

Trigger Learning and ECG Parameter Customization for Remote Cardiac Clinical Care Information System

Mohamed Ezzeldin A. Bashir, Dong Gyu Lee, Meijing Li, Jang-Whan Bae, Ho Sun Shon, Myung Chan Cho, and Keun Ho Ryu, *Member, IEEE*

Abstract—Coronary heart disease is being identified as the largest single cause of death along the world. The aim of a cardiac clinical information system is to achieve the best possible diagnosis of cardiac arrhythmias by electronic data processing. Cardiac information system that is designed to offer remote monitoring of patient who needed continues follow up is demanding. However, intra- and interpatient electrocardiogram (ECG) morphological descriptors are varying through the time as well as the computational limits pose significant challenges for practical implementations. The former requires that the classification model be adjusted continuously, and the latter requires a reduction in the number and types of ECG features, and thus, the computational burden, necessary to classify different arrhythmias. We propose the use of adaptive learning to automatically train the classifier on up-to-date ECG data, and employ adaptive feature selection to define unique feature subsets pertinent to different types of arrhythmia. Experimental results show that this hybrid technique outperforms conventional approaches and is, therefore, a promising new intelligent diagnostic tool.

Index Terms—Arrhythmia, electrocardiogram (ECG), health-care information system, remote cardiac clinical care information system.

I. INTRODUCTION

AN integrated information system or application landscape is a combination of technologies, and human effort to offer best management and decision making. It is wildly designed in many areas to ensure efficiency and effectiveness [1]–[3].

The healthcare information system is interrelated components that provide considerable arguments for providing proper medical and administrative operation [4]–[6]. The cardiac clinical information system, as a part of this, is gaining a very particular

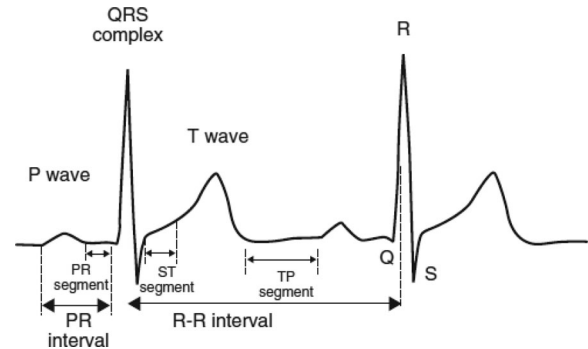


Fig. 1. Normal ECG signal and its different components.

interest since specifying the patient clinical problem clearly and in the right time play a major rule for saving people life.

Electrocardiogram (ECG) signal contains important information that can help medical diagnosis, reflecting cardiac activity, if it is normal or failing heart that has certain pathologies. ECG represents the electrical activity of the heart, as a waveform graph sensed by several electrodes, or leads placed on the body. ECG contains five characteristic peaks and valleys, arbitrarily labeled with successive letters of the alphabet: P, Q, R, S, and T, as shown in Fig. 1 the P wave represents activation of the upper chambers of the heart, the atria, whereas the QRS wave (or complex) and T wave represent excitation of the ventricles or the lower chambers of the heart [7].

Indeed, continuous remote cardiac monitoring for those who require follow-up is demanding. However, morphological characteristics of ECGs vary from person to person and even for a single individual over time. Thus, the current model is run out of domain after a moment, and the accuracy dropped down. Therefore, to build an accurate model, huge amount of training data are required. However, building such a database can be a very costly endeavor and will still only detect a limited number of arrhythmias with limited accuracy. Moreover, since all ECG features must be considered, the computational load would be impractical for fast analysis on a computer with limited resources. Therefore, there has been considerable interest in developing methods to select a subset of features sufficient for accurate classification [8].

Here, we propose a hybrid technique comprising an active learning technique called trigger learning and a method for adaptive feature selection to achieve accurate, arrhythmia detection. The former trains the classifier model with updated data, while the latter selects a unique subset of ECG features related to the QRS complex as well as the P or T waves for each

Manuscript received August 16, 2011; revised December 27, 2011; accepted February 10, 2012. Date of publication February 23, 2012; date of current version July 5, 2012. This work was supported by the National Research Foundation of Korea (NRF) under the Korea Government (MEST) under Grant 2011-0001044.

M. E. A. Bashir, D. G. Lee, M. Li, and H. S. Shon are with the Database/Bioinformatics Laboratory, School of Electrical and Computer Engineering, Chungbuk National University, Cheongju 361-763, Korea (e-mail: mohamed@dblab.chungbuk.ac.kr; dglee@dblab.chungbuk.ac.kr; mjlee@dblab.chungbuk.ac.kr; shon0621@dblab.chungbuk.ac.kr).

J.-W. Bae and M. C. Cho are with the Department of Internal Medicine, Chungbuk National University School of Medicine, Cheongju 361-763, Korea (e-mail: drcorazonmcho@chungbuk.ac.kr; mcho@chungbuk.ac.kr).

K. H. Ryu is with the Database/Bioinformatics Laboratory, School of Electrical and Computer Engineering, Chungbuk National University, Cheongju 361-763, Korea (e-mail: khryu@dblab.chungbuk.ac.kr, corresponding author).

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/TITB.2012.2188812

type of arrhythmia. Together, the two methods achieve sensitive detection with a low computational complexity. This paper is a combination of [9] and [10] but with further discussion, analyses, and experimental results. The outstanding performance of the proposed hybrid technique was demonstrated using various approaches. Experimental results confirm the effectiveness of the proposed technique.

In the remainder of this paper, we provide a brief description of related work, the proposed hybrid technique, and the experimental results. Finally, we summarize our findings and present the main conclusions.

II. RELATED WORK

A. Arrhythmias Detection Training Datasets

The two main approaches to constructing classifiers are the global and the local methods. Global classifiers are built from a large database of ECGs and are the most common solutions used in automatic ECG analysis [11]. In brief, large ECG datasets are randomly divided into training and validation datasets of different sizes; the former is used to train the classifier and the latter to validate it. For example, Rodriguez *et al.* attempted to build an accurate model for classifying cardiac arrhythmias based on feature extraction [12] using (66%) for training and (33%) for validation. However, one main challenge faced with this technique is that the morphologies of ECG waveforms vary widely from patient to patient. Thus, a classifier learned from data specific to one patient will perform very well when tested on data for the same patient, but will often fail on data for other patients. To overcome this problem, the common trend seen in the literature is to increase the size of the training dataset by as much as possible. This trend is also seen in commercial products introduced by various device vendors. However, such an approach has several different drawbacks. First, the huge amount of ECG records necessary to build the classifier will necessitate complex development, maintenance, and update procedures. Second, it is difficult to learn the classifier using abnormal ECGs collected during the monitoring process. Therefore, there is a possibility that specific arrhythmias will not be detected when applying that model to patient records. Moreover, it is impossible to introduce all ECG waveforms from all expected patients [8].

The second approach, the local method, is customized to a specific patient. In other words, the classifier is learned only using datasets collected for that specific patient [13]. The goal is to ensure that the classification model is adapted to the unique characteristics of each patient. Although this technique may alleviate the problem with the learning process, it suffers from a clear disadvantage in terms of the time consuming and labor intensive nature of creating cardiologist-labeled patient-specific training sets. Moreover, only few patients can be expected to be involved in the development of the ECG processing method. Thus, there are limitations to the advantages provided by such technique among the expected audience, even if it is permissible. Hu *et al.* [14] overcome this problem by utilizing a mixture-of-experts (MOE) approach that combines global and local classifiers to realize patient adaptation. This did away with the need to manually label the entire database, thus reducing time and effort. How-

ever, their approach still suffers from several pitfalls: a lack of sensitivity due to comparison between two experts (global classifier and patient-specific local classifier), and considerable cost to develop a local expert for each individual patient. Moreover, it is error prone because of the dependence on different classifiers. We previously suggested a nested ensemble technique to solve the problem of creating an appropriate training dataset. Specifically, we proposed modifying the training dataset with up-to-date data and selecting an adequate set of ECG features for better accuracy [15]. However, despite favorable results, synchronizing the two steps was computationally expensive, which precluded a practical implementation. Moreover, the technique was static to some extent.

B. ECG Parameter Selection

Several methods have been used to extract features as inputs for the classifier: digital filtering [16], Fourier transform [17], [18], wavelet transform [19]–[21], mathematical base technique [22], [23], principal component analysis [24], [25], and independent component analysis (ICA) [26], [27]. ICA, in particular, has been shown to outperform the others especially when applied to the ECG data. Overall, the attractiveness of ICA lies in its lack of use of any strong assumptions on the data. Unlike other approaches, ICA methods do not impose constraints on shape and may, thus, detect responses that would otherwise be ignored by a model-based framework. Moreover, ICA shows a good performance with noise data [28].

Among the various features, most techniques use the QRS complex, mainly the R wave, and ignore the other features (the P and T waves) because the QRS complex is usually quite well defined. From the QRS complex, the RR interval can be determined, which is critical in the diagnosis of many arrhythmias such as premature ventricular contractions, left and right bundled branch blocks, and paced beats. However, there are still a large number of arrhythmias that cannot be detected without considering the P and T waves [29]. In addition, arrhythmias that have different causes may manifest in similar ways on the ECG, taking into account the two main types of arrhythmias: ventricular and supraventricular arrhythmias. The former occur in the ventricles and are recognized because of the abnormal QRS morphology, while the latter occur in the atrium and can only be determined from their effect on the ventricular rhythm. For example, prematurity is used to detect nonsinus beats, sudden pauses as indicators of atrioventricular (AV) conduction disturbances or sinus pauses, and irregularity as a measure of the presence of atrial fibrillation (AF) or flutter. Accordingly, supraventricular abnormalities causing no, or only gradual, changes in ventricular rhythm are not detected by current analysis methods that only refer to the QRS complex for tracing cardiac activity [8].

Most descriptors of QRS complex morphology were developed using pattern recognition techniques [30], which can realize very high accuracy, but is extremely unadaptive to intra- and interpatient ECG morphological disparity. Moreover, the number of morphological descriptors greatly affects computational cost and speed [8]. Such computation can be too complex to

achieve with wireless sensors, which have limited power and can suffer from large noise.

C. Arrhythmias Classification Methods

Automated arrhythmia classification using ECG features (P, QRS, and T) is performed either using supervised and nonsupervised methods. Supervised training requires building a model for classifying the ECG data. The classifier model maps the input features to required output classes on the basis of features specified during training. Several data mining techniques are used for this purpose, with one of the most famous being the decision-tree technique [31]. Much effort has been made to apply artificial neural networks (ANNs) as well. ANNs have good noise tolerance and high efficiency when dealing with nonlinear problems [32], [33], but suffer from many drawbacks. For example, only a limited number of arrhythmias that can be detected due to the restricted number of genuine arrhythmia shapes that can be saved in memory. Moreover, the computational complexity rises rapidly with the number of arrhythmias that are being categorized, which makes the technique impractical. Other methods have also been employed, including support vector machine [34], [35], nearest neighbor analysis [36]–[38], rule-based classifiers [39], fuzzy adaptive classification [40], rule-based rough-set decision system [41], fuzzy neural network [42], multilayer perceptron neural network [43], [44], independent components analysis [45], power spectral density [46], [47], modified MOE [48], image-based technique [49], linear discriminant classifiers [50], Hermit functions [51], high-order spectral analysis [52], Markov approach [53], autoregressive modeling [54], adaptive filtering [55], and genetic algorithm [56]. Also clustering for the purposes of arrhythmias identification is introduced [57]. Recent studies that apply immersing patterns to detect arrhythmias were also applied [58].

III. PROPOSED HYBRID TECHNIQUE

The proposed hybrid technique is composed of two main parts, as shown in Fig. 2 the trigger learning method, and the parameter customization method. These two components work independently, but in a well-synchronized manner. ECGs are sent to the trigger learning method to build an updated training model, and also to the parameter customization system for adopting the features according to the arrhythmias. The hybrid model integrates the two methods to enhance accuracy in real time.

A. Trigger Learning

Conventionally, the computation process to detect arrhythmias starts with detecting the ECG signal, filtering and extracting the useful features, training the classifiers, and then identifying the type of rhythm from among a limited number of labels. In these approaches, errors at early stages such as feature extraction affect the overall performance. Thus, ambiguous outputs persist and might not be resolved using a single learning technique. Moreover, dependence on only one learning process often leads to errors that are apparent in a classifier model.

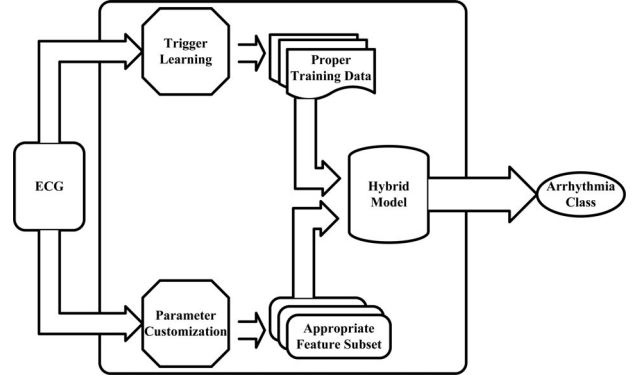


Fig. 2. Proposed hybrid technique components.

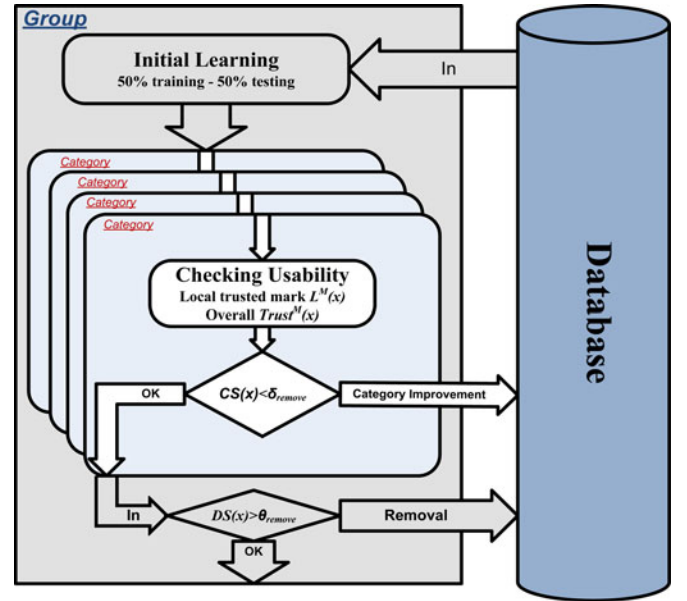


Fig. 3. Trigger learning process flow.

The trigger method was developed to detect arrhythmias in very efficient manner. In essence, it involves to learning the classifier model with up-to-date training data to reflect changes in the morphological descriptors with time. The conventional learning techniques as shown in Fig. 3 (left) try to learn each label assignment process, that is, study the available features with specific class labels to predict future data. By contrast, trigger learning as shown in Fig. 3 (right) is a continuous process that keeps the classifier up-to-date. Partial changes are made to the training dataset when there are insufficient high-quality training data, and complete changes are made when very few high-quality training data are available. That is, new features are introduced to the current training group to update it, or all the present data may be dumped to begin with a fresh dataset if a considerable number of modifications occur. Trigger learning can provide very high accuracy and reduce the computational cost to some extent since the modifications are not conducted in all situations.

The trigger technique has four steps, as shown in Fig. 4 the initial learning stage involves learning from a random set of data

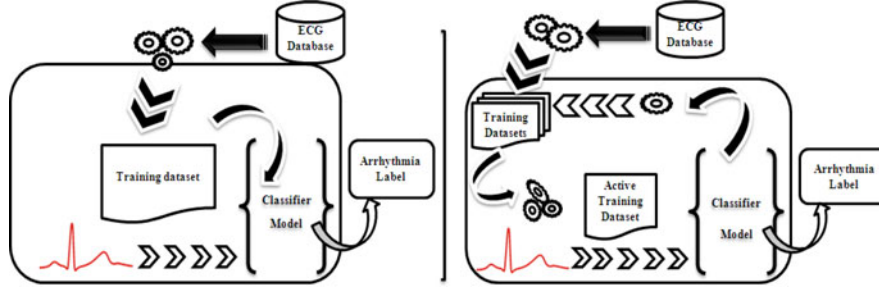


Fig. 4. Traditional training approach and the proposed trigger approach.

without any further considerations. The classifier performance is then evaluated (check) and updated (improve) for consistency. Finally, low-quality data are removed to avoid poor results.

1) *Initial Learning*: First, we start the learning process with a random group of records (categories), which represent (50%) of the overall dataset without considering any factors or any details to start the process of labeling (detecting arrhythmia types). The (check) and (improve) steps are later performed to ensure the correctness of the arrhythmia assignment process when applying the classifier model to testing data that represent (50%) of the overall dataset.

2) *Checking Usability*: After the initial step, the assigned labels are checked on randomly selected categories. This is conducted using an overall trust index $\text{Trust}^M(x)$, which is calculated using the local trust index $L^M(x)$ obtained using a label assigned to a specific category with a specific feature vector. If the label of the same category (with the same feature set) is assigned to the target category, the local trust index $L^M(x)$ will increase. This index $L^M(x)$ is calculated with the following formula:

$$L^M(x) = \sum_{f \in \text{features}} \beta_f(F, i) \cdot C^S(x) \quad (1)$$

where f is the feature number, F is the contribution of the feature, and $C^S(x)$ represents the category score when labeled as arrhythmia (i), which calculated as follows:

$$C^S(x) = \sum_{f \in \text{features}} \beta_f(F, i). \quad (2)$$

The function $\beta_f(F, i)$ checks the set of features (F) in specific category labeled as arrhythmia (i). It returns “+1” if the label (i) is assigned to category (x), otherwise it returns “-1”

$$\beta_f(F, i) = \begin{cases} +1 & \text{if label}(x) = i \\ -1 & \text{otherwise.} \end{cases} \quad (3)$$

The local trust index $L^M(x)$ is considered in determining the overall trust index $\text{Trust}^M(X)$, which is defined using a sigmoid function $\text{Sigmoid}(X)$ ($0.5 < \text{Trust}^M(X) < 1$):

$$\text{Trust}^M(X) = \text{sigmoid} \sum_{x=1}^n L^M(x) \quad (4)$$

$$\text{sigmoid}(x) = \frac{1}{1 + \exp(x)}. \quad (5)$$

The overall $\text{Trust}^M(X)$ is utilized as a likelihood that indicates the usability of the training set (X). If $\text{Trust}^M(X)$ is greater than some arbitrarily chosen threshold, (X) is judged to be reliable, i.e., effective, and otherwise (X) is judged to be unreliable, i.e., ineffective. The unreliable (X) is either improved or removed. The overall $\text{Trust}^M(X)$ fluctuates continuously in relation to the overall performance of the classifier model and its ability to detect different types of arrhythmias.

3) *Improvement*: The checking step ends with one of two judgments: either the current training set is reliable or not for different classes of arrhythmias. Accordingly, unreliable sets must be modified with new data. This process has two parts: first, specifying the useless category or categories; and second, replacing it or them with newly selected one(s). In the first step, category (x) in the active training set (X) is removed if the category score $C^S(x)$ is less than a threshold δ_{remove} . The removal process is as follows:

$$\begin{aligned} &\text{if } C^S(x) < \delta_{\text{remove}} \\ &\text{then remove.} \end{aligned} \quad (6)$$

Second, a new category is selected randomly depending on the probability $p^C(x)$ that a specific category (x) will be used in updating the current training set (X). The probability $p^C(x)$ is relative to the overall $\text{Trust}^M(X)$ calculated in (4)

$$p^C(x_{\text{selected/removed}}) = \frac{C^S(x_{\text{selected/removed}})}{\sum_j C^S(x_j)}. \quad (7)$$

We calculate both P^C for the substitute category (x_{selected}) and the removed category (x_{removed}), and then compare them to avoid selecting the removed one. The selected category is newly assigned to the active training group (active X). Then, the process returns to the loop of the check and improvement steps.

The replacement of the impractical category could be executed several times during the check and update steps. Categories that are removed from the current active training set (X) could be selected in the subsequent update steps for reactivation, which means all categories, could be assigned, regardless of the removal process.

4) *Removal*: The improvement step is useful when there is a limited number of bad labeling using the current group (X), while is useless when there are multiple defects among the categories, which requires an iterative improvement process. This can be very expensive in terms of time and, thus, negatively affect the

performance of the classifier model. Therefore, the removal step is introduced.

All categories in (X) are removed, i.e., the active training set is removed, if it has a *defect score* $D^S(X)$ (the number of removed categories) greater than a threshold θ_{remove} . The removal process is as follows:

$$\begin{aligned} &\text{if } D^S(X) > \theta_{\text{remove}} \\ &\quad \text{then remove.} \end{aligned} \quad (8)$$

In this case, the initial learning step will restart again with the same procedures. However, a new group of categories (not random) should be selected, which can be achieved using (7). Note that the ratio of training to validation data does not affected by improvement or removal steps.

B. ECG Parameters Customization

As mentioned, the aim of this method is to design a unique feature set (distributed through ECG parameters P, QRS, and T) that can be employed to describe arrhythmias in a very sensitive manner. The selection processes identifies one or two parameters in addition to the *QRS* complex. In our design, we accomplish sensitive adaptation on the basis of the necessity of features to specifically detect a specific arrhythmia class. Consequently, considerable accuracy and lower computation complexity are achieved.

Similar arrhythmias often share similar features. Therefore, it is useful to predict the required features to detect different types of arrhythmias. The method uses similar arrhythmias collected from the training data. A parameter score PS is used to quantify the pertinence of a parameter. The overall features list, which represents the arrhythmia class label, is created from the collected group of similar cases. The parameter (P, QRS, and T) with high PS are grouped together to generate an overall features list, which indicates the possibilities of assigning a given arrhythmia class to a case with a specific feature set (distributed through different parameters included in the overall feature list). Accordingly, there will be a different feature lists for each arrhythmia, which enhances the accuracy, and at the same time, reduces the computational burden.

The collected cases are used to calculate PS . First, the ten most similar arrhythmia cases are collected. Then collected categories are manually labeled with binary maps BM s, which indicate the presence “1” or absence “0” of feature F related to a specific parameter in representing a specific type of arrhythmia:

$$BM_{\text{Arrhythmia}}(F) = \begin{cases} 1, & \text{if } F \text{ is positive} \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

Thirty binary labeled maps BM s (ten for each parameter P, QRS, and T) are combined together to create one general PS for any arrhythmia. As shown in Fig. 5, the general PS is created through four steps: Gaussian-weighted sum for BM s, first maximization process O^{1P} , Gaussian-weighted average O^{2P} , and final maximization process O^{3P} .

1) *Weighted Sum*: The ten binary maps BM_p for each parameter $p \in \{P, QRS, T\}$ are smoothed out using an isotropic

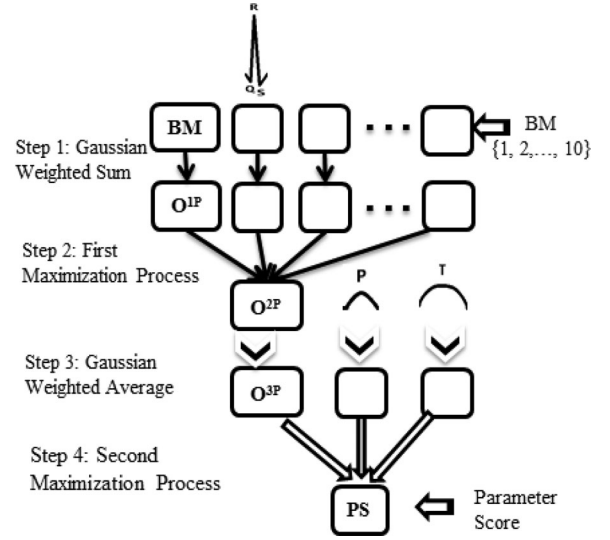


Fig. 5. ECG parameter customization steps.

Gaussian function $g\sigma_{\text{sum}}$ for each feature F

$$O^{1P}(BM_p) = \sum_{f=1}^n g\sigma_{\text{sum}}[f] BM_p[f]. \quad (10)$$

This affords the summation of the weighted features related to each BM_p , which can be used to detect an arrhythmia.

2) *First Maximization Process*: The maximum value among the ten outputs $O^{1P}(BM_p)$ is taken for every parameter p to detect an arrhythmia:

$$O^{2P}(p) = \text{MAX}_p O^{1P}(BM_p). \quad (11)$$

3) *Gaussian Weighted Average*: The output O^{2P} is smoothed using a Gaussian function $g\sigma_{\text{avg}}(p)$ whose mean is the target parameter p :

$$O^{3P}(p) = \frac{1}{S} [g\sigma_{\text{avg}}(p) O^{2P}(p)] \quad (12)$$

where $g\sigma_{\text{avg}}(p)$ is the standard deviation for each parameter p , and S is the total number of features used to describe a specific arrhythmia. This affords a smooth distribution of scores centered on the target parameter p .

4) *Final Maximization Process*: The maximum value among $O^{3P}(p)$ for the three parameters is taken

$$O^{4P}(p) = \text{MAX}_p O^{3P}(p) \quad (13)$$

As described earlier, PS indicates the importance of a parameter p in detecting a specific type of arrhythmia. Therefore, we take the parameter with the highest PS and consider it as the main parameter. Then, we calculate the ratio of the other two parameters to the main parameter. If the ratio is more than or equal to 75%, we consider that parameter as also necessary to detect that type of arrhythmia. Consequently, the unique feature set to describe any arrhythmia in a very sensitive manner is obtained.

TABLE I
ACCURACY ACCORDING TO SPECIFIC ECG PARAMETERS

Parameter	OneR	J48	Naïve Bayes	Dagging	Bagging
QRS only	60.4	91.2	76.5	63.5	81.0
QRS + P	60.4	91.4	77	62.4	81.6
QRS + T	61.3	91.2	76.7	63.0	82.3
QRS + P + T	61.1	92.3	77.7	64.2	83.0

TABLE II
PSS OBTAINED BY PARAMETER CUSTOMIZATION METHOD

Arrhythmia	PS (P)	PS (QRS)	PS (T)
Normal rhythm	85.5	93.7	71.0
Ischemic changes (Coronary Artery Disease)	62.1	87.2	55.6
Old Anterior Myocardial Infarction	66.2	89.4	60.5
Old Inferior Myocardial Infarction	67.4	91.6	63.9
Sinus tachycardia	76.9	88.9	61.7
Sinus bradycardia	78.7	90.7	67.3
Ventricular Premature Contraction (PVC)	86.8	95.0	82.8
Supraventricular Premature Contraction	89.9	67.0	52.0
Left bundle branch block	71.0	97.8	69.1
Right bundle branch block	70.6	94.9	70.7
Left ventricle hypertrophy	81.7	96.6	71.5
Atrial Fibrillation or Flutter	87.9	94.4	68.2
Others	83.2	92.1	78.6

IV. EXPERIMENTAL ENVIRONMENT

We used a database generated at the University of California, Irvine [59], containing 279 attributes and 452 instances [60]. Classes from 01 to 15 were distributed to describe normal rhythm, ischemic changes (coronary artery disease), old anterior myocardial infarction, old inferior myocardial infarction, sinus tachycardia, sinus bradycardia, ventricular premature Contraction (PVC), supraventricular premature contraction, left bundle branch block, right bundle branch block, first-degree AV block, second-degree AV block, third-degree AV block, left ventricle hypertrophy, AF or flutter, and others types of arrhythmias, respectively. Some instances related to specific arrhythmia classes were duplicated, generating overall 573 instances. The experiments were conducted in the WEKA 3.6.1 environment on a PC with an Intel Core 2 Duo processor running at 2.40 GHz with 2.00 GB RAM.

Due to tradeoff between accuracy and speed, we specified the values of $\delta_{\text{remove}} = 1.0$ in (6), and $\theta_{\text{remove}} = 5.0$ in (8). However, the weight $g\sigma_{\text{sum}}$ is the summation of ones in the binary map in order to represent its effect within all binary maps, while $g\sigma_{\text{avg}}(p)$ is the standard deviation of the related first maximization $O^2P(p)$ within the three different parameters (P, QRS, and T).

V. RESULTS

In this section, we will present the results obtained when integrating trigger learning and parameter customization methods to offer clinical information required for efficient cardiac arrhythmia diagnosing system. We use two different types of evaluation mechanisms, when identifying 15 arrhythmias and when detect-

ing only AF arrhythmia. In each type accuracy and speed will be measured, to show the usability in practical implementation.

A. Necessity for Including All ECG Parameters

First, we prove the necessity for including the *P* and *T* waves in conjunction with the QRS complex to evaluate arrhythmias correctly. We measured the performance of five different algorithms with different sets of features: OneR, J48, naïve Bayes, dagging, and bagging. Table I summarizes the accuracy obtained by each algorithm.

B. ECG Parameter Customization

Second, as shown in Table II, we calculated the PSSs related to each arrhythmia in the database [59] obtained by the parameter customization method. Fig. 6 illustrates the specifications of the selected PSSs among the three parameters, depending on the percentage of each in relation to the main (maximum) PS. We consider only the parameters with a ratio to the maximum that is equal to or greater than 0.75.

We found that 23.1% of the cases require P, QRS, and T; 38.5% require only the QRS; 30.8% require P and QRS; and the last 7.6% requires P only. This means that each arrhythmia can be described in much a more accurate manner using just the parameters specified.

C. Arrhythmias Detection

Fig. 7 compares the accuracies achieved by the OneR, J48, naïve Bayes, dagging, and bagging methods when using the hybrid technique, trigger learning, and parameter customization. We also show their original performance without the proposed method for comparison.

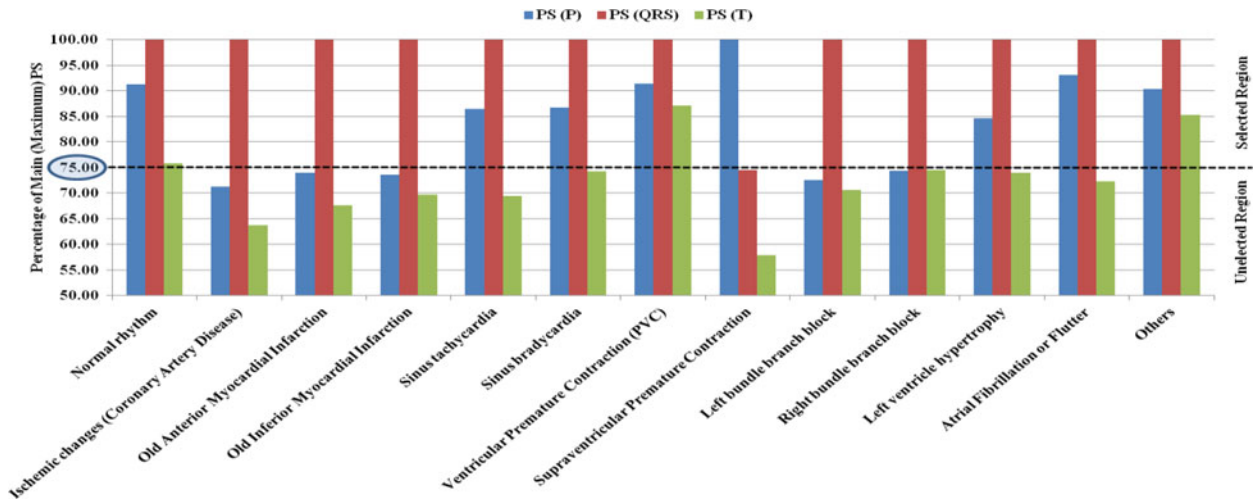


Fig. 6. ECG features selection based on the ratio of PS of a certain parameter to that of the main parameter.

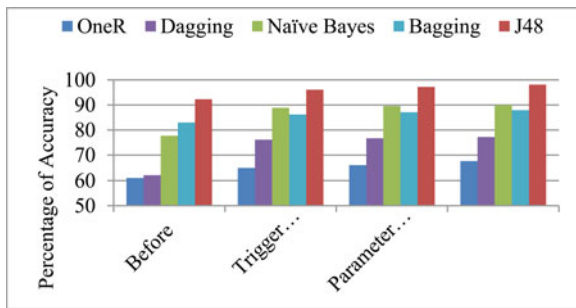


Fig. 7. Accuracy achieved by different methods when using hybrid technique and its components.

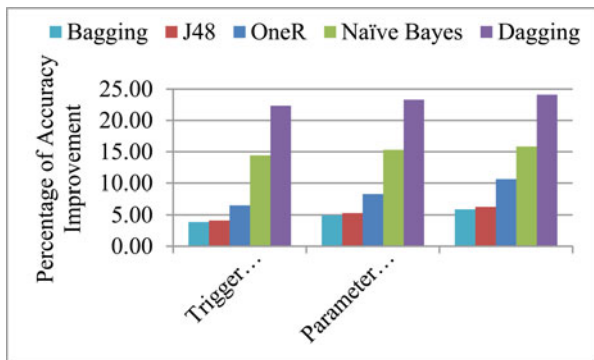


Fig. 8. Accuracy improvement achieved by hybrid technique and its components.

Fig. 8 illustrates the improvements due to the proposed trigger learning, parameter customization, and hybrid techniques in all algorithms tested here. We specifically compare the best-case accuracies when including all features related to the *P*, *QRS*, and *T* waves with that obtained after using the hybrid technique or just one of its components (trigger learning and parameter customization)

These figures clearly show that the trigger learning, customizing parameter, and hybrid methods improve the detection accuracy for the different types of arrhythmia. The improvement is

noticeable for all the algorithms with different weights due to their mechanisms. Specifically, improvements of 6.5%, 5.8%, 12.3%, 15.0%, and 4.9% percentage were achieved in performance for OneR, J48, naïve Bayes, dagging, and bagging, respectively, when applying the hybrid technique. In general, these are significant improvements.

It is also interesting to compare the accuracy of our hybrid technique using the J48 algorithm with that of other methods presented in the literature. Methods from ten representative studies were chosen for this comparison, the including patient-adaptive model [14], Fourier transform and neural network [17], statistical features and fuzzy hybrid neural network [35], principle component with independent component analysis [26], wavelet transform and neural network [19], neuro-SVM-KNN hybrid classifier with virtual QRS image-based geometrical features [30], ECG classification by combining three different kinds of features and neuro-fuzzy network [61], wavelet-based features and neural network [44], fuzzy K-nearest neighbors and neural networks combined with a fuzzy system [40], and independent component analysis with neural network (ICANN) [27]. Table III summarizes the comparative results of these methods, in which the last row lists the results of our model. Among the ten methods, the proposed method outperforms the other methods with an impressive accuracy of 98.1% in discriminating 15 ECG beat types. Further, although, ICNN achieves 0.6% greater accuracy, it can only identify half the number of arrhythmias.

D. Speed

Fig. 9 shows the training and validation times for the J48 classifier with the three methods: trigger learning, parameter customization, and the hybrid technique. As can be seen, parameter customization greatly reduces the computational time for training and validation. Analogously, a smaller number of training samples also lead to a decrease in time required for classifying unknown samples. However, trigger learning takes the most computational time because the process of selecting the right group of data is very complicated. Consequently, the speed

TABLE III
COMPARATIVE RESULT OF DIFFERENT METHOD

Method	Number of Arrhythmia	Accuracy
PAM	4	94.0
WFNN	3	97.1
FTNN	3	98.0
SFHNN	7	96.1
PCICA	5	85.0
WTNN	13	96.8
FNFN	4	98.0
ICNNN	8	98.7
FKNF	6	98.0
NVMKN	7	98.1
Proposed technique	15	98.1

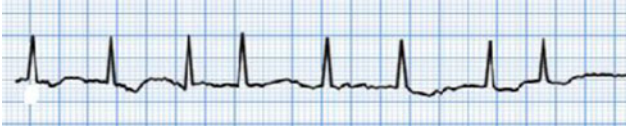


Fig. 9. AF Arrhythmia Shape.

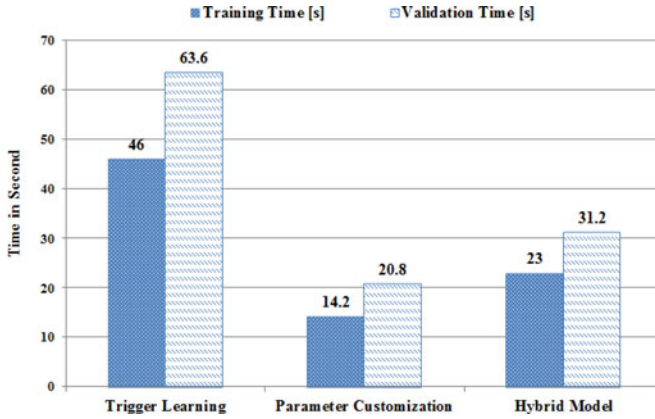


Fig. 10. Speed to detect 15 Arrhythmias.

of the hybrid technique is affected by the negative performance of trigger learning.

E. AF Detection

We measure the performance of the hybrid technique to detect only one arrhythmia. Accordingly, we selected the AF, which is the most frequently occurring cardiac arrhythmia. It is the major cause of morbidity in populations over the age of 75 [62]. AF causes the heart to beat irregularly, leading to inefficient pumping of the blood and changing the blood flow dynamics. The AF chaotic nature affects all the ECG parameters by modifying their shapes and intervals as shown in Fig. 10.

We generate a subset of the same database [59] that contains only AF arrhythmia and normal rhythm, with a total of 266 instances, including 21 cases of AF and with the rest a normal rhythm. We applied the hybrid technique using the J48 algorithm. Sensitivity, specificity, and accuracy were obtained for a detailed performance analysis of the trigger learning, parameter customization, and hybrid technique. The classification performance is generally presented by a confusion matrix where, TP,

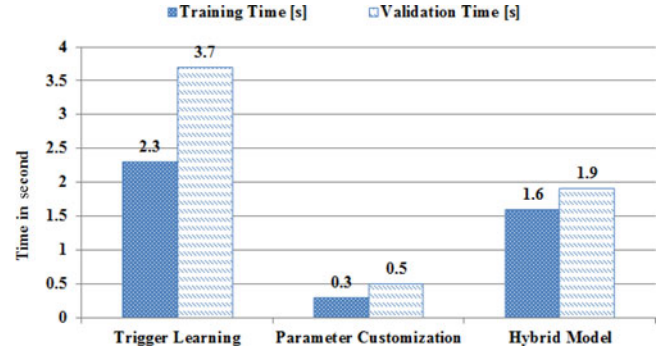


Fig. 11. Speed to detect AF.

TN, FP, and FN stand for true positive, true negative, false positive, and false negative, respectively. Accordingly, we evaluated: the accuracy, expressed in percentage of the division of the sum of correctly detected AF (TP + TN) by the sum of all parameters (TP + TN + FP + FN), resulting in a measure of the precision of the algorithm. Sensitivity, expressed in percentage of the division of all true AF (TP) by the sum of TP + FN provides a measure of the capacity of the technique to detect AF. Specificity, expressed in percentage of the division of all non-AF (TN) by the sum of TN + FP, provides a measure of the capacity of the technique to confirm the nonpresence of AF episodes in the ECG.

1) *Accuracy*: Table IV shows the figures obtained when applying trigger learning, parameter customization, and hybrid model in conjunction with the J48 to detect AF. The results imply that they have good predictive abilities and generalization performance. Based on the results, the customization parameter model has provided slightly better performance than the trigger learning method. In contrast, when we combine the two techniques in a mixture framework we achieve outstanding performance. In particular, the sensitivity and specificity of the proposed hybrid framework on the testing data are 95.2% and 99.6% respectively, and its accuracy is 99.2%.

Several researchers have addressed the AF arrhythmia detection problem using the ECG signals directly or by analyzing the heart rate variability signal [63]–[66]. Generally, all these techniques utilize either QRS complex mainly the R wave, or the P waves. The literature never shows the employment of other ECG parameters and their intervals to detect AF arrhythmia. Table V summarizes the testing results obtained by different methods. It can be observed from this table that the models derived using hybrid technique contain the trigger learning and parameter customization methods provide better accuracy than those obtained by other methods, which are reported in the literature.

2) *Speed*: Also in this section, we will detail the training and test times of the J48 classifiers with the three methods trigger learning, parameter customization, and hybrid model when applying to detect the AF.

Fig. 11 illustrates the computational time for the training and the testing process that needed to detect AF.

Again the performance of parameter customization method is much better than trigger learning method and hybrid model.

TABLE IV
PERFORMANCES TRIGGER LEARNING, PARAMETER CUSTOMIZATION, AND HYBRID TECHNIQUE

Method	TP	FN	FP	TN	Sensitivity	Specificity	Accuracy
Trigger learning	17	3	4	242	85.0%	98.4%	97.4%
Parameter customization	18	3	3	242	85.7%	98.8%	97.7%
Hybrid technique	20	1	1	244	95.2%	99.6%	99.2%

TABLE V
COMPARATIVE RESULTS OF DIFFERENT AF DETECTION METHODS

Author	Database	Method	Sensitivity	Specificity	Accuracy
Christov et al. [63]	ECG	P wave	95.7%	-	98.8%
Logan and Healey [64]	HRV	RR irregularity	96%	89%	-
Rodriguez and Silveira [65]	HRV	RR interval	91.4%	-	-
Fukunami et al. [66]	ECG	P wave	91%	76%	-
Hybrid technique	ECG	Trigger learning and parameter customization	95.2%	99.6%	99.2%

The reason is discussed clearly in Section V-D when applied the proposed methods to detect 15 arrhythmias.

VI. CONCLUSION

Clinical Cardiac health information system is a challenging problem in the field of data mining and extracting knowledge that has received a great deal of attention over the past few years because of its importance to save people's lives and reduce risk. Specially, systems designed to offer continuous remote cardiac monitoring. However, current remote cardiac clinical information system is far from adequate and efficient performances. This is partly because of inter- and inpatient (ECG) morphological descriptors, which are varying through the time. Thus, developing one classifier model to satisfy all patients in different situations using static training datasets is not practical. Furthermore, since all ECG features must be considered, the computational load would be impractical for fast analysis on a computer with limited resources.

In this paper, we presented a hybrid technique comprising an active learning technique called trigger learning and a method for adaptive feature selection to achieve efficient, cardiac clinical information system. The former trains the classifier model with updated data, while the latter selects a unique subset of ECG features related to the QRS complex as well as the P or T waves for each type of arrhythmia. Together, the two methods achieve sensitive detection with a low computational complexity. The performance of our framework was evaluated using various approaches, which demonstrate their effectiveness. In future, we plan to perform more experiments to account for interrelated ECG features, and measure the responsiveness of the model to noise with a different percentage.

REFERENCES

- [1] L. Wang, L. Xu, X. Wang, W. J. You, and W. Tan, "Knowledge portal construction and resources integration for a large scale hydropower dam," *Syst. Res. Behav. Sci.*, vol. 26, no. 3, pp. 357–366, 2009.
- [2] L. Xu, "Editorial: Inaugural Issue," *Enterprise Inf. Syst.*, vol. 1, no. 1, pp. 1–2, 2007.
- [3] L. Xu, "Enterprise systems: State-of-the-art and future trends," *IEEE Trans. Ind. Inf.*, vol. 7, no. 4, pp. 630–640, Nov. 2011.
- [4] R. Perez-Castillo, B. Weber, J. Pinggera, S. Zugal, I. Guzman, and M. Piattini, "Generating event logs from non-process-aware systems enabling business process mining," *Enterprise Inf. Syst.*, vol. 5, no. 3, pp. 301–335, 2011.
- [5] L. Duan, W. N. Street, and E. Xu, "Healthcare information systems: Data mining methods in the creation of a clinical recommender system," *Enterprise Inf. Syst.*, vol. 5, no. 2, pp. 169–181, 2011.
- [6] E. Xu, M. Wermus, and B. Bauman, "Development of an integrated medical supply information system," *Enterprise Inf. Syst.*, vol. 5, no. 3, pp. 385–399, 2011.
- [7] G. Clifford, F. Azuaje, and P. McSharr, *Advanced Methods and Tools for ECG Data Analysis*. Norwood, MA: Artech house, 2006, pp. 4–12.
- [8] M. Ezzeldin A. Bashir, M. Akasha, D. G. Lee, G. M. Yi, K. H. Ryu, J.-W. Bae, M. C. Cho, and E. J. Cha, "Highlighting the current issues with pride suggestions for improving the performance of real time cardiac health monitoring," in *Proc. 1st Int. Technol. Bio Med. Inf. Conf.*, Bilbao, Spain, 2010, pp. 226–233.
- [9] M. Ezzeldin A. Bashir, K. S. Ryu, S. H. Park, D. G. Lee, J.-W. Bae, H. S. Shon, and K. H. Ryu, "Superiority real-time cardiac arrhythmias detection using trigger learning method," in *Proc. 2nd Int. Technol. Bio Med. Inf. Conf.*, Toulouse, France, 2011, pp. 53–65.
- [10] M. Ezzeldin A. Bashir, K. S. Ryu, S. H. Park, D. G. Lee, J.-W. Bae, H. S. Shon, and K. H. Ryu, "Pro-detection of atrial fibrillation with ECG parameters mining technique," in *Proc. 2nd Int. Conf. Electr. Electron. Eng.*, Macao, China, 2011, pp. 717–724.
- [11] G. Bortolan, I. Jekova, and I. Christov, "Comparison of four methods for premature ventricular contractions and normal beats clustering," *Comp. Card.*, vol. 30, pp. 921–924, 2005.
- [12] U. Rajendra, P. Subbann, S. Iyengar, A. Raod, and S. Dua, "Classification of heart rate data using artificial neural network and fuzzy equivalence relation," *Pattern Recog.*, vol. 36, pp. 61–68, 2003.
- [13] F. de Chazal and R. B. Reilly, "A patient adapting heart beat classifier using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2535–2543, Dec. 2006.
- [14] Y. H. Hu, S. Palreddy, and W. J. Tompkins, "Patient adaptable ECG beat classification using mixture of experts," in *IEEE Trans. Biomed. Eng.*, vol. 44, no. 9, pp. 891–900, Sep. 1997.
- [15] M. Ezzeldin A. Bashir, M. Akasha, D. G. Lee, Min Yi, K. H. Ryu, E. J. Bae, M. Cho, and C. Yoo, "Nested ensemble technique for excellence real time cardiac health monitoring," in *Proc. World Congr. Comput. Sci. Comput. Eng. Appl. Comput.*, Las Vegas, NV, 2010, pp. 519–525.
- [16] V. S. Vyas and P. P. Rege, "Real coded genetic algorithm for optimal parameters selection in circular Mellin feature extractors," *Int. J. Comput. Appl.*, vol. 30, no. 2, pp. 65–72, 2008.
- [17] K. Minami, H. Nakajima, and T. Toyoshima, "Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 2, pp. 179–185, Feb. 1999.

- [18] A. Yildiz, M. Akin, and M. Poyraz, "An expert system for automated recognition of patients with obstructive sleep apnea using electrocardiogram recordings," *Expert Syst. Appl.*, vol. 38, no. 10, pp. 12880–12890, 2011.
- [19] G. K. Prasad and J. S. Sahambi, "Classification of ECG arrhythmias using multi-resolution analysis and neural networks," in *Proc. IEEE Conf. Convergent Technol.*, Bangalore, India, 2003, pp. 227–231.
- [20] K. Sternickel, "Automatic pattern recognition in ECG timeseries," *Comput. Meth. Prog. Biomed.*, vol. 68, pp. 109–115, 2002.
- [21] Z. Dokur and T. Olmez, "ECG beat classification by a novel hybrid neural network," *Comput. Meth. Prog. Biomed.*, vol. 66, pp. 167–181, 2001.
- [22] X. Wang and K. Paliwal, "Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition," *Pattern Recog.*, vol. 36, pp. 2429–2439, 2003.
- [23] H. X. Li, L. D. Xu, J. Y. Wang, and Z. W. Mo, "Feature space theory in data mining: Transformations between extensions and intensions in knowledge representation," *Exp. Syst.*, vol. 20, no. 2, pp. 60–71, 2003.
- [24] H. X. Li and L. D. Xu, "Feature space theory—a mathematical foundation for data mining," *Kno.-Bas. Syst.*, vol. 14, pp. 253–257, 2001.
- [25] Y. Ozbay, R. Ceylan, and B. Karlik, "A fuzzy clustering neural network architecture for classification of ECG arrhythmias," *Comput. Biol. Med.*, vol. 36, pp. 376–388, 2006.
- [26] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999.
- [27] M. I. Owis, A. B. Youssef, and Y. M. Kadah, "Characterization of ECG signals based on blind source separation," *Med. Biol. Eng. Comput.*, vol. 40, pp. 557–564, 2002.
- [28] Y. Sung-Nien and C. Kuan-To, "Integration of independent component analysis and neural networks for ECG beat classification," *Expert Syst. Appl.*, vol. 34, pp. 2841–2846, 2008.
- [29] L. Sörnmo and P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Amsterdam, The Netherlands: Elsevier, 2005, ch. 6.
- [30] M. R. Homaeinezhad, S. A. Atyabi, E. Tavakkoli, H. N. Toosi, A. Ghaffari, and R. Ebrahimpour, "ECG arrhythmia recognition via a neuro-SVM-KNN hybrid classifier with virtual QRS image-based geometrical features," *Expert Syst. Appl.*, vol. 39, no. 2, pp. 2047–2058, 2012.
- [31] S. Osowski and T. Linh, "ECG beat recognition using fuzzy hybrid neural network," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 11, pp. 1265–1271, Nov. 2001.
- [32] T. P. Exarchosa, M. G. Tsipourasa, C. P. Exarchosa, C. Papaloukasa, D. I. Fotiadis, and L. K. Michalis, "A methodology for the automated creation of fuzzy expert systems for ischaemic and arrhythmic beat classification based on a set of rules obtained by a decision tree," *Artificial Intell. Med.*, vol. 40, no. 3, pp. 187–200, 2007.
- [33] H. Li and L. Li, "Representing diverse mathematical problems using neural networks in hybrid intelligent systems," *Expert Syst.*, vol. 16, no. 4, pp. 262–272, 1999.
- [34] F. Sheng-wei, "Diagnostic study on arrhythmia cordis based on particle swarm optimization-based support vector machine," *Expert Syst. Appl.*, vol. 37, no. 10, pp. 6748–6752, 2010.
- [35] R. Yuan, Z. Li, X. Guan, and L. Xu, "An SVM-based machine learning method for accurate internet traffic classification," *Inf. Syst. Frontiers*, vol. 12, no. 2, pp. 149–156, 2010.
- [36] A. Kampouraki, G. Manis, and C. Nikou, "Heartbeat time series classification with support vector machines," *IEEE Trans. Informat. Technol. Biomed.*, vol. 13, no. 4, pp. 512–518, Jul. 2009.
- [37] L. Duan, L. Xu, F. Guo, J. Lee, and B. Yan, "A local-density based spatial clustering algorithm with noise," *Inf. Syst.*, vol. 32, pp. 978–986, 2007.
- [38] L. Duan, L. Xu, Y. Liu, and J. Lee, "Cluster-based outlier detection," *Ann. Oper. Res.*, vol. 168, pp. 151–168, 2009.
- [39] P. Chazal, M. Dwyer, and R. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196–1206, Jul. 2004.
- [40] O. Castillo, P. Melin, E. Ramirez, and J. Soria, "Hybrid intelligent system for cardiac arrhythmia classification with Fuzzy K-Nearest Neighbors and neural networks combined with a fuzzy system," *Expert Syst. Appl.*, vol. 39, pp. 2947–2955, 2012.
- [41] S. Mitra, M. Mitra, and B. B. Chaudhuri, "A rough set-based inference engine for ECG classification," *IEEE Trans. Instrum. Meas.*, vol. 55, no. 12, pp. 2198–2206, Dec. 2006.
- [42] L. Y. Shyu, Y. H. Wu, and W. C. Hu, "Using wavelet transform and fuzzy neural network for VPC detection from the Holter ECG," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1269–1273, Jul. 2004.
- [43] H. G. Hosseini, K. J. Reynolds, and D. Powers, "A multi-stage neural network classifier for ECG events," in *Proc. 23rd Int. Conf. IEEE Eng. Med. Biol. Soc.*, New Zealand, 2001, pp. 1672–1675.
- [44] A. E. Zadeh, A. Khazaee, and V. Ranaee, "Classification of the electrocardiogram signals using supervised classifiers and efficient features," *Comput. Methods Programs Biomed.*, vol. 99, pp. 179–194, 2010.
- [45] S. N. Yu and K. T. Chou, "Selection of significant for ECG beat classification," *Expert Syst. Appl.*, vol. 36, pp. 2088–2096, 2009.
- [46] U. R. Acharya, M. Sankaranarayanan, J. Nayak, C. Xiang, and T. Tamura, "Automatic identification of cardiac health using modeling techniques: A comparative study," *Inform. Sci.*, vol. 178, pp. 4571–4582, 2008.
- [47] C. H. Lin, "Frequency-domain features for ECG beat discrimination using grey relational analysis based classifier," *Comput. Math. Appl.*, vol. 55, pp. 680–690, 2008.
- [48] E. D. Ubeyli, "Features for analysis of electrocardiographic changes in partial epileptic patients," *Exp. Syst. Appl.*, vol. 36, pp. 6780–6789, 2009.
- [49] R. Sarvestani, R. Boostani, and M. Roopaei, "VT and VF classification using trajectory analysis," *Nonlinear Anal.: Theory, Method Appl.*, vol. 71, no. 12, pp. e55–e61, 2009.
- [50] P. Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196–1206, Jul. 2004.
- [51] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sörnmo, "Clustering ECG complexes using Hermite functions and self-organizing maps," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 7, pp. 838–848, Jul. 2000.
- [52] L. Khadra, A. S. Al-Fahoum, and S. Binajaj, "A quantitative analysis approach for cardiac arrhythmia classification using higher order spectral techniques," *IEEE Trans. Bio. Eng.*, vol. 52, no. 11, pp. 1840–1845, Nov. 2005.
- [53] R. V. Andreao, B. Dorizzi, and J. Boudy, "ECG signal analysis through hidden Markov models," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 8, pp. 1541–1549, Aug. 2006.
- [54] M. Corduas and D. Piccolo, "Time series clustering and classification by the autoregressive metric," *Comput. Statist. Data Anal.*, vol. 52, no. 4, pp. 1860–1872, 2008.
- [55] A. Sengur, "An expert system based on linear discriminant analysis and adaptive neuro-fuzzy inference system to diagnosis heart valve diseases," *Expert Syst. Appl.*, vol. 35, no. 1–2, pp. 214–222, 2008.
- [56] S. Osowski, K. Siwek, and R. Siroic, "Neural system for heartbeats recognition using genetically integrated ensemble of classifiers," *Comput. Biol. Med.*, vol. 41, pp. 173–180, 2011.
- [57] I. M. Ishag, A. F. A. Dafa-Alla, G. Min Yi, D. G. Lee, J.-W. Bae, and K. Ho Ryu, "Agglomeration, cluster accuracy, hierarchical clustering. On-body sensor," *Adv. Exp. Med. Biol.*, vol. 680, no. 1, pp. 83–88, 2010.
- [58] H. G. Lee, K. Noh, B. J. Lee, H. S. Shon, and K. Ho Ryu, "Cardiovascular disease diagnosis method by emerging patterns," *Adv. Data Mining Appl.*, vol. 4093, pp. 819–826, 2006.
- [59] UCI Machine Learning Repository, (2010). [Online]. Available: <http://www.ics.uci.edu/~mllearn/MLRepository.html>
- [60] WEKA web site, (2010). [Online]. Available: <http://www.cs.waikato.ac.nz/~ml/weka/index.html>
- [61] M. Engin, "ECG beat classification using neuro-fuzzy network," *Pattern Recog. Lett.*, vol. 25, pp. 1715–1722, 2004.
- [62] N. M. Wheeldon, "Atrial fibrillation and anticoagulant therapy," *Euro. Heart J.*, vol. 16, pp. 302–312, 1995.
- [63] I. Christov, G. Bortolan, and I. Daskalov, "Sequential analysis for automatic detection of atrial fibrillation and flutter," in *Proc. Comput. Cardiol.*, 2001, pp. 293–296.
- [64] B. Logan and J. Healey, "Robust detection of atrial fibrillation for a long term telemonitoring system," in *Proc. Comput. Cardiol.*, 2005, pp. 619–622.
- [65] C. A. R. Rodriguez and M. A. H. Silveira, "Multi-thread implementation of a fuzzy neural network for automatic ECG arrhythmia detection," in *Proc. Comput. Cardiol.*, 2001, pp. 297–300.
- [66] M. Fukunami, T. Yamada, M. Ohmori, K. Kumagai, K. Umamoto, A. Sakai, N. Kondoh, T. Minamino, and N. Hoki, "Detection of patients at risk for paroxysmal atrial fibrillation during sinus rhythm by P wave-triggered signal-averaged electrocardiogram," *Circulation*, vol. 83, pp. 162–169, 1991.



Mohamed Ezzeldin A. Bashir received the Post Graduate Diploma and the M.S. degree in computer science and information technology from Gazera University, Wad Madani, Sudan, and the M.S. degree in business administration science from the University of Medical Sciences and Technology, Khartoum State, Sudan, and the Ph.D. degree in computer science from Chungbuk National University, Cheongju, Korea, in 2012.

He is a Professor in Faculty of Computer Science and Information Technology, University of Medical Sciences and Technology, Khartoum State, Sudan. His research interests include data mining, information retrieval, database systems, bioinformatics, and biomedicine.



Ho Sun Shon received the B.S. and M.S. degrees in statistics from the Sungshin Women University of Nature Science, Seoul, Korea, in 1986 and 1992, respectively, and the Ph.D. degree in computer science from Chungbuk National University, Cheongju, Korea, in 2010.

She is currently with the Database/Bioinformatics Laboratory, School of Electrical and Computer Engineering, Chungbuk National University. Her research interests include the area of bioinformatics, biomedicine, data mining, and pattern recognition.



Dong Gyu Lee received the B.S. degree in the Department of Information Communication Engineering from Seowon University, Chungbuk, Korea, in 2005, and the M.S. degree in the Department of Computer Science from Chungbuk National University, Cheongju, Korea, in 2007. He was a Research Student in the University of Aizu, Aizu-Wakamatsu, Japan, in 2011. He is currently working toward the Ph.D. degree in Database and Bioinformatics Laboratory, Department of Computer Science, Chungbuk National University.

His research interests include data mining, medical informatics, and geographic information systems.



Myeong Chan Cho received the MD degree and Bachelor of Medical Science in 1983, Post Graduate diploma and the Ph.D. degree in Medicine in 1996 from Seoul National University, Seoul, Korea.

Currently, he is a Professor in medicine (cardiology subspecialty) in Chungbuk National University School of Medicine, Cheongju, Korea, and works as the President of the Korean National Institute of Health, Chungcheongbuk-do, Korea. His research interests include heart failure, vascular biology, GWAS and stem cell therapy.



Meijing Li received the B.S. degree in the School of Information and Computing Science from Dalian University, Dalian, China, in 2007, and the M.S. degree in the Department of Bio Information Technology from Chungbuk National University, Cheongju, Korea, in 2010, where she is currently working toward the Ph.D. degree in the Department of Computer Science.

Her major research interests include bioinformatics and data mining.



Keun Ho Ryu (M'82) received the Ph.D. degree from Yonsei University, Seoul, Korea, in 1988.

He is a Professor in the School of Electrical and Computer Engineering, the Chungbuk National University, Cheongju, Korea. He was a Postdoctoral Researcher at the University of Arizona and also a Research Scientist at Electronics and Telecommunications Research Institute, Daejeon, Korea. His research interests include temporal databases, spatiotemporal databases, temporal GIS, ubiquitous computing and stream data processing, knowledgebase information

retrieval, database security, data mining, and bioinformatics.

Dr. Ryu is a member of the ACM since 1983.



Jang-Whan Bae received the MD degree and Bachelor of Medical Science in 1994, Post Graduate diploma and the Ph.D. degree in Medicine in 2004 from Chungbuk National University, Korea.

Currently, he is an Assistant Professor in medicine (cardiology subspecialty) in the Department of Internal Medicine, Chungbuk National University School of Medicine, Cheongju, Korea. His research interests include ischemic heart disease, lipid metabolism, and stem cell therapy.