Real-Time Automated Cardiac Health Monitoring by Combination of Active Learning and Adaptive Feature Selection

Mohamed Ezzeldin A. Bashir[1], Ho Sun Shon[2], Dong Gyu Lee[2], Hyeongsoo Kim[2] and Keun Ho Ryu[2]*

[1]Faculty of Computer Science & Information Technology, University of Medical Sciences & Technology
Khartoum, Sudan
[2] School of Electrical & Computer Engineering, Chungbuk National University
Cheongju, 361-763 - South Korea
[1][e-mail: m.ezzo@umst-edu.org]
[2][e-mail: {shon0621, dglee, hskim, khryu}@dblab.chungbuk.ac.kr]
*Corresponding author

Received May 13, 2012; revised September 20, 2012; accepted November 19, 2012; published January 29, 2013

Abstract

Electrocardiograms (ECGs) are widely used by clinicians to identify the functional status of the heart. Thus, there is considerable interest in automated systems for real-time monitoring of arrhythmia. However, intra- and inter-patient variability as well as the computational limits of real-time monitoring poses 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.

Keywords: Arrhythmia, Real-time cardiac health monitoring, Adaptive classification, Adaptive feature selection, Electrocardiogram

A preliminary version of this paper appeared in DEXA 2011, August 29 - September 2, Toulouse, France. Also in ICBBB 2011, Februrary 26-28, Singapore. This version intigrates the two previouse versions with extera expermental works and analysis. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012-0000478). Its also supported by the University of Medical sciences and Technology (UMST) Khartoum, Sudan.

http://dx.doi.org/10.3837/tiis.2013.01.007

1. Introduction

Electrocardiogram (ECGs) are a series of waves and deflections representing cardiac (heart) electrical activity, as sensed by several electrodes, or leads. placed on the body. ECGs contain 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.

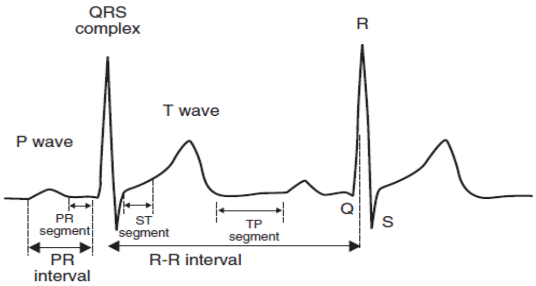


Fig. 1. Normal ECG signal showing temporal and amplitude characteristics of different components [1]

ECGs are a very important clinical tool for characterizing the functional status of the heart. Heart arrhythmias can be accurately identified by expert clinicians simply on the basis of changes in the characteristics of the P, QRS, and T components [1]. Indeed, wired ECG monitoring in hospitals is very crucial in saving lives. However, such monitoring is inadequate for patients with coronary heart disease, who require continuous follow-up and monitoring. In addition, the morphological characteristics of ECGs vary from person to person and even for a single individual over time. Thus, to build an accurate automated classification model, a huge amount of training data is 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 [2].

Here we propose a hybrid technique comprising an active learning technique and a method for adaptive feature selection to achieve accurate, real-time 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 type of arrhythmia. Together, the two methods achieve sensitive detection with a low computational complexity. This paper is a combination of [3] and [4] 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.

2. Related Work

2.1 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 (e.g., see [5] and [6]). 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 [7]. They randomly divided a global dataset into a training (66%) and validation (33%) set and used the Waikato Environment for Knowledge Analysis (WEKA) and SPSS AnswerTree tools for learning. Sixteen methods were used in the experiments. 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 [9]. 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. [10] 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. However, 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 [11]. However, despite favorable results, synchronizing the two steps was computational expensive, which precluded a real–time implementation. Moreover, the technique was static to some extent.

2.2 Feature Extraction

Several methods have been used to extract features as inputs for the classifier: digital filtering [12], Fourier transform [13, 14], wavelet transform [15, 16, 17], principal component analysis (PCA) [18, 19], and independent component analysis (ICA) [20, 21]. ICA, in particular, has been shown to outperform the others. This method identifies the underlying factors or components from multivariate (multidimensional) statistical data. What distinguishes ICA from other methods is that it finds components that are both statistically independent and non-Gaussian. ICA has been successfully applied to numerous signal processing problems in areas such as biomedicine, communications, finance, and remote sensing. In addition, it has recently found applications in the study of ECG data [22, 23], where it has been used in an exploratory manner to detect consistent patterns of heart activation with a common time course. 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 can isolate sources of structured noise that may otherwise be too complex to model as confounds in the multiple regression framework of the general linear model (GLM) [24].

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 [25]. 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 as a feature to detect non-sinus beats, sudden pauses as indicators of atrioventricular conduction disturbances or sinus pauses, and irregularity as a measure of the presence of atrial fibrillation 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 [26].

Most descriptors of QRS complex morphology were developed using pattern recognition techniques [27]. Measuring the diversity between the sequential and frequency characteristics of the QRS complex waveform has also been attempted, such as by using Karhunen–Loeve transforms [28], Hermite functions [29], and the wavelet transform [30]. Recently, methods of determining the adaptive time–frequency transform of ECG signals and calculating the applicable time–frequency features, which reveal the structures of the signals, have been introduced [31, 32]. The most popular approaches are based on pattern recognition techniques using morphological features, which can realize very high accuracy, but there are several disadvantages. First, a very large database of templates must be stored in memory for matching. Second, the accuracy relies on threshold-based segmentation to discriminate components of the ECG signal, which is extremely unadaptive to intra- and interpatient ECG morphological disparity. Finally, with such features, only a limited numbers of classes of waveforms can be extracted to describe specific cardiac arrhythmias. Moreover, the number of morphological descriptors greatly affects computational cost and speed [33]. Such computation can be too complex to achieve with wireless sensors, which have limited power and can suffer from large noise. On the other hand, in some studies, time–frequency analysis of ECG waveforms has been used to detect the abnormal cardiac conditions [34, 35].

2.3 Arrhythmia Classification

Automated arrhythmia classification using ECG features (P, QRS, and T) is performed either using supervised and non-supervised methods [36, 37]. 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 [38, 39]. Several efforts have been made to apply artificial neural networks (ANNs) as well. ANNs have good noise tolerance and high efficiency when dealing with non-linear problems [26, 40, 41], 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 [42, 43], nearest neighbor analysis [44, 45], rule-based classifiers [46], fuzzy adaptive classification [47, 48], rule-based rough-set decision system [49], clustering for the purposes of arrhythmias identification is introduced [50]. Recent studies that apply immerging patterns to detect arrhythmias were also applied [51].

Generally, these methods are general and can be applied to any classification task. The techniques can be evaluated on the basis of accuracy: correct descriptions of arrhythmias, effectiveness: sensitivity to abnormalities, efficiency: speed and reliability: determine how far doctors can trust a model. These factors vary from one method to the other.

3. Proposed Hybrid Technique

The proposed hybrid technique is composed of two main parts: the active learning method, and the feature selection method. These two components work independently, but in a well synchronized manner. ECGs are sent to the active learning method to build an updated training model, and also to the feature selection system for tuning the features according to the arrhythmias. The hybrid model integrates the two methods to enhance accuracy in real time.

3.1 Active 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.

The active 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 try to learn each label assignment process, that is, study the available features with specific class labels to predict future data. By contrast, active learning 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.

For the efficiency and the effectiveness requirements, the technique is based upon in-between process which is called double impact so the change always takes place through the double impact. Active learning provides very high accuracy and reduces the computational cost to some extent since the modifications are not conducted in all situations.

The active technique has four steps, as shown in **Fig. 2.** The initial learning stage involves learning from a random set of data without any further considerations. The classifier performance is then evaluated (check) and updated (improve) for consistency. Finally, low-quality data is removed to avoid poor results [39]. The double impact is used to substitute the partial or complete modification of the current active training data set. The Double impact conducts the improvement and the removal process in two different files. The objective is to minimize hitting the main database as much as possible, so as to save time furthermore enhance accuracy by making double filtering.

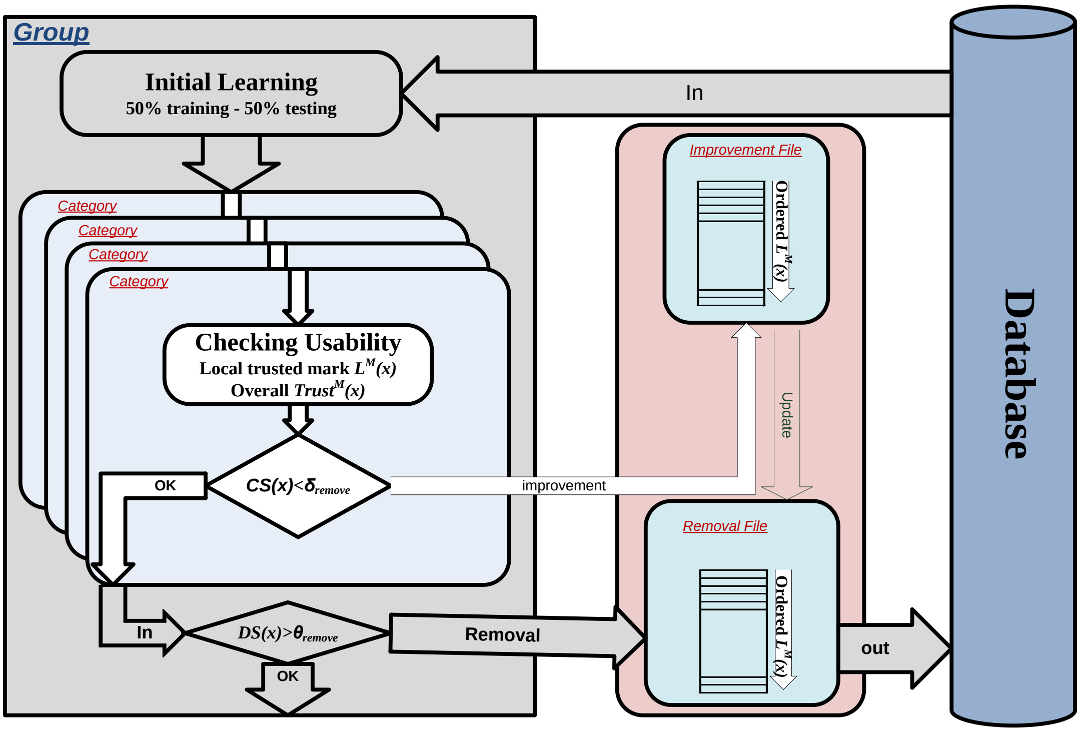


Fig. 2. Active learning process flow [52]

3.1.1 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.

3.1.2 Checking Usability

After the initial step, the assigned labels are checked on randomly selected categories. This is conducted using an overall trust index $Trust^M(x)$, 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:

$$l^M(x) = \sum_{f \in features} \beta_f(F,i).C^s(x) \quad [39, 52] \tag{1}$$

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:

$$C^S(x) = \sum_{f \in features} \beta_f(F,i) \quad [39, 52] \tag{2}$$

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."

$$\beta_f(F,i) = \begin{cases} +1 & \text{if label(x)} = \text{i} \\ -1 & \text{other wise} \end{cases} \quad [39, 52] \tag{3}$$

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

$$Trust^M(X) = sigmoid \sum_{x=1}^{n} l^m(x) \quad [39, 52] \tag{4}$$

$$sigmoid(x) = \frac{1}{1+\exp(x)} \quad [39, 52] \tag{5}$$

The overall $Trust^M(X)$ is utilized as a likelihood that indicates the usability of the training set (X). If $Trust^M(X)$ 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 $Trust^M(X)$ fluctuates continuously in relation to the overall performance of the classifier model and its ability to detect different types of arrhythmias.

3.1.3 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:

$$\text{if } C^S(x) < \delta_{remove} \text{ then remove} \quad [39, 52] \tag{6}$$

The removed category or categories will send to the improvement file in the double impact area not to the main database. So the removed category will send with their $C^S(x)$ in a sending order manner. Accordingly, the saving process in the improvement file is conducted depending on the category score (the high score in the top). The size of the improvement file is fixed so as not to be exceeding the 50% of the whole.

Second, a new a category is selected randomly from the main database 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 $Trust^M(X)$ calculated in equation (4).

$$p^C(x_{selected/removed}) = \frac{C^S(x_{selected/removed})}{\sum_j C^S(x_j)} \quad [39, 52] \tag{7}$$

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.

Starting from the second modification process, the substitutions take place from the improvement file in the double impact area not from the main database. The selection of the substituted category is depending on the $C^S(x)$, so the category with high score will be selected i.e. selecting the category on the top of the list. The category will remove from the improvement file, if selected two times and removed from the active learning set. In this case, it replaced by fresh category from the main database using equation (7) to avoid selecting the same one.

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.

3.1.4 Removal

The improvement step is usful 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* $D^S(X)$ greater than a threshold θ_{remove}. The removal process is as follows:

$$\text{if } D^S(X) > \theta_{remove} \text{ then remove} \quad [39, 52] \tag{8}$$

As it happened with improvement stage the transfer and the replacement processes are conducted through the double impact area, but this time using the removal file. Except in the first time the initial learning step will restart again with the same procedures.

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

equation (7). Note that the ratio of training to validation data does not affected by improvement or removal steps.
The removal file generated by updating the improvement file. The best categories with highest categoryscore $C^S(x)$ in the improvement file, will copied to the removal file. The sizes of the removal file exactly represent the size of the training data. The different between the two files in the double impact area (improvement and the removal files) is that, the old `good categories with high $C^S(x)$, which their performnaces droubled sudenlly will no removed from the removal file as it happened in the improvement file. They will be active but in the bottom of the list with specific sign to specify them. If the improvement is no achived in the next coming testing proces, then, they will be swapted with new categories from the main database using equation (7) to insure random selection and avoide selecting the samed ones.

3.2 ECG Features Selection

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 feature list of each arrhythmia is predicted from similar cases collected from the training data based on general features. 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 BMs, which indicate the presence “1” or absence “0” of feature *F* related to a specific parameter in representing a specific type of arrhythmia:

$$BM_{Arrhythmia}(F) = \begin{cases} 1 & \text{if } F \text{ is positive} \\ 0 & \text{otherwise} \end{cases} \qquad [39, 52] \qquad (9)$$

Thirty binary labeled maps *BMs* (ten for each parameter P, QRS, and T) are combined together to create one general *PS* for any arrhythmia. As shown in **Fig. 3**, the general *PS* is created through four steps: Gaussian-weighted sum for *BMs*, first maximization process O^{1P}, Gaussian-weighted average O^{2P}, and final maximization process O^{3P} [52].

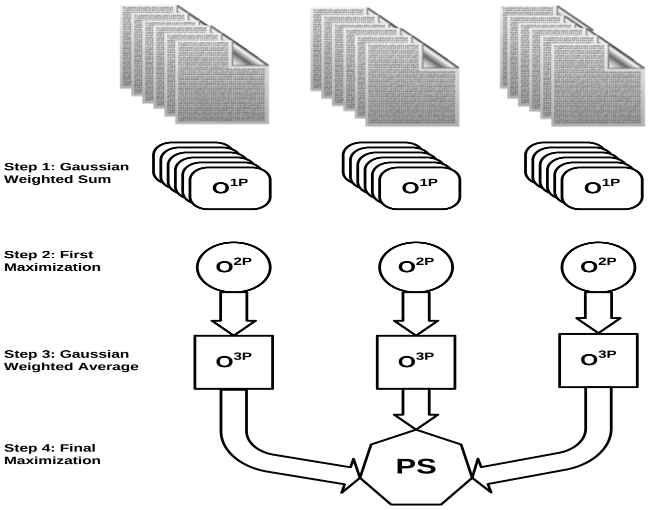


Fig. 3. ECG feature selection steps [39]

3.2.1 Weighted Sum

The ten binary maps BM_p for each parameter $P \in \{P, QRS, T\}$ are smoothed out using an isotropic Gaussian function $g\sigma_{sum}$ for each feature F.

$$O^{1P}(BM_p) = \sum_{f=1}^{n} g\sigma_{sum}[f]\ BM_p[f] \qquad [39, 52] \qquad (10)$$

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

3.2.2 First Maximization Process

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

$$O^{2P}(p) = MAX_p\, O^{1P}(BM_p) \qquad [39, 52] \qquad (11)$$

3.2.3 Gaussian Weighted Average

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

$$O^{3P}(p) = \frac{1}{f}[g\sigma_{avg}(p) O^{2P}(p)] \qquad [39, 52] \qquad (12)$$

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

3.2.4 Final Maximization Process

The maximum value among O^{3P} (p) for the three parameters is taken.

$$O^{4P}(p) = MAX_p\, O^{3P}(p) \quad [39, 52] \tag{13}$$

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.

4. Experimental Results and Analysis

We used a database generated at the University of California, Irvine [53], containing 279 attributes and 452 instances [54]. 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 atrioventricular (AV) block, second-degree AV block, third-degree AV block, left ventricle hypertrophy, atrial fibrillation 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. The parameters were set as follows: in equation (6), δ_{remove} = 1.0, and in equation (8), θ_{remove} = 5.0.

4.1 Necessity for including all ECG Features

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 1** summarizes the accuracy obtained by each algorithm.

Table 1. Accuracy of different algorithms according to ECG parameters included [39, 52]

Parameters	*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

4.2 ECG Features Selection

Second, as shown in **Table 2**, we calculated the PSs related to each arrhythmia in the database [53] obtained by the feature selection method. The specifications of the selected PSs among the three parameters are considered depending on the percentage of each PS 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.

Table 2. PSs obtained by feature selection method [39]

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

4.3 Arrhythmia Detection

Fig. 4 compares the accuracies achieved by the OneR, J48, naïve Bayes, dagging, and bagging methods when using the hybrid technique, active learning, and feature selection. We also show their original performance without the proposed method for comparison.

Fig. 5 illustrates the improvements due to the proposed active learning, feature selection, 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 (active learning and feature selection)

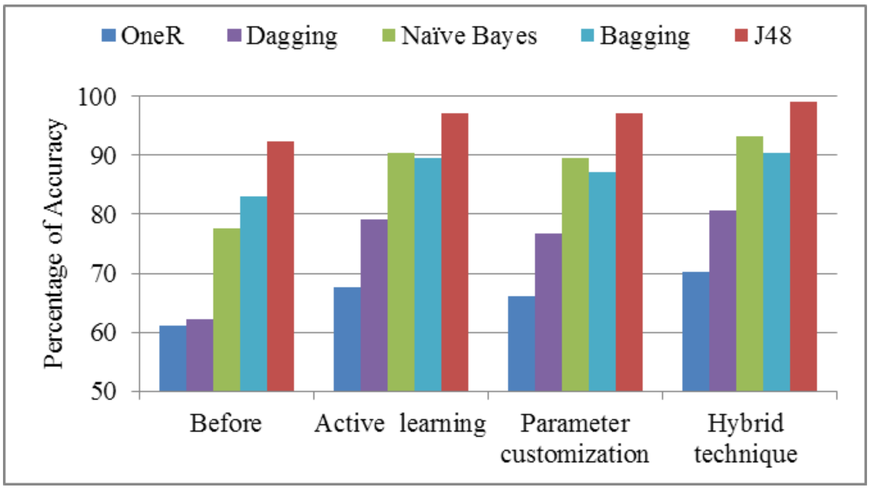


Fig. 4. Accuracy achieved by different methods when using hybrid technique and its components

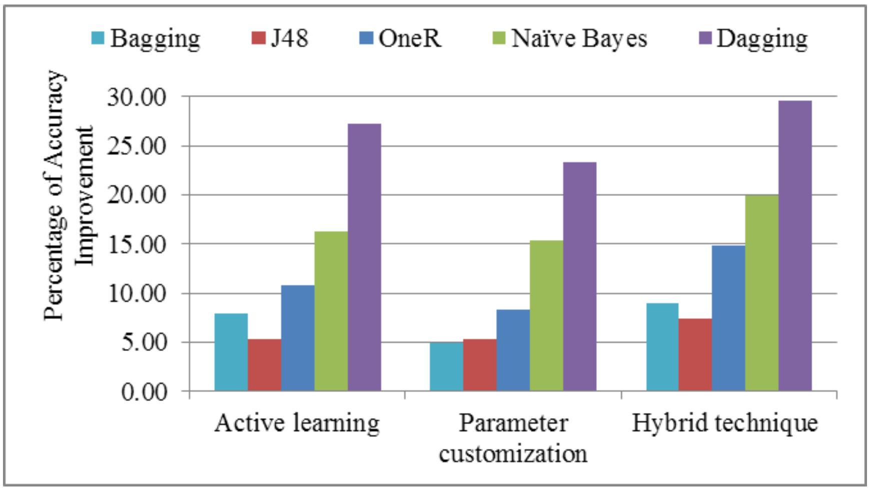


Fig. 5. Accuracy improvement achieved by hybrid techniche and its components

These figures clearly show that the active learning, feature selection, 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 14.89, 7.37, 19.95, 29.58, and 8.92 % 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 eight representative studies were chosen for this comparison, the including patient-adaptive model (PAM) [10], Fourier transform and neural network (FTNN) [13], dynamic learning and parameter tuning with decision tree (DLPTT) [39], statistical features and fuzzy hybrid neural network (SFHNN) [41], principle component with independent component analysis (PCICA) [22], wavelet transform and neural network (WTNN) [15], ECG classification by combining three different kinds of features and neuro-fuzzy network (FNFN) [55], and independent component analysis with neural network (ICANN) [23]. **Table 3** summarizes the comparative results of these methods, in which the last row lists the results of our model. Among the eight methods, the proposed method outperforms the other methods with an impressive accuracy of 99.1% in discriminating 15 ECG beat types.

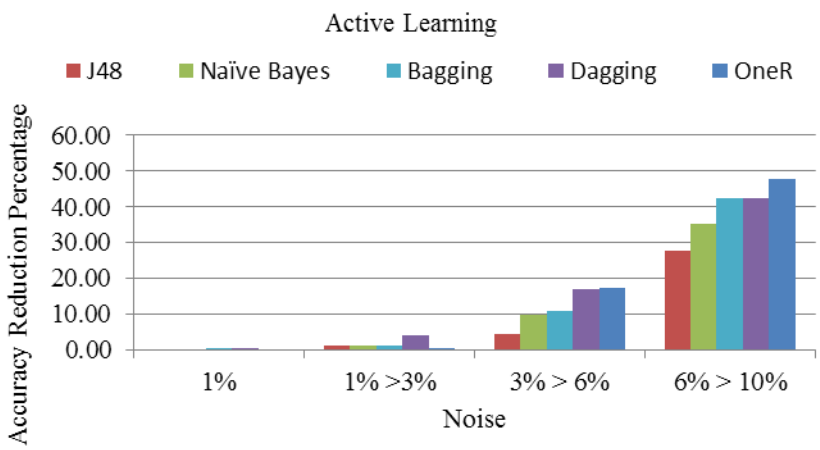
Table 3. Accuracy comparison with other methods

Method	*Number of arrhythmia*	*Accuracy %*
PAM	4	94.0
FTNN	3	98.0
DLPTT	15	98.1
SFHNN	7	96.1
PCICA	5	85.0
WTNN	13	96.8
FNFN	4	98.0
ICNNN	8	98.7
Proposed model	15	99.1

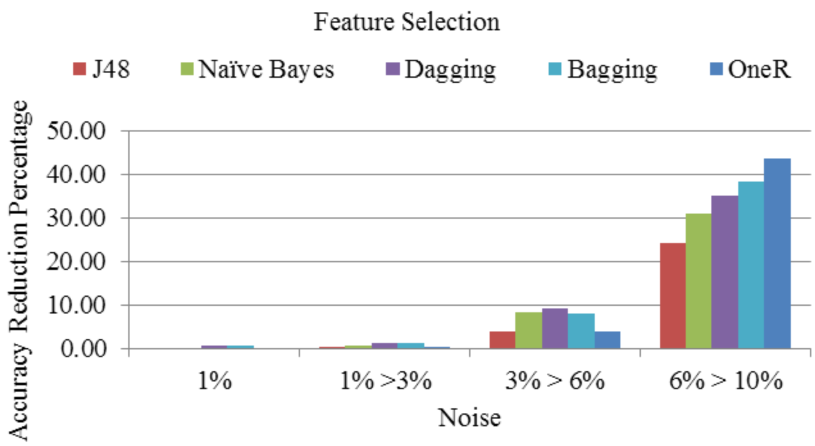
4.4 Noise Effects

The performance of the proposed hybrid technique and its components was further tested in the presence of noisy data. For this purpose, random noise was applied to the datasets for both training and validation. Different noise levels were investigated: 1%, greater than 1%, and less than or equal to 3%, greater than 3%, and less than or equal to 6% and greater than 6% and less than or equal to 10%. Furthermore, the noise was applied to all ECG parameters: P, QRS, and T waves. The measurement of the accuracy after applying each degree of noise was calculated by taking averages.

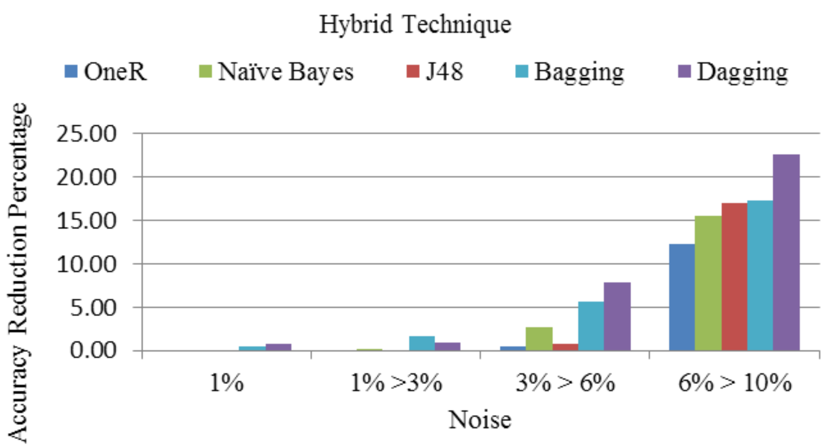
The results obtained in the presence of the noisy data, as presented in **Fig. 6,** show a reduction in accuracy in the presence of noise. The experiment was also conducted using the same methods OneR, J48, naïve Bayes, bagging, and dagging. We found that OneR performed the worst when using the individual components alone. However, dagging was the worst with the hybrid technique. Furthermore, J48 was the best when using active learning, and feature selection methods, and the hybrid technique.



(A)



(B)



(C)

Fig. 6. Accuracy reduction under noise

The performance of the hybrid technique, active learning, and feature selection methods at different noise levels are summarized in **Table 4**.

Table 4. **S**ummary of the reduction in accuracy for hybrid technique and its components under different levels of noise

Method	*1%*	*1% >3%*	*3% > 6%*	*6% > 10%*
Active learning	no	no	small	Very large
Feature Selection	no	no	Very Small	Very large
Hybrid technique	no	no	Very small	Large

As seen in the table, 1% noise had no influence on the performance of the hybrid technique, active learning, or feature selection methods. Noise levels greater than 1% and less than or equal 3% also has no affects on the performance of active learning neither feature selection nor the hybrid technique. Between 3% and 6%, the performance of active learning start declining significantly while feature selection was only affected slightly, but the hybrid technique very slightly. At between 6% and 10% noise, however, all methods were significantly affected. Generally, active learning is very sensitive to noise because of the mechanism of learning, which involves selecting a proper group of training data. Thus, when searching a data set that does not matching the overall data at all, the performance decreases significantly. Feature selection is much more robust to noise because there are always a limited number of features. The performance is also much better with the hybrid technique, since it is a combination of active learning and feature selection. Both components are implemented in parallel, and so there is no need for synchronization. Thus, it is very rare that noises affect the same features and are evaluated in the wrong way by both methods.

4.5 Speed

Fig. 7 shows the training and validation times for the J48 classifier with the three methods: active learning, feature selection, and the hybrid technique. As can be seen, feature selection greatly reduces the computational time for training and validation. Analogously, a smaller number of training samples also leads to a decrease in time required for classifying unknown samples. However, active learning takes the most computational time because the process of selecting the right group of data is very complicated. Consequently, the speed of the hybrid technique is affected by the negative performance of active learning.

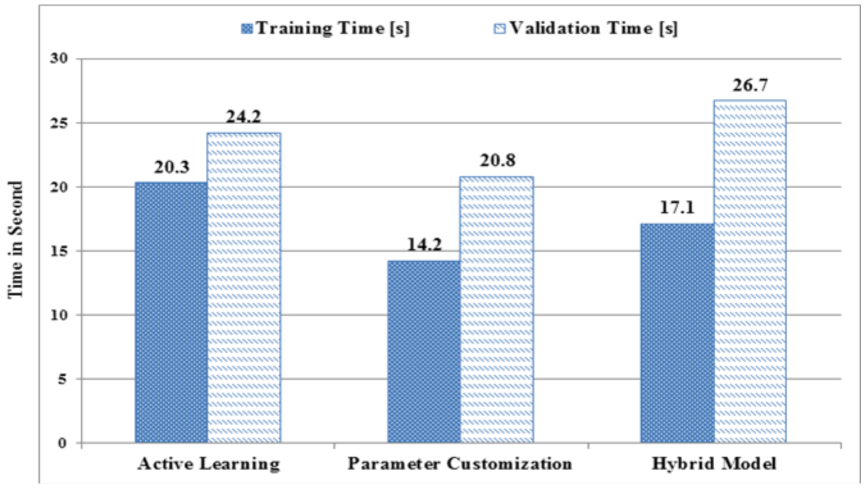


Fig. 7. Training and validation time to detect 15 arrhythmias using hybrid technique and its components

5. Conclusion

Cardiac health monitoring is a challenging problem in the field of data mining and knowledge extraction, and has received considerable attention over the past few years because of its importance in saving lives and reducing health risks. Today, cardiac health monitoring has reached a level of maturity when operating directly on or off-line. However, current methods are far from adequate for automated, remote cardiac health monitoring by detecting arrhythmias in real time. This is partly because of inter- and intra-patient variability. Thus, developing one classifier model to satisfy all patients in different situations using static training datasets is not practical. Furthermore, analyzing the QRS, P-wave, and other elements of ECG, and measuring the time interval between these elements, is necessary for real-time cardiac monitoring. This is technically infeasible with current systems because of computational limitations.

In this paper, we presented a hybrid technique as a proposed solution to solve these problems. 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.

References

[1] M. D. Dale Dubin, *rapid interpretation of EKG's*, 6th Edition, Cover Inc, USA, 2000.

[2] M. E. A. Bashir, M. Akasha, D. G. Lee, Min Yi, K. H. Ryu, E. J. Bae, M. Cho, and C. Yoo, " Highlighting the Current Issues with Pride Suggestions for Improving the Performance of Real Time Cardiac Health Monitoring," in *Proc. of 1st int. conf. on Inf. Technology. In Bio and Medical Informatics*, pp. 226-233, 30 August - 3 September, 2010.

[3] M. E. A. Bashir, K. S. Ryu, S. H. Park, D. G. Lee, J-W. Bae, H. S. Shon, and K. Ho Ryu, "Superiority Real-Time Cardiac Arrhythmias Detection using Trigger Learning Method," in *Proc. of 2nd Int. conf. on Inf. Technology. In Bio and Medical Informatics*, pp. 53-65, August 29 - September 2, 2011.

[4] M. E. A. Bashir, G. Min Yi, P. Minghao, H. S. Shon, and K. Ho Ryu, "Fine-tuning ECG Parameters Technique for Precise A bnormalities Detection," in *Proc. of Int. Conf. on Bioscience, Biochemistry and Bioinformatics*, pp.305-309, February 26-28, 2011.

[5] M. E. A. Bashir, K. S. Ryu, S. H. Park, D. G. Lee, J-W. Bae, H. Sun Shon, and K. Ho Ryu, "Pro-Detection of Atrial Fibrillation with ECG Parameters Mining Technique," in *Proc. of Int. Conf. on Electrical and Electronics Eng.*, December 22-23, 2011.

[6] G. Bortolan, I. Jekova and I. Christov, " Comparison of four methods for premature ventricular contractions and normal beats clustering," *IEEE Comp. in Card.*, vol. 32, , pp.921-924, 2005.

[7] U. Rajendra, P. Subbann, S. Iyengar, A. Raod, and S. Dua, "Classification of heart rate data using artificial neural network and fuzzy equivalence relation," *Patt. Recog.*, vol. 61 pp.61-68, 2003.

[8] M. E. 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. of 1. World Cong. in Comp. Sci. Comp. Eng. and Appl. Comput.*, pp.519-5225, July 12-15, 2010.

[9] G. Clifford, F. Azuaje, P. McSharrg, *advanced methods and tools for ECG data analysis*, Artech house, 2006.

[10] Y. H. Hu, S. Palreddy, and W. J. Tompkins, Eds., "Patient adaptable ECG beat classification using mixture of experts," in *Neural Network for Signal Processing V. Piscataway, NJ: IEEE Press*, vol. 44, no. 9, pp.495–463, 1995.

[11] O. Pichler, A. Teuner and B.J. Hosticka, "A comparison of texture feature extraction using adaptive Gabor filtering, pyramidal and tree structured wavelet transforms," *Pattern Recognition*, vol. 29, no. 5, pp.733-742,1996.

[12] K. Minami, H. Nakajima, and T. Toyoshima, "Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network," *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 2, pp. 179-185, 1999.

[13] Z. Dokur, T. Ölmez and E. Yazgan, "Comparison of discrete wavelet and Fourier transforms for ECG beat classification," *Elect. Lett.*, vol. 35, no. 18, pp.1502-1504, 1999.

[14] G. K.Prasad, and J. S Sahambi, "Classification of ECG arrhythmias using multi-resolution analysis and neural networks," in *Proc. of IEEE Conference on Convergent Technologies*, pp.227-231, October 15-17, 2003.

[15] K. Sternickel, "Automatic pattern recognition in ECG timeseries," *Comput. Meth. and Prog. in Biomed.*, vol. 68, pp.109-115, 2002.
[16] Z. Dokur and T. Olmez, "ECG beat classification by a novel hybrid neural network," *Comput. Meth. And Prog. In Biomed.*, vol. 66, , pp. 167–181, 2002.
[17] X. Wang and K. Paliwal, "Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition," *Pattern Recognition*, vol. 36, pp.2429-2439, 2003.
[18] Y. Ozbay, R. Ceylan and B. Karlik, "A fuzzy clustering neural network architecture for classification of ECG arrhythmias," *Comp. in Biology and Medicine*, vol. 36, pp. 376-388, 2006.
[19] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Tran. on Neural Networks*, vol. 10, no. 3, pp. 626-634, 1999.
[20] Z. Wang, Z. He and J. Z. Chen, "Blind EGG separation using ICA neural networks," in *Proc. of 19th int. conf. of the IEEE-EMBS*, pp.1351-1354, October 30 – November 2, 1997.
[21] C. Thomas, R. Harshman, and R. Menon, "Noise reduction in BOLDbased fMRI using component analysis," *Neuroimage*, vol. 17, pp.1521–1537, 2002.
[22] M. I. Owis, AB. Youssef, and Y M. Kadah, "Characterization of ECG signals based on blind source separation," *Medical and Biological Eng. and Computing*, vol. 40, pp.557-564, 2002.
[23] Y. Sung-Nien, C. Kuan-To, "Integration of independent component analysis and neural networks for ECG beat classification," *Expert Systems with Applications*, vol. 34, pp.2841-2846, 2008.
[24] L. S¨ornmo and P. Laguna, "*Bioelectrical Signal Processing in Cardiac and Neurological Applications*", Inc. Hoboken, New Jersey, 2005.
[25] J. de Bie, "P-wave trending: A valuable tool for documenting supraventricular arrhythmias and AV conduction disturbances," in *Proc. of IEEE Conf. on Comp. in Card.*, pp.511-514, September 23-26, 1990.
[26] J. Millet, M. Pkrez, G. Joseph, A. Mocholi, and J. Chorro, "Previous identification of QRS Onset and Offset is not essential for classifying QRS complex in a single lead," *Com. In card.*, vol. 24, , pp.299-302, 1997.
[27] G. Moody and R. Mark, "QRS Morphology Representation and Noise Estimation using the Karhunen-Loève Transform," in *Proc. of IEEE Conf. on Comp. in Card.*, pp.269-272, September 19-22, 1989,
[28] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sörnmo, "Clustering ECG complex using Hermite Functions and selforganizing maps," *T. On B. med. Eng*, vol. 47, pp.838–848, 2000.
[29] L. Senhadii, G. Carrault, J. Bellanger, and G. Passariello, "Comparing wavelet transforms for recognizing cardiac patterns," IEEE Eng. in Med. and Bio. Mag., vol. 14, no. 2, pp.167-173, 1995.
[30] G. Herrero, A. Gotchev, I. Christov, K. Egiazarian, "Heartbeat classification using independent component analysis and matching Pursuits," in *Proc. of int. Conf. on Acoustics, Speech, and Signal processing*, pp.725–728, March 18-23, 2005.
[31] I. Christov and G. Bortolan, "Ranking of pattern recognition parameters for premature ventricular contractions classification by neural networks," *Physiological Measurement*, vol. 25, no. 5, pp.1281-1290, 2004.
[32] I. M. Ishag, A. F. A. Dafa-Alla, G. Min Yi, D. G. Lee, J-W. Bae, K. Ho Ryu, "Agglomeration, Cluster accuracy, Hierarchical clustering. On-body sensor," *Advances in Experimental Medicine and Biology,* vol. 680, no. 1, pp. 83-88, 2010.
[33] S. Osowski and T. Linh, "ECG beat recognition using fuzzy hybrid neural network," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 11, pp.1265-1271, 2001.
[34] J. P. Abenstein, "Algorithms for real time ambulatory ECG monitoring," *Biomed. Sci. Instrum*, vol. 14, pp.73-79, 1978.
[35] H. G. Lee, K. Noh, B. J. Lee, H. S. Shon, K. Ho Ryu, "Cardiovascular Disease Diagnosis Method by Emerging Patterns," *Advanced Data Mining and Applications* ,vol. 4093, pp.819-826, 2006.
[36] M. Chen, S. Gonzalez, A. Vasilakos, H. Cao and V. C. M. Leung, "Body Area Networks: A Survey", ACM/Springer Mobile Networks and Applications (MONET), Vol. 16, No. 2, pp. 171-193, April 2011.

[37] S. Gonzalez, M. Chen, V. Leung, "Mobility Support for Health Monitoring at Home Using Wearable Sensors", IEEE Transactions on Information Technology in Biomedicine, vol.15, no.4, pp.539-549, July 2011.

[38] G. Bortolan, R. Degani, and J. L. Willems, "ECG classification with neural networks and cluster analysis," in *Proc. of IEEE Conf. on Com. In Card.*, pp.177-180, September 23-26, 1991.

[39] M. E. A. Bashir, D. G. Lee, M. Li, J. W. Bae, H. S. Shon, M. C. Cho, and K. H. Ryu, "Trigger Learning and ECG Parameter Tuning for Real-Time Cardiac Clinical Information System, IEEE Trans. In Bio Med., Vol. 16, No. 4, pp.561-571, July 2012

[40] S. Osowski, and T. H. Linh, "ECG beat recognition using fuzzy hybrid neural network," *IEEE Trans. on Bio. Engineering*, vol. 48, no. 11, pp.1265-1271, 2001.

[41] T. Yang, B. Devine, and P. Macfarlane, "Artificial neural networks for the diagnosis of atrial fibrillation," *Med. and Bio. Engineering and computing*, vol. 32, no. 6, pp.615-619, 1994.

[42] A. Kampouraki, G. Manis, and C. Nikou, "heartbeat time series classification with support vector machines," *IEEE Inf. Tech. in Biomedicine*, vol. 13, no, 4, pp.512-518, 2009.

[43] F. Sheng-wei, "Diagnostic study on arrhythmia cordis based on particle swarm optimization-based support vector machine," Expert Systems with Applications, vol. 37, no. 10, 2010, pp.6748–6752.

[44] 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, 2004.

[45] A. Kampouraki, G. Manis, and C. Nikou, "heartbeat time series classification with support vector machines," Eng. In Med. And Bio. Sc., 2009, pp. 512 – 518

[46] K. Birman, "Rule-Based Learning for More Accurate ECG Analysis," *Tran. On Puttern analysis and Mach. Intelligence.*, vol. 4, no. 4, pp.369-380, 1982.

[47] 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.

[48] 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 Systems with Applications, vol. 39, 2012, pp. 2947–2955.

[49] S. Mitra, M. Mitra, B.B. Chaudhuri, A rough set-based inference engine for ECG classification, IEEE Trans. Instrum. Meas., vol.55 (December), 2006, pp. 2198–2206.

[50] I. M. Ishag, A. F. A. Dafa-Alla, G. Min Yi, D. G. Lee, J-W. Bae, K. Ho Ryu, "Agglomeration, Cluster accuracy, Hierarchical clustering. On-body sensor," Advances in Experimental Medicine and Biology, vol. 680, no. 1, 2010, pp. 83-88.

[51] H. G. Lee, K. Noh, B. J. Lee, H. S. Shon, K. Ho Ryu, "Cardiovascular Disease Diagnosis Method by Emerging Patterns," Advanced Data Mining and Applications ,vol. 4093, 2006, pp.819-826.

[52] M. E. A. Basheer, K. S. Ryu, U. Yun, and K. H. Ryu, "Pro-Detection of Atrial Fibrillation Using Mixture of Experts," will be published to IEICE Transaction, 2012.

[53] UCI Machine Learning Repository, "http://www.ics.uci.edu/~mlearn/MLRepository.html".

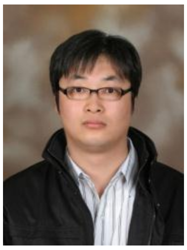
[54] WEKA web site, "http://www.cs.waikato.ac.nz/~ml/weka/index.html".

[55] M. Engin, "ECG beat classification using neuro-fuzzy network," *Pattern Recognition Letters*, vol. 25, pp.1715-1722, 2004.

Mohamed Ezzeldin A. Bashir received a Post Graduate Diploma and M.Sc. Degree in Computer Science and Information Technology from University of Gezira, Sudan, Also received M.Sc. Degree in Business Administration, from University of Medical Sciences and Technology, Sudan, and PhD from the School of Electrical & Computer Engineering, in Chungbuk National University, South Korea. He is assistant professor at University of Medical Sciences and Technology, Sudan. His research interests include ubiquitous computing and stream data processing, knowledgebase information retrieval, data mining, information retrieval, database systems, Bioinformatics, and Biomedicine.

Ho Sun Shon received a BS in Statistics from the Sung Shin Women University of Nature Science, South Korea in 1986. And she obtained her MS Degree in Statistics from Sung Shin Women University, South Korea in 1992. And she obtained her PhD Degree in Computer Science from Chungbuk National University, South Korea in 2010. Her research interests are in the area of bioinformatics, data mining and biomedical statistics.

Dong Gyu Lee received the bachelor degree of engineering from department of information communication engineering, Seowon University, Korea, in 2005. He received the master degree of engineering from department of computer science, Chungbuk National University in 2007. Also, he studied as research student in University of Aizu, Japan in 2011. And currently he is a Ph.D. student in Database & Bioinformatics laboratory, department of computer science, Chungbuk National University. His research interests include data mining, medical informatics, and geographic information systems.



Hyeong soo Kim received a Bachelor of Computer and obtained Master's degree of Computer Science from Chungbuk National University, South Korea in each of 2008 and 2010. He is currently studying to get doctoral degree at Chungbuk National University. His research interests include GIS, data mining, Medical mining and Geo-healthcare.

Keun Ho Ryu received the Ph.D degree from Yonsei University, Korea, in 1988. He is a professor at Chungbuk National University in Korea and a leader of database and bioinformatics laboratory. He served Korean Army as ROTC. He worked not only at University of Arizona as Post-doc and research scientist but also at Electronics& Telecommunications Research Institute, Korea. His research interests are included in temporal databases, spatiotemporal databases, ubiquitous computing and stream data processing, knowledgebase information retrieval, database security, data mining, biomedical and bioinformatics. He is a member of the IEEE and member of the ACM, since 1983