping and Adaptive Learning," in *IEEE Global Communications Conference (GLOBECOM)*, 2017.

[32] K. Vo, D. Nguyen, H. H. Kha, and E. Dutkiewicz, "Dynamic Stopping Using eSVM Scores Analysis for Event-Related Potential Brain-Computer Interfaces," in *11th International Symposium on Medical Information and Communication Technology (ISMICT)*, 2017.

[33] J. C. T. B. Moraes, M. M. Freitas, F. N. Vilani, and E. V. Costa, "A QRS complex detection algorithm using electrocardiogram leads," in *Proceedings of the International Conference on Computers in Cardiology*, pp. 205–208, 2002.

[34] P. Hamilton, "Open source ECG analysis," in *Proceedings of the International Conference on Computers in Cardiology*, pp. 101–104, 2002.

[35] M. Bahoura, M. Hassani, and M. Hubin, "DSP implementation of wavelet transform for real time ECG wave forms detection and heart rate analysis," *Computer Methods and Programs in Biomedicine*, vol. 52, no. 1, pp. 35–44, 1997.

[36] J. Lewandowski, H. E. Arochena, R. N. G. Naguib, and K. M. Chao, "A simple real-time QRS detection algorithm utilizing curve-length concept with combined adaptive threshold for electrocardiogram signal classification," in *TENCON 2012 IEEE Region 10 Conference*, pp. 1–6, 2012.

[37] D. Castells-Rufas and J. Carrabina, "Simple real-time QRS detector with the MaMeMi filter," *Biomedical Signal Processing and Control*, vol. 21, no. Supplement C, pp. 137–145, 2015.

[38] J. Kim and H. Shin, "Simple and Robust Realtime QRS Detection Algorithm Based on Spatiotemporal Characteristic of the QRS Complex," *PLoS ONE*, vol. 11, no. 3, p. e0150144, 2016.

8