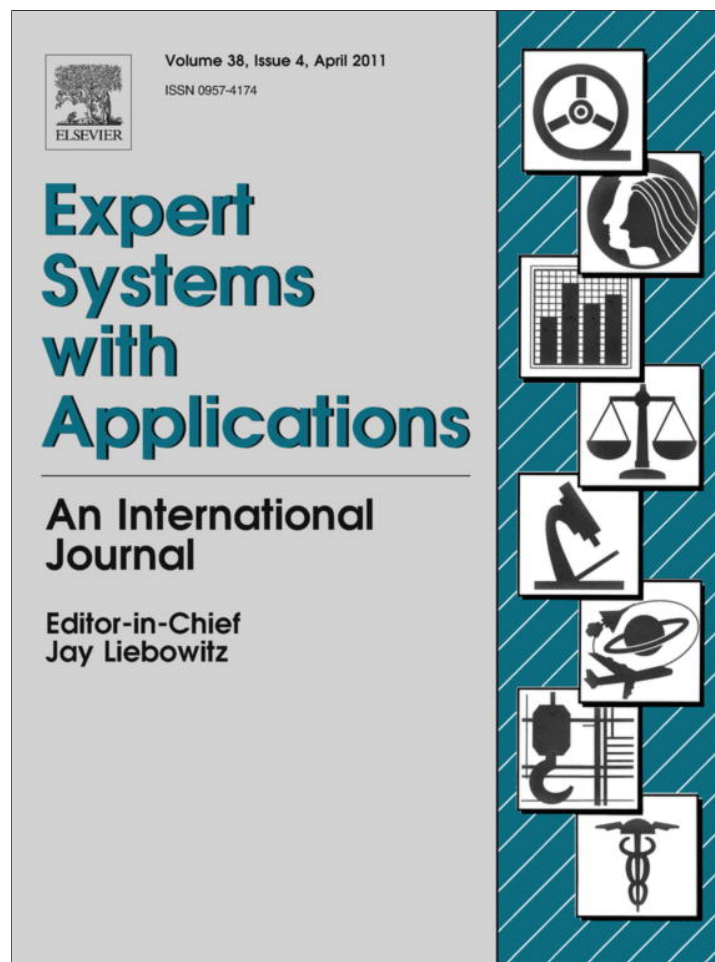


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Personalized long-term ECG classification: A systematic approach

Serkan Kiranyaz^{a,*}, Turker Ince^b, Jenni Pulkkinen^a, Moncef Gabbouj^a^a Department of Signal Processing, Tampere University of Technology, Tampere, Finland^b Department of Electronics and Telecommunications Engineering, Izmir University of Economics, Izmir, Turkey

ARTICLE INFO

Keywords:

Personalized long-term ECG classification
 Exhaustive *K*-means clustering
 Holter registers

ABSTRACT

This paper presents a personalized long-term electrocardiogram (ECG) classification framework, which addresses the problem within a long-term ECG signal, known as *Holter* register, recorded from an individual patient. Due to the massive amount of ECG beats in a *Holter* register, visual inspection is quite difficult and cumbersome, if not impossible. Therefore, the proposed system helps professionals to quickly and accurately diagnose any latent heart disease by examining only the representative beats (the so-called master key-beats) each of which is automatically extracted from a time frame of homogeneous (similar) beats. We tested the system on a benchmark database where beats of each *Holter* register have been manually labeled by cardiologists. The selection of the right master key-beats is the key factor for achieving a highly accurate classification and thus we used *exhaustive K*-means clustering in order to find out (near-) optimal number of key-beats as well as the master key-beats. The classification process produced results that were consistent with the manual labels with over 99% average accuracy, which basically shows the efficiency and the robustness of the proposed system over massive data (feature) collections in high dimensions.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Long-term continuous electrocardiogram (ECG) monitoring and recording, also known as *Holter* electrocardiogram or *Holter* register (Holter, 1961), is needed for detection of some diseases, such as *cardiac arrhythmias*, *transient ischemic episodes* and *silent myocardial ischemia*, and for arrhythmic risk assessment of patients (Paoletti & Marchesi, 2006). Since visual analysis of long-term recordings of the heart activity, with more than 100,000 ECG beats in a single recording, is difficult to diagnose and can be highly error prone, automated computer analysis is of major importance. Most of the *Holter* classification techniques presented up-to-date mainly suffer from the usage of sub-optimal clustering algorithms, such as Max–Min in Syed, Guttag, and Stultz (2007), *k-medians* in Cuesta-Frau, Pe'rez-Corte's, and Andreu-Garci'a (2003) and SOMs in Lagerholm, Peterson, Braccini, Edenbrandt, and Sörnmo (2000), some of which require *a priori* setting of some thresholds or parameters, such as $\theta = 50$ in Syed et al. (2007). Particularly, the performance of the approach in Lagerholm et al. (2000) is limited due to small number of Hermite expansion coefficients used for the approximation of the heartbeats. It is worth noting that although all these techniques claim to address the problem of long-term (*Holter*)

ECG classification, *none* has been applied to a real *Holter* register, probably due to such limitations.

In order to alleviate the problems of the aforementioned sub-optimum clustering schemes, in this paper we used exhaustive *K*-means clustering with the purpose of finding out the true (optimal) number of clusters and their centroids. To assess and find out the best *K*-means run among an exhaustive number of trials, we performed cluster validity analysis, which is the assessment of the clustering method's output using a specific criterion for optimality, i.e. the so-called clustering validity index (CVI). Hence, we used a simple yet efficient CVI in order to assess the clustering performance of each *K*-means run with a given *K* value, which is also varied within a practical range. The particular *K*-means run with the best CVI score is then used for determining the representative beats, or the so-called key-beats. The proposed clustering approach was applied over a real (benchmark) dataset, which contains seven long-term electrocardiogram (ECG) recordings (Physio-Bank) to obtain semi-automatic classification (labeling). Such ambulatory ECG recordings with a typical duration of 24–48 h, are particularly useful for estimating the risk of *ventricular arrhythmias*, such as *premature ventricular contractions (PVCs)*, in patients with heart disease, which may not be detected by a short-time ECG (Paoletti & Marchesi, 2006). Yet any process that requires humans or even an expert cardiologist to examine more than a small amount of data can be highly error prone. A single record of a *Holter* register is usually more than 100,000 beats, which makes the visual inspection almost infeasible, if not impossible. Therefore,

* Corresponding author.

E-mail addresses: serkan.kiranyaz@tut.fi, serkan@cs.tut.fi (S. Kiranyaz), turker.ince@ieu.edu.tr (T. Ince), jenni.pulkkinen@tut.fi (J. Pulkkinen), moncef.gabbouj@tut.fi (M. Gabbouj).

the need for automatic techniques for analyzing such a massive data is imminent and in that, it is crucial not to leave out significant beats since the diagnosis may depend on just a few of them. However, the dynamic range and intra-signal variation in a typical *Holter* register are quite low and abnormal beats, which may indicate the presence of a potential disease, can be scattered along the signal. So based on the proposed exhaustive K -means clustering, a systematic approach is developed, which can summarize a long-term ECG record by discovering the so-called master key-beats that are the representative or the prototype beats from different clusters. With a great reduction in effort, the cardiologist can then perform a quick and accurate diagnosis by examining and labeling only the master key-beats, which in duration are no longer than 15 min of ECG record (for a *Holter* register of 24–48 h). The expert labels over the master key-beats are then back-propagated over the entire ECG record to obtain a patient-specific, long-term ECG classification.

The rest of the paper is organized as follows. Section 2 presents the related work on ECG analysis and surveys related work on data clustering whilst explaining the exhaustive K -means method in detail. The proposed framework for personalized long-term ECG (*Holter*) classification is presented in detail in Section 3. Section 4 provides the experiments conducted over a benchmark *Holter* dataset and discusses the results. Finally, Section 5 concludes the paper.

2. Related work

2.1. ECG classification

Electrocardiogram (ECG) analysis has proven to be an important method routinely used in clinical practice for continuous monitoring of cardiac activities. ECG analysis can be used to detect cardiac arrhythmias, which often arise as a consequence of a cardiac disease and may be life-threatening and require immediate therapy. According to (Mele, 2008), with an estimated 300 millions of ECGs performed every year, there is clearly a need for accurate and reliable ECG interpretation. Computer analysis of ECG data can be of great assistance to the experts in detecting cardiac abnormalities both for real-time clinical monitoring and long-term (24–48 h) monitoring in intensive care unit (ICU) and ambulatory sensors. Many computer-based methods have been proposed for automated analysis and interpretation of ECGs. However, automated classification of ECG beats is a challenging problem as the morphological and temporal characteristics of ECG signals show significant variations for different patients and under different temporal and physical conditions (Hoekema, Uijen, & Oosterom, 2001). This is the reason in practice for the underperformance of many fully automated ECG processing systems, which hence make them unreliable to be widely used clinically (de Chazal & Reilly, 2006). It is also known that accuracy and efficiency of these systems degrade significantly for large datasets (Lee, 1989).

In the past, a number of methods have been proposed for feature extraction from ECG signals including heartbeat temporal intervals (de Chazal, O'Dwyer, & Reilly, 2004), time-domain morphological features (de Chazal, O'Dwyer, & Reilly, 2004), frequency domain features (Minami, Nakajima, & Toyoshima, 1999), wavelet transform features (Inan, Giovangrandi, & Kovacs, 2006), and Hermite transform coefficients (Lagerholm et al., 2000). Accordingly, several techniques have been developed by researchers for ECG data analysis. In (Lagerholm et al., 2000), a method for unsupervised characterization of ECG signals is presented. Their approach involves Hermite function representation of ECG beats (specifically QRS complex) and self-organized neural networks (SOMs) for beat clustering. Application to all (48) 30-min records from the MIT-BIH

arrhythmia database results in 25 clusters and by classifying each cluster according to an expert's annotation of one typical beat, a total misclassification error of 1.5% is achieved. The method proposed in Cuesta-Frau, Pe'rez-Corte's, and Andreu-Garci'a (2003) consists of nonlinear temporal alignment, trace segmentation as feature extraction and k -medians as the clustering algorithm. Its primary goal is to extract accurately significant beats, which can be examined by a physician for the diagnosis. From the results of their experimental studies using 27 registers (for a total 27412 beats) from the MIT-BIH database, k -medians performs better than the *Max-Min* clustering algorithm achieving a clustering error of ~7% in the best case. Syed et al. (2007) describe a new approach for analyzing large amounts of cardiovascular data, for example multiple days of continuous high-resolution ECG data, based on symbolic representations of cardiovascular signals and morphology-based *Max-Min* clustering. It was tested over cardiologist-annotated ECG data (30-min recordings) from 48 patients from the MIT-BIH arrhythmia database achieving 98.6% overall correct classification. This approach has the advantage of using no *a priori* knowledge about disease states allowing for discovery of unexpected events (patterns). The goal of the work in Korurek and Nizam (2008) is to achieve better clustering analysis of ECG complexes using a novel Ant Colony Optimization (ACO) based clustering algorithm. In this study, time dependent morphological parameters extracted from two consecutive periods of an ECG signal are used as specific features. The method is tested using a total of 8771 ECG periods taken from the MIT-BIH database resulting in a total sensitivity of 94.4% to all six arrhythmia types.

2.2. Exhaustive K -means clustering

As the process of identifying natural groupings in a multi-dimensional data based on some distance metric (e.g. *Euclidean*), data clustering can be divided into two main categories: hierarchical and partitional (Frigui & Krishnapuram, 1997). Each category then has a wealth of sub-categories and different algorithmic approaches for finding the clusters. Clustering can also be performed in two different modes: hard (or crisp) and fuzzy. In the former mode, clusters are disjoint, non-overlapping and any data point belongs to a single cluster whereas in the latter case it can belong to all the clusters with some degree of membership (Jain, Murthy, & Flynn, 1999). K -means (Tou & Gonzalez, 1974) is a well-known and widely used clustering method, which first assigns each data point to the closest of the K cluster centroids and then updates the centroids to the mean of their associated points. Starting from a random set of K centroids, this cycle is iteratively performed until the convergence criteria, $\Delta_{Kmeans} < \epsilon$ is reached where the objective function, Δ_{Kmeans} can be expressed as,

$$\Delta_{Kmeans} = \sum_{k=1}^K \sum_{x_p \in c_k} \|c_k - x_p\|^2 \quad (1)$$

where c_k is the k th cluster center, x_p is the p th data point in cluster c_k and $\|\cdot\|^2$ is the distance metric in the *Euclidean* space. As a hard clustering method, K -means is one of the fastest, i.e. $O(n)$, method but suffers from the following drawbacks:

- The number of clusters, K , needs to be set in advance.
- The performance of the method depends on the initial (random) centroid positions as the method converges to the closest local optima.
- The method is also dependent on the data distribution.

The fuzzy version of K -means, the so-called Fuzzy C-means (FCM) (yet sometimes also called as Fuzzy K -means) was proposed by Bezdek (1981) and has become the most popular fuzzy clustering method so far. It is a fuzzy extension of K -means whilst FCM

usually achieves a better performance than *K-means* (Hammerly, 2003) and is less data dependent; however, it still suffers from the same drawbacks, i.e. the number of clusters has to be fixed *a priori* and unfortunately it may also converge to local optima (Jain et al., 1999). Zhang and Hsu (1999) proposed a novel fuzzy clustering technique, the so-called *K* harmonic means (KHM), which is less sensitive to initial conditions and promises further improvements. Experimental results demonstrate that KHM outperforms both *K-means* and FCM (Hammerly & Elkan, 2002; Zhang & Hsu, 1999). The Self Organizing Map (SOM) is a promising approach for data clustering as it gives an intuitive 2D map of a multi-dimensional dataset; however, similar to other deterministic approaches, SOM performs a sub-optimal partitioning depending on the selection of the initial weights and its convergence is strictly dependent on several parameters. Furthermore, it has a *stability* problem as an input pattern may fire different output units and it is only suitable for detecting *hyperspherical* clusters, which are not too likely in high dimensions, (Jain et al., 1999). There are other variants that are skipped. An extensive survey can be found in Jain et al. (1999) and Pal and Biswas (1997).

As a conclusion none of the major clustering techniques can guarantee an optimal solution for clustering massive datasets where it is crucial to find the true number of clusters and the exact cluster centroids. Therefore, we turn the attention back to *K-means*, which is the simplest and the fastest; however, it has those aforementioned drawbacks, particularly a single run of *K-means*, where the centroids are randomly initialized over *N-D* data space, is bound to get trapped in to the closest local optimum, and the optimal *K* is an unknown that should be determined within the process. Taking its speed advantage into account, we run *K-means exhaustively* (significant number of times, e.g. >100 with random initializations) for each *K* within a certain range, i.e. $K_{min} \leq K \leq K_{max}$ in order to increase significantly the probability of converging to a (near-) optimal solution. Among all, we then use the “best” *K-means* run to find out the true (number of) clusters. For the assessment of the clustering performance, the following CVI is used to obtain computational simplicity with minimal or no parameter dependency,

$$f(K, Z) = Q_e K^\alpha \quad \text{where } Q_e = \frac{1}{K} \sum_{j=1}^K \frac{\sum_{z_p \in C_j} \|c_j - z_p\|^2}{N(C_j)} \quad (2)$$

where Q_e is the quantization error (or the average intra-cluster distance) as the *Compactness* term, K^α is the *Separation* term, which simply penalizes higher cluster numbers with an exponential,

$\alpha \geq 0$ and $N(C_j)$ is the number of items in cluster C_j . For $\alpha = 0$, CVI simply becomes the Q_e and using $\alpha = 1$ yields the simplest form (i.e. only the numerator of Q_e). On the other hand, (hard) clustering has some constraints. Let $C_j = \{c_j\}$ be the set of data points assigned to a (potential) cluster centroid. The partitions $C_j, \forall j \in [1, K]$ should maintain the following constraints:

- (1) Each data point should be assigned to one cluster set, i.e. $\bigcup_{j=1}^K C_j = Z$.
- (2) Each cluster should contain at least one data point, i.e. $C_j \neq \{\phi\}, \forall j \in [1, K]$.
- (3) Two clusters should have no common data points, i.e. $C_i \cap C_j = \{\phi\}, \forall i, j \in [1, K] \wedge i \neq j$.

As a hard clustering method, *K-means* is immune to the 1st and 3rd constraints; however, it may fail the 2nd constraint (the so-called under-clustering) especially in high dimensions. Therefore, if any *K-means* run violates this constraint, that run is simply discarded.

3. The proposed framework for personalized long-term ECG classification

In this section we shall describe in detail the systematic approach for personalized classification of long-term ECG data. As the overview shown in Fig. 1, the proposed system addresses the problem within the entire life-time of a long-term ECG signal recorded from an individual patient, i.e. starting from data acquisition and pre-processing, to the temporal segmentation, followed by a master key-beat extraction by two-pass exhaustive *K-means* clustering and finally, classification of the entire ECG data by back propagating the expert cardiologist labels over the master key-beats. As a *personalized* approach, the objective is to minimize as much as possible the amount of data from each individual patient by selecting the most relevant data, which will be subject to manual classification, so that the cardiologist can quickly and accurately diagnose any latent disease by examining only the representative beats (the master key-beats) each from a cluster of homogeneous (similar) beats. This justifies the application of the proposed clustering approach, which aims to extract the optimal (number of) clusters within a diverse dataset. Recall that *optimality* here can only be assessed according to the CVI, the *feature extraction* (data representation) and the *distance* (similarity) metric used. Therefore, the clustering performance can further be

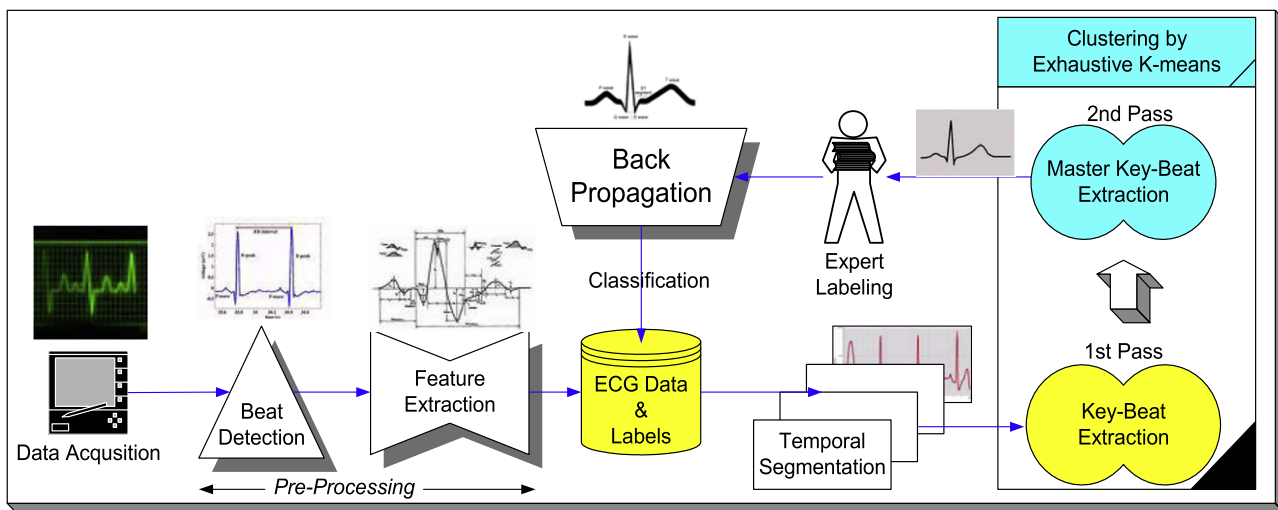


Fig. 1. The overview of the proposed system for long-term ECG classification.

improved by using superior alternatives than the basic and simple ones intentionally used in the current work with the sole purpose of demonstrating the basic performance level of the proposed approach. For both passes the clustering validity assessment is performed using the same CVI given in Eq. (2) with $\alpha = 0$. Recall that this is entirely parameter-free and in addition, L_2 Minkowski norm (Euclidean) is used as the distance metric in the feature space.

As shown in Fig. 1, after the data acquisition is completed, the pre-processing stage basically contains beat detection and feature extraction of the sampled and quantized ECG signal. Before beat detection, all ECG signals are filtered to remove baseline wander, unwanted power-line interference and high-frequency noise from the original signal. This filtering unit can be utilized as part of the heartbeat detection process (for example, the detectors based on wavelet transforms (Li, Zheng, & Tai, 1995)). For all records, we used the first-lead signals and the annotation information (provided with the MIT-BIH database PhysioBank) to locate beats in ECG signals. Beat detection process is beyond the scope of this paper, as many beat detection algorithms achieving over 99% accuracy have been reported in the literature, e.g. Li et al. (1995) and Pan and Tompkins (1985). Before feature extraction, the ECG signal is normalized to have a zero-mean and unit variance to eliminate the effect of dc offset and amplitude biases. Following the detection of each beat of the cardiac cycle within *quasiperiodic* ECG signals based on the R-peak detection and RR-intervals, morphological and temporal features are extracted as suggested in de Chazal et al. (2004) and Hu, Palreddy, and Tompkins, 1997, and combined into a single characteristic feature vector to represent each heartbeat. As shown in Fig. 2, the temporal features relating to heartbeat *fiducial* point intervals and morphology of the ECG signals are extracted by sampling the signals. They are calculated separately from the first-lead signals for each heartbeat. Since the detection of some arrhythmia (such as *Bradycardia*, *Tachycardia* and *premature ventricular contraction*) depends on the timing sequence of two or more ECG signal periods (Tompkins & Webster, 1981), four temporal interval features are considered in our study. They are extracted from heartbeat *fiducial* point intervals (RR-intervals), as follows:

- *pre-RR-interval*: the RR-interval between a given heartbeat and the previous heartbeat,
- *post-RR-interval*: the RR-interval between a given heartbeat and the following heartbeat,

- *local average RR-interval*: the average of the 10 RR-intervals surrounding a heartbeat,
- *average-RR interval*: the mean of the RR-intervals for an ECG recording.

In addition to temporal features, ECG morphology features are extracted from two sampling windows in each heartbeat formation. The sampling windows are formed based on the heartbeat *fiducial* points (maximum of R-wave or minimum of S-wave in Fig. 2). Specifically, the morphology of the QRS complex is extracted using a 150-ms window and 60-Hz sampling rate, resulting in nine ECG samples as features. The eight ECG samples representing the low-frequency T-wave morphology are extracted using a 350-ms window and 20-Hz sampling rate. Each dimension of the 17 dimensional morphological and four dimensional temporal interval features is scaled individually within $[-1, 1]$ interval by applying a nonlinear function, $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$, following a linear transformation. The final 21 dimensional feature vector for each heartbeat is then formed by combining those normalized 17 morphological and four temporal features.

Once the 21 dimensional (21-D) feature vectors composed from the temporal and morphological characteristics of ECG beats are extracted, the entire ECG data is temporally segmented into fixed size frames (segments) in order to achieve mainly two objectives. On one hand, the massive size of ECG data makes it almost infeasible to perform an efficient clustering and on the other hand, *outliers*, which are significantly different from the typical (normal) beats and thus may indicate the presence of an abnormal heart activity, may get lost due to their low-frequency of occurrence. Therefore, we adopt a typical approach, which is frequently performed in audio processing, that is, temporally segmenting data into homogeneous frames. Due to the dynamic characteristics of an audio signal, the frame duration is typically chosen between 20 and 50 ms in order to get as a homogeneous signal as possible. Accordingly, for a 24–48 h long *Holter* register, we have chosen ~5 min long (300 beats) duration for time segments since the intra-segment variation along the time axis is often quite low. So performing a clustering operation within such homogeneous segments will yield only one or few clusters except perhaps the transition segments where a *change*, morphological or temporal, occurs on the normal form of the ECG signal. No matter how minor or insignificant duration this abnormal *change* might take, in such a

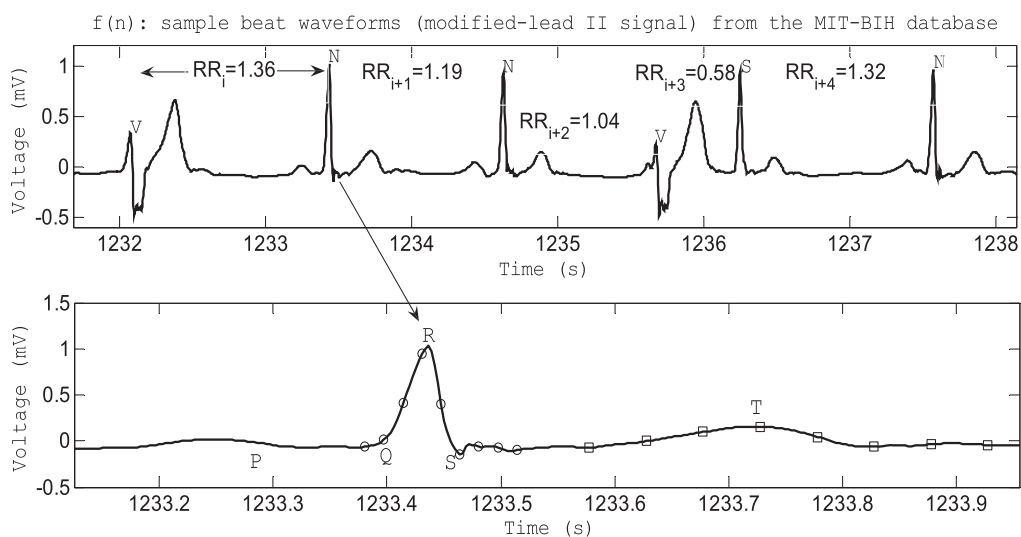


Fig. 2. Sample beat waveforms, including Normal (N), PVC (V), and APC (S) AAMI (Anonymous, 1987) heartbeat classes from the MIT-BIH database. Heartbeat fiducial point intervals (RR-intervals) and ECG morphology features (samples of QRS complex and T-wave) are extracted.

limited time segment, the proposed exhaustive K -means clustering can separate those “different” beats from the normal ones and group them into a distinct cluster. One key-beat, which is the closest to the cluster centroid with respect to the distance metric used in 21-D feature space, is then chosen as the “prototype” to represent all beats in that cluster. Since the optimal number of clusters is extracted within each time segment, only necessary and sufficient number of key-beats is thus used to represent all 300 beats in a time segment. Note that the possibility of missing outliers is thus reduced significantly with this approach since one key-beat is equally selected either from an outlier or a typical cluster without considering their size. Yet redundancy among the key-beats of consecutive segments still exists since it is highly probable that similar key-beats shall occur among different segments. This is the main reason for having the 2nd pass, which performs the exhaustive K -means clustering over key-beats to finally extract the master key-beats. They are basically the “elite” prototypes representing all possible physiological heart activities occurring during a long-term ECG recording.

Since this is a personalized approach, each patient has, in general, normal beats with possibly one or few abnormal periods, indicating a potential heart disease or disorder. Therefore, ideally speaking only a few master key-beats would be expected at the end, each representing a cluster of similar beats from each type. For instance one cluster may contain *ventricular* beats arising from ventricular cavities in the heart and another may contain only *junctional* beats arising from *atrioventricular junction* of the heart. Yet due to the lack of discrimination power of the morphological or temporal features or the similarity (distance) metric used, the clustering operation may create more than one cluster for each anomaly. Furthermore, the normal beats have a broad range of morphological characteristics (Syed et al., 2007) and within a long time span of 24 h or longer, it is obvious that the temporal characteristics of the normal beats may also significantly vary. Therefore, it is reasonable to represent normal beats with multiple clusters rather than only one. In short, several master key-beats may represent the same physiological type of heart activity. The presentation of the master key-beats to the expert cardiologist can be performed with any appropriate way, but this is a *visualization* detail and hence beyond the scope of this work. Finally, the overall classification of the entire ECG data can be automatically accomplished by back propagating the master key-beats' labels in such a way that a beat closest to a particular master key-beat (using the same distance metric in 21-D feature space) is assigned the corresponding label.

4. Experimental results

The systematic approach presented in Section 3 is applied to long-term ECG data in the Physionet MIT-BIH Long-Term database (PhysioBank), which contains six two-channel ECG signals sampled at 128 Hz per channel with 12-bit resolution, and one three-channel ECG sampled at 128 Hz per channel with 10-bit resolution. The duration of the seven recordings changes from 14 to 24 h each and a total of 668, 486 heartbeats in the whole database are used in this study. The database contains annotation for both timing information and beat class information manually reviewed by independent experts. The WFDB (Waveform Database) software package with library functions (from PhysioToolkit) is used for reading digitized signals with annotations. In this study, for all records, we used the first-lead signals and utilized the annotation to locate beats in ECG signals. The CVI, the feature extraction and the distance metric are already presented in Section 2.

Following the pre-processing that consists of the formation of heartbeats using the RR-intervals and the feature extraction there-

after, the patient's long-term ECG data is temporally segmented into homogenous frames of 300 beats (~ 5 min duration) as described in Section 3. With 100 runs for each time frame, the exhaustive K -means clustering is then performed in 21-D feature space to extract the true number of clusters. We used $\alpha = 0$, to make the CVI in Eq. (2) completely parameter-free. The CVI then becomes the quantization error, Q_e and the range for K is set as $2 \leq K \leq 25$. The number of clusters, that is identical to the number of key-beats found automatically for each time frame depends on distinct physiological heartbeat types in each patient's ECG record. As a result, the proposed systematic approach by temporal segmentation and the dynamic clustering technique produces such key-beats that represent all possible physiological heart activities in a patient's ECG data. Therefore, finding the true number of clusters by the proposed systematic approach is the key factor that differentiates it from some earlier works such as Cuesta-Frau et al. (2003) and Syed et al. (2007), both of which iteratively determine this number by an empirical *threshold* parameter. In the proposed method no parameters or threshold values are used.

Table 1 shows the overall results of the proposed systematic approach over all patients from the MIT-BIH Long-Term ECG database. Labels manually annotated by the experts are used only for the master key-beats selected by the proposed system. The classification of the entire ECG data, or in other words, the labeling of all the beats contained therein is then automatically accomplished by back propagation of the master key-beat labels, as explained in Section 3. The performance results tabulated in Table 1 are calculated based on the differences between the labels generated by the proposed approach and the expert supplied labels provided with the database. The Association for the Advancement of Medical Instrumentation (AAMI) provides standards and recommended practices for reporting performance results of automated arrhythmia detection algorithms (AAMI, 1987). In this study, according to the AAMI recommended practice, each ECG beat is classified into the following five heartbeat types: N (beats originating in the sinus mode), S (supraventricular ectopic beats), V (ventricular ectopic beats), F (fusion beats), and Q (unclassifiable beats). In the overall, the proposed systematic approach labeled heartbeats consistent with the cardiologist supplied annotations over 99% of the time within the entire benchmark dataset.

From the results in Tables 1 and 2 it can be seen that the proposed systematic approach performed with very high accuracy for detection of normal (N) and ventricular (V) groups of beats. Specifically, accurate detection of premature ventricular contractions (PVCs) from the ventricular group (V) in long-term ECG data is essential for patients with heart disease since it may lead to possible life-threatening cardiac conditions (Iwasa, Hwa, Hassankhani, Liu, & Narayan, 2005). On the other hand, for supraventricular ectopic (S) beats and some cases of fusion of ventricular (V) and fusion (F) beats, the proposed method did not form a separate cluster corresponding to each type of beat due to the fact that their morphological and temporal features are indeed quite similar to

Table 1

Overall results for each patient in the MIT-BIH Long-Term database using the proposed system. For each class, the number of correctly detected beats is shown relative to the total beats originally present.

Patient	N	S	V	F
14046	105289/105405	1/1	9102/9765	73/95
14134	38548/38766	3/29	9711/9835	744/994
14149	144498/144534	0/0	243/264	0/0
14157	82698/83412	104/244	4334/4368	57/63
14172	57182/58315	401/1003	6526/6527	1/1
14184	77606/78096	13/39	22479/23383	11/11
15814	91129/91617	20/34	9706/9941	1601/1744
Total	596950/600145	542/1350	62101/64083	2487/2908

Table 2

Classification accuracies for each patient in the MIT-BIH Long-Term database using the proposed system.

Patient	14,046	14,134	14,149	14,157	14,172	14,184	15,814	Average
Accuracy	99.31%	98.75%	99.96%	98.99%	97.36%	98.60%	99.15%	99.04%

normal (N) beats. Therefore, we can conclude that a more accurate separation of both S and F beats from the N beats requires a feature extraction technique superior to the one used in the current work.

The (average) classification error, <1%, can further be divided into *critical* and *non-critical* errors, which can be defined as follows: all normal (N) beats that are misclassified as one of the anomaly classes (S, V or F) contributes to *non-critical* error because the expert cardiologist, who is false-alerted by the presence of such beats that indicate a potential cardiac disease, can review them and correct the classification. This is also true for such misclassification among the anomaly classes since those beats are anyway classified as *not normal*, but with a wrong class type and thus they shall be all subject to the expert's attention, following with a manual correction. The *critical* errors occur when a beat in one of the anomaly classes is misclassified as *normal* (N) since this is the case where the expert is not alerted to a potential heart disease. So the consequence of such *critical* errors might be fatal. According to the overall confusion matrix given in Table 3, the (average) *critical* error level is ~0.35% where the majority of the critical errors occurred for the beats within class S due to the above mentioned reasons specific to the feature extraction method used in this study. Overall, no beat is classified as Q type and in reality there are no unclassifiable beats in the benchmark database.

In addition to accuracy, which measures the overall system performance over all classes of beats, three other standard metrics found in the literature (Hu et al., 1997), sensitivity (*Sen*), specificity (*Spe*), and positive predictivity (*Ppr*), are used to quantify the performance of the proposed system with respect to detection of each class of beat. Their respective definitions using true positive (*TP*), true negative (*TN*), false positive (*FP*), and false negative (*FN*), all of which can be obtained from the confusion matrix given in Table 3, are as follows: sensitivity is the rate of correctly classified events among all events, $Sen = TP/(TP + FN)$; specificity is the rate of correctly classified nonevents among all nonevents, $Spe = TN/(TN + FP)$; and positive predictivity is the rate of correctly classified events in all detected events, $Ppr = TP/(TP + FP)$. Accuracy is usually the most crucial metric for determining overall system performance, however due to large variation in the number of beats from

different classes in the long-term ECG dataset, sensitivity, specificity, and positive predictivity can too be critical and relevant performance criteria for medical diagnosis. Table 4 presents performance results of the proposed system in these three areas for each class of beat. Overall, for normal (N) and ventricular (V) groups of beats the proposed system shows high performance, however its sensitivity and positive predictivity for supraventricular ectopic (S) beats are low as both morphology and temporal information features for S beats closely resemble F and N beats.

Since the proposed optimization technique is stochastic in nature, to test the *repeatability* and *robustness* of the proposed system, we performed 10 independent runs on patient record 14172, from which we obtained the lowest performance with the average classification accuracy 97.36%. All runs led to similar accuracy levels with only a slight deviation of ~0.3%. It is also worth mentioning that using a computer with P5 2.4 GHz CPU, the extraction of key-beats in a ~5 min time frame typically takes less than 3 min. Therefore, the proposed system is quite suitable for a real-time application, that is, the key-beats can be extracted in *real-time* with a proper hardware implementation during the recording of a *Holter* ECG.

5. Conclusions

In this paper we proposed a long-term, personalized ECG classification system, which addresses the problem within the entire life-time of a long-term ECG signal recorded from an individual patient and it is tested over a real (benchmark) database containing a total of 668, 486 (manually) labeled heartbeats. To our knowledge this is the first work ever applied to a real full *Holter* database; since most of the earlier works tested only over regular half-hour excerpts from ambulatory ECG records with duration of 30 min or even less, from the benchmark MIT-BIH arrhythmia database (Mark & Moody). As a *personalized* approach with an expert labeling of only 5–15 min long (clinically) distinctive ECG beats from each patient's long-term ECG recording, we achieved an average of above 99% classification accuracy. In a typical 24–48 h long *Holter* register, selection of the right prototype beats, which can yield such a high accuracy level and a great reduction in effort, is mainly due to two key operations. The first one, the so-called temporal segmentation, partitions the entire data into homogenous time segments that can be represented by *minimal* amount of key-beats. The following two-pass exhaustive *K*-means operations first extract the key-beats and then the master key-beats among them. Such a delicate classification accuracy indicates that, in both operations, the proposed approach successfully extracts the true (number of) clusters in a 21-D feature (data) space. Although the outcome of a single *K*-means run is purely random, repeating the classification process several times over the benchmark dataset yields almost identical accuracy levels only with insignificant variations, thus indicating a high level of robustness as well.

Moreover, such a systematic approach apparently promises a high level of *insensitivity* to the length (duration) of the data since the duration of the time segments is fixed and the number of clusters (master key-beats) found in the second pass is not related whatsoever with the number of key-beats in the first pass. Although the proposed system is intended and purposefully developed for analysis of long-term datasets by helping professionals focus on the most relevant patterns, it can also provide efficient and

Table 3

The overall confusion matrix.

Truth	Classification results			
	N	S	V	F
N	596950	2700	114	371
S	801	542	5	2
V	1368	20	62093	602
F	200	1	220	2487

Table 4

Performance of the proposed system for detection of each beat class.

	Sen (%)	Spe (%)	Ppr (%)
N	99.47	96.49	99.60
S	40.15	99.78	16.61
V	96.89	99.38	99.46
F	85.52	99.49	71.84

robust solutions for much shorter ECG datasets. Besides classification, with some proper annotation, master key-beats can also be used for the *summarization* of any long-term ECG data for a fast and efficient visual inspection, and they can further be useful for indexing *Holter* databases, for a fast and accurate information retrieval. On the other hand, $\sim 0.35\%$ critical error rate, although may seem quite insignificant for a short ECG dataset, can still be in practice high for *Holter* registers because it corresponds to several hundreds of misclassified beats, some of which might be important for a critical diagnosis. Yet recall that the optimality of the clustering algorithm depends on the CVI, the feature extraction method and the distance metric, in that, we purposefully use simple and typical ones so as to obtain a basic or unbiased performance level. Therefore, by using for instance a more efficient CVI and better alternatives for distance metric, as in Cuesta-Frau et al. (2003) and Syed et al. (2007), performance may be improved further. Instead of *K*-means, a better clustering method can also be used; however, note that the computational complexity may then become a serious drawback especially when used exhaustively as in the proposed approach. For better and more discriminative features, superior techniques can also be sought within computational biology and information theory, all of which are subject to our future work.

References

- Recommended practice for testing and reporting performance results of ventricular arrhythmia detection algorithms (1987). *Association for the Advancement of Medical Instrumentation*. Arlington, VA.
- Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum.
- Cuesta-Frau, D., Pe'rez-Corte's, J. C., & Andreu-Garci'a, G. (2003). Clustering of electrocardiograph signals in computer-aided *Holter* analysis. *Computer Methods and Programs in Biomedicine*, 72(3), 179–196.
- de Chazal, P., O'Dwyer, M., & Reilly, R. B. (2004). Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, 51(7), 1196–1206.
- de Chazal, P., & Reilly, R. B. (2006). A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, 53(12), 2535–2543.
- Frigui, H., & Krishnapuram, R. (1997). Clustering by competitive agglomeration. *Pattern Recognition*, 30, 1109–1119.
- Hammerly, G. (2003). *Learning structure and concepts in data through data clustering*. PhD thesis.
- Hammerly, G., & Elkan, C. (2002). Alternatives to the *k*-means algorithm that find better clusterings. In *Proc. of the 11th ACM CIKM* (pp. 600–607).
- Hoekema, R., Uijen, G. J. H., & Oosterom, A. v. (2001). Geometrical aspects of the interindividual variability of multilead ECG recordings. *IEEE Transactions on Biomedical Engineering*, 48(5), 551–559.
- Holter, N. J. (1961). New methods for heart studies. *Science*, 134(3486), 1214–1220.
- Hu, Y., Palreddy, S., & Tompkins, W. J. (1997). A patient-adaptable ECG beat classifier using a mixture of experts approach. *IEEE Transactions on Biomedical Engineering*, 44(9), 891–900.
- Inan, O. T., Giovanrandi, L., & Kovacs, G. T. A. (2006). Robust neural-network based classification of PVCs using wavelet transform and timing interval features. *IEEE Transactions on Biomedical Engineering*, 53(12), 2507–2515.
- Iwasa, A., Hwa, M., Hassankhani, A., Liu, T., & Narayan, S. M. (2005). Abnormal heart rate turbulence predicts the initiation of ventricular arrhythmias. *Pace-Pacing and Clinical Electrophysiology*, 28(11), 1189–1197.
- Jain, A. K., Murthy, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Reviews*.
- Korurek, M., & Nizam, A. (2008). A new arrhythmia clustering technique based on ant colony optimization. *Journal of Biomedical Informatics*. doi:10.1016/j.jbi.2008.01.014.
- Lagerholm, M., Peterson, C., Braccini, G., Edenbrandt, L., & Sörnmo, L. (2000). Clustering ECG complexes using Hermite functions and self-organizing maps. *IEEE Transactions on Biomedical Engineering*, 47(7), 838–848.
- Lee, S. C. (1989). Using a translation-invariant neural network to diagnose heart arrhythmia. In *IEEE proc. conf. neural information processing systems*.
- Li, C., Zheng, C. X., & Tai, C. F. (1995). Detection of ECG characteristic points using wavelet transforms. *IEEE Transactions on Biomedical Engineering*, 42(1), 21–28.
- Mark, R., & Moody, G. *MIT-BIH arrhythmia database directory* [Online]. Available: <http://ecg.mit.edu/dbinfo.html>.
- Mele, P. (2008). Improving electrocardiogram interpretation in the clinical setting. *Journal of Electrocardiology*. doi:10.1016/j.jelectrocard.2008.04.003.
- Minami, K., Nakajima, H., & Toyoshima, T. (1999). Real-Time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network. *IEEE Transactions on Biomedical Engineering*, 46(2), 179–185.
- Pal, N. R., & Biswas, J. (1997). Cluster validation using graph theoretic concepts. *Pattern Recognition*, 847–857.
- Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, 32(3), 230–236.
- Paoletti, M., & Marchesi, C. (2006). Discovering dangerous patterns in long-term ambulatory ECG recordings using a fast QRS detection algorithm and explorative data analysis. *Computer Methods and Programs in Biomedicine*, 82, 20–30.
- PhysioBank. *MIT-BIH long-term database directory* [Online]. Available: <http://www.physionet.org/physiobank/database/ltldb/>.
- PhysioToolkit. *The WFDB software package* [Online]. Available: <http://www.physionet.org/physiotools/wfdb.shtml>.
- Syed, Z., Gutttag, J., & Stultz, C. (2007). Clustering and symbolic analysis of cardiovascular signals: discovery and visualization of medically relevant patterns in long-term data using limited prior knowledge. *EURASIP Journal of Application Signal Process*.
- Tompkins, W. J., & Webster, J. G. (1981). Design of microcomputer-based medical instrumentation. *Englewood Cliffs, NJ 07632* (pp. 398–3999). Prentice Hall Inc. ISBN: 0-13-201244-8.
- Tou, J. T., & Gonzalez, R. C. (1974). *Pattern recognition principles*. London: Addison-Wesley.
- Zhang, B., & Hsu, M. 1999. *K-harmonic means – a data clustering algorithm*. Hewlett-Packard Labs Technical Report HPL-1999-124.