IET Signal Processing

Research Article

Hybrid hierarchical method for electrocardiogram heartbeat classification

ISSN 1751-9675
Received on 15th March 2017
Revised 16th November 2017
Accepted on 1st December 2017
E-First on 23rd March 2018
doi: 10.1049/iet-spr.2017.0108
www.ietdl.org

Hadeer El-Saadawy[1] ✉, Manal Tantawi[1], Howida A. Shedeed[1], Mohamed F. Tolba[1]
[1]Scientific Computing, Ain Shams University, Abbassia, Cairo, Egypt
✉ *E-mail: hadeer_ibrahim@cis.asu.edu.eg*

Abstract: This paper proposes an automatic reliable two-stage hybrid hierarchical method for ECG heartbeat classification. The heartbeats are segmented dynamically to avoid the consequences of the heart rate variability. Discrete Wavelet Transform (DWT) is utilized to extract morphological features. The extracted features are then reduced by using Principle Component Analysis (PCA). Subsequently, the resulted features along with four RR features are fed into Support Vector Machine (SVM) to classify five categories. Thereafter, the heartbeats are further classified to one of the classes belonging to the assigned category. Two different strategies for classification have been investigated: One versus All and One versus One. The proposed method has been applied on data from lead 1 and lead 2. A new fusion step is introduced, where stacked generalisation algorithm is applied and different types of classifiers have been examined. Experiments have been carried out using a MIT_BIH database. The best overall and average accuracies obtained by the first stage are 98.40% and 97.50% respectively. For the second stage, 94.94% and 93.19% are the best overall and average accuracies obtained respectively. The best results are achieved using SVM with one versus one classification strategy for both stages and decision trees classifier for the fusion step.

1 Introduction

The electrocardiogram (ECG) records the electrical activity of the heart over time. Each recorded heartbeat shows certain characteristics that are prevalent in all normal subjects. Most clear amongst these characteristics are a series of excursions from a baseline which reflects the contraction and relaxation of the heart chambers. These excursions manifest as waves that are readily visible in the ECG. In particular, a typical ECG heartbeat will include three successive prominent waves called P, QRS and T waves [1]. These waves have been employed to evaluate the cardio-physiological state of health of an individual in clinical situations for over a century. Cardiac arrhythmias can be detected by any change in the characteristics of any of these three waves [2, 3].

Cardiac arrhythmias means abnormal activities in the heart upon certain conditions and mainly consists of two types. One of them is life threatening and can cause death, for example ventricular fibrillation and flutter [4]. This type has been well investigated in [5–7]. On the other hand, the other type is cardiac arrhythmia which is our interest in this study. It needs attention to avoid deterioration, but it is not critically as life threatening as the first one [2].

Hence, ECG heartbeats should be continuously examined and classified [8]. However, beat by beat manual classification is time consuming and too difficult of a process [2]. Moreover, arrhythmias needs accurate and early detection, especially the life threatening types [9]. Thus, the automation of ECG investigation is a very demanding issue [10, 11] which has resulted in a lot of publications [2, 9–19] with promising results. However, most of the existing studies consider only a few arrhythmia classes. In addition, although ECG is recorded from different positions using leads, the data of only one lead is utilised by most of the studies. Finally, most of the existing studies have utilised MIT_BIH database [20] for evaluation and used the overall accuracy measure which is not reliable enough since the data available for the different classes is not equally distributed. Thus, some classes with large data will dominate the overall result.

In this paper, an automatic ECG diagnosis two-stage hybrid hierarchical method is proposed. The proposed method considers all the 15 classes that exist in MIT_BIH database. The heartbeats are dynamically segmented to avoid heart rate variability. The segmented beats are then decomposed using discrete wavelet transform (DWT) and projected to low-dimension sub-space by using the principle component analysis (PCA) algorithm. The resulted features are concatenated with four dynamic RR interval features and fed into support vector machine (SVM) to be generally classified into five main categories. Thereafter, the heartbeats are transferred to the second stage of the proposed hierarchy to be specifically classified as one of the classes that belongs to the category determined from the previous stage for each heartbeat. Two strategies: one versus all and one versus one for classification have been examined in this study. The proposed method has been applied to data from the two leads 1 and 2 provided by the MIT_BIH database. A new fusion step based on a stacked generalisation method is suggested by this study to benefit from the information of both leads. Finally, this study not only considers the overall accuracy but also the average accuracy of each category to guarantee a robust reliable method for ECG classification.

The rest of this paper is organised as follows. Section 2 provides a review of the existing studies. The proposed method is discussed in detail in Section 3. Section 4 presents the achieved results and finally, Section 5 provides the conclusion and future work.

2 Related work

The automation of analysing a rich source of information like ECG for diagnostic purposes is very crucial, since it helps 24 h monitoring and instant discovering of cardiac disorders which needs rapid medical aid in clinical situations. Hence, in the last decade, many studies have emerged to investigate this issue intensively. In this section, a brief survey of some key published studies is presented below.

Alickovic and Subasi [19] utilised DWT, different classification algorithms and different datasets to classify only five classes. 99.33% is the best accuracy achieved using random forest classification algorithm and MIT-BIH dataset.

Yazdanian *et al.* [10] considered only five classes. Important points of each segmented heartbeat have been derived and then

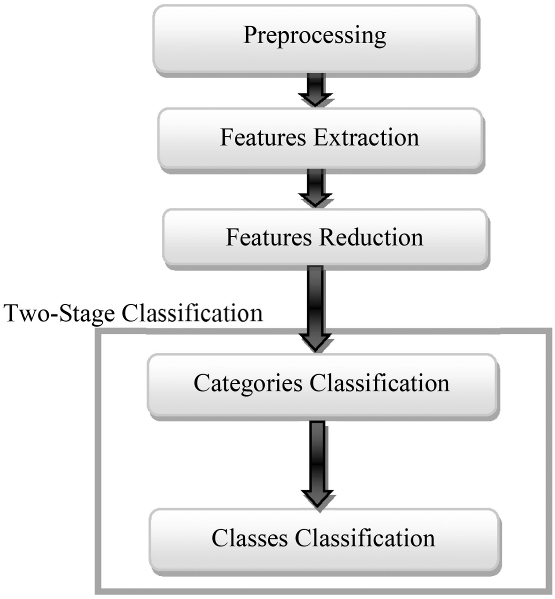

Fig. 1 *Main steps of the proposed hierarchical method*

Table 1 Training and testing percentages used in the experiments [2]

Heartbeat type	Training ratio, %	Training beats number
normal	13	9753
left bundle branch block	40	3229
right bundle branch block	40	2902
atrial premature contraction	40	1019
premature ventricular contraction	40	2852
aberrated atrial premature	50	75
ventricular flutter wave	50	236
fusion of ventricular and normal	50	401
blocked atrial premature	50	97
nodal (junctional) escape	50	115
fusion of paced and normal	50	491
ventricular escape	50	53
nodal (junctional) premature	50	42
atrial escape	50	8
unclassifiable	50	7
total	21.89	21,290

classified using SVM. An overall accuracy of 96.67% has been achieved using MIT-BIH as a validation database.

Bulusu *et al.* [12] proposed an improved morphological feature vector including ST-segment information to classify six different heartbeat types. The extracted features are then classified using SVM, resulting in an overall accuracy of 96.35% using MIT-BIH as a validation database.

DWT and independent component analysis (ICA) techniques have been considered by Ye *et al.* [2] to extract the morphological features. RR dynamic features are also added to represent each heartbeat. SVM has been utilised for classification. The method has been applied to both lead 1 and lead 2 separately. Subsequently, a fusion step has been considered to achieve the final decision by using the rejection and Bayesian approaches. An 86.4% overall accuracy has been achieved using MIT-BIH as a validation database using subject oriented evaluation for classifying the five categories. On the other hand, a 99.3% overall accuracy has been achieved using the same methodology to classify the 16 classes based on class-oriented evaluation.

Features from the QRS wave along with DWT coefficients have been utilised by Martis *et al.* [9] to classify 15 classes mapped to five main categories. The extracted features have been reduced using three different algorithms ICA, PCA and linear discriminate analysis to feed into three different classifiers probabilistic neural network (PNN), neural network (NN) and SVM. The combination between ICA and PNN achieved the best results to get an overall accuracy 99.28% using MIT-BIH as a validation database.

Generally speaking, the following observations can be concluded from the existing studies in the literature: (i) most of the existing studies except [2, 9, 16] have considered very few classes (only three to six classes). (ii) MIT-BIH has been considered as a validation database by most studies. (iii) For heartbeat extraction, static heartbeat segmentation has been utilised by the existing studies, which is not robust enough for heart rate variations. (iv) Regarding the description of the heartbeats, different features have been investigated including high-order statistics features [21, 22], Hermite coefficients [21, 23, 24], linear discriminates [25–27] and wavelet features [2, 3, 28]. (v) Different classification algorithms have been utilised successfully such as: artificial neural network [24, 28], decision trees [21], self-organising map [23] and SVM [2, 10–12, 16]. (vi) Although lead 1 and 2 record ECG from different positions, which means they do not provide the same information, the existing studies except [2] have utilised information from only one lead and neglected the other. (vii) Finally, the observation that can be considered as a main drawback for all research in this application is that, only the overall accuracy has been considered as an evaluation measure to test the reliability of the system. However, it is totally inconvenient for the evaluation as the tested samples of MIT-BIH database are not equally represented. Some categories have thousands of testing samples (normal classes) while other categories (classes that represent cardiac disorders) have only hundreds. So the overall accuracy may be biased towards the categories with a huge amount of data. Motivated by these observations, in this work, a hybrid hierarchical method that is invariant to heart rate variability and considers all categories with their associated classes (not only few classes) has been proposed. The efficiency of the proposed method in recognising all classes has been carefully examined with information from both leads.

3 Methodology

The proposed hierarchical method consists of five main steps including a hybrid hierarchical classification step that encompasses two stages as shown in Fig. 1. Each step is applied separately on both lead 1 and lead 2 till the last step, where the results from the two leads are fused to get the final decision in each stage. A detailed description of the utilised database and the proposed method will be discussed in the next sub-sections.

3.1 Dataset

As previously mentioned in Section 2, the MIT-BIH database [20] is the most popular database utilised by the existing studies [2, 9–16] for training and testing purposes. The MIT-BIH database consists of 48 records, each of them is a 30 min long sample with a sampling frequency of 360 Hz but according to the ANSI/AAMI EC57: 1998 standard [29], only 44 records can be utilised as there are four paced records.

Moreover, attached with each record a file contains the annotations of the beats and the locations of the R-peaks. In this study, the mentioned given data is utilised as a ground truth in the training and testing processes. Moreover, the data from both lead 1 and lead 2 have been utilised to make the final decision.

The data of the 44 records mentioned above is divided into training and testing portions. The data division has been done exactly as proposed in [2, 16] for comparison. The total number of samples for each class is not equally distributed which is why the percentage of the training and testing portions are not the same in all the classes. The training percentage for the normal class is 13% of the total number of beats as it contains thousands of numbers of beats, 40% for the classes that have lower number of beats and only 50% for the classes that have a very low number of beats as shown in Table 1. Furthermore, according to the ANSI/AAMI EC57: 1998 standard [29], the mentioned classes in Table 1 are then mapped to five main categories as shown in Table 2.

3.2 Pre-processing

The aim of this step is to improve the signal-to-noise ratio and extract the heartbeats. Noise has been reduced by removing both high and low frequencies out of the ECG spectra. Hence, Butterworth bandpass filter with a range 0.5–40 Hz is applied and subsequently, ECG records are segmented into heartbeats. Each heartbeat encompasses three main waves P, QRS and T waves. In the literature, due to the difficulty of the detection of these three waves, a fixed segmentation strategy is usually applied. However, it is not reliable, since it cannot take into consideration the heart rate variations.

Accordingly, in this study, a dynamic segmentation strategy invariant to heart rate variability and as proposed in our previous work [3] is followed. The number of samples considered before and after the R peak is derived according to (1). Fig. 2 shows an example for a segmented heartbeat where the previous RR is the number of samples between the current R peak and the previous one. On the other hand, the next RR is the number of samples between the current R peak and the next one

$$\text{Before R peak} = \frac{1}{3} * \text{Max}(\text{RR previous}, \text{RR next}) \tag{1a}$$

$$\text{After R peak} = \frac{2}{3} * \text{Max}(\text{RR previous}, \text{RR next}) \tag{1b}$$

Finally, all the heartbeats are then adjusted to have the same length of 300 samples (0.83 s according to sampling rate) as done in [2, 3].

Table 2 Mapping to the five main categories according to the ANSI/AAMI EC57: 1998 standard [2]

ANSI/AAMI classes	MIT-BIH classes
N	NOR, LBBB, RBBB, AE, NE
S	APC, AP, BAP, NP
V	PVC, VE, VF
F	VFN
Q	FPN, UN

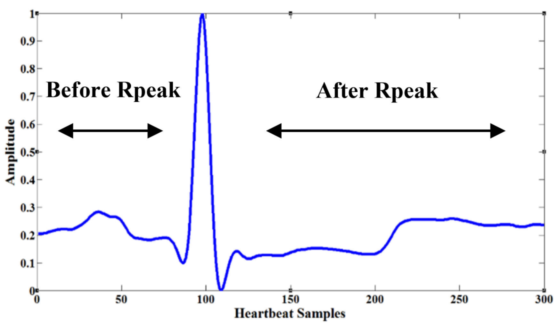

Fig. 2 *Segmented heartbeat using dynamic segmentation strategy*

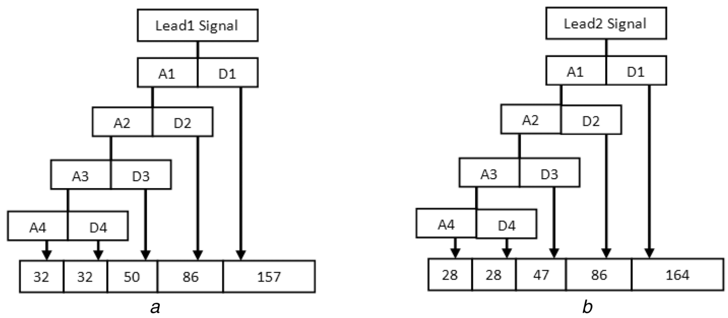

Fig. 3 *DWT levels and the resulted wavelet coefficients structure for* ***(a)*** Lead 1, ***(b)*** Lead 2

3.3 Features extraction

Both morphological and dynamic features are utilised in this paper to describe each heartbeat using DWT and RR features.

3.3.1 Discrete wavelet transform: Due to the non-stationary nature of the biomedical signals, DWT is preferable due to its efficiency to describe the QRS characteristics [2, 3].

Hence, different mother wavelet families (i.e. Biorthogonal and *Daubechies*) have been examined to get the best description for the characteristics of the QRS wave formed by each lead separately. The ECG leads are placed on different positions on the patient's limbs and on the chest surface. Consequently, each lead from its position can display some valuable details in the three heartbeat waves (P, QRS and T) that do not appear in the other leads. Hence, the shape of QRS complex in lead 1 carries some information that does not appear in lead 2 and vice versa. Thus, each lead can have its best-fit mother wavelet. Our initial experiments have revealed the efficacy of Daubechies of order 8 (db8) and Biorthogonal (bior 2.4) mother wavelets in representing the heartbeats formed by lead 1 and lead 2, respectively. Following [2, 3], a four-level decomposition is applied to both leads. The first level decomposes the heartbeat into both low (approximation A1) and high (details D1) parts and then the approximation part is further decomposed into approximation A2 and detailed parts D2 of the second level and so on until the last level. The resulted coefficients from all detailed parts and the approximation of the last level, construct the corresponding wavelet coefficients as shown in Figs. 3*a* and *b*. In this study, wavelet decomposition is applied using Matlab DWT function. This DWT function has the advantage that it is not limited to dyadic length and is based on a simple scheme: convolution and down-sampling. As usual, finite-length convolution causes border distortions. Hence, few extra coefficients are computed at each stage of the decomposition process yielding a number of coefficients that is not power of 2 [30].

3.3.2 RR features: The RR features have been successfully considered by Ye *et al.* [2]. Thus, besides the DWT coefficients, RR features are also examined in this study to describe the dynamic features for each heartbeat. Hence, following Ye *et al.* [2], four main RR features are considered as follows: previous RR, post RR, local RR and average RR intervals. Previous RR is the interval between the current RR and the previous one; the post RR is the interval between the current one and the next one; local RR is obtained by calculating the average of all the RR intervals in the past 10 s episodes and finally the average RR interval is obtained by calculating the average of all the RR intervals in the past 5 min [2].

3.4 Features reduction

The features reduction step is needed due to the high dimension of the extracted features which may cause redundancies and irrelevant information that may affect the classification results. PCA is suggested in the proposed method to find a sub-space that preserves most of the context of the features. Thus, PCA is applied only on the features extracted from DWT step. The resulted

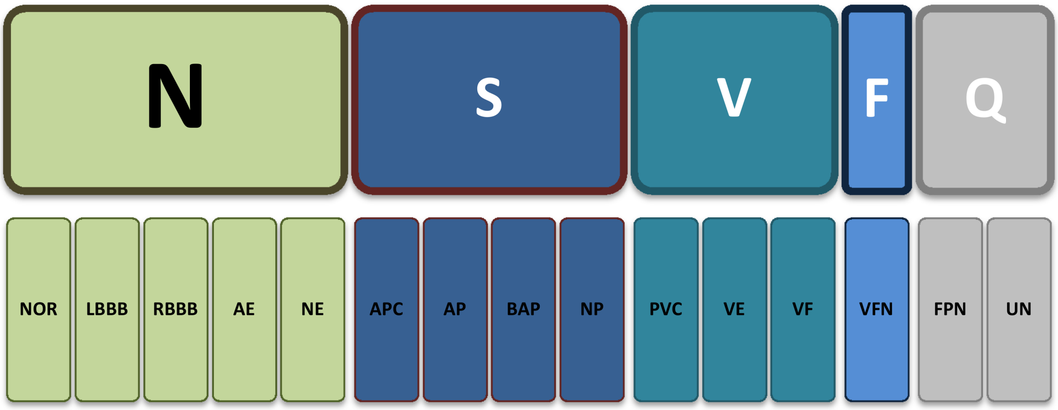

Fig. 4 *Block diagram that shows the classes associated with each category*

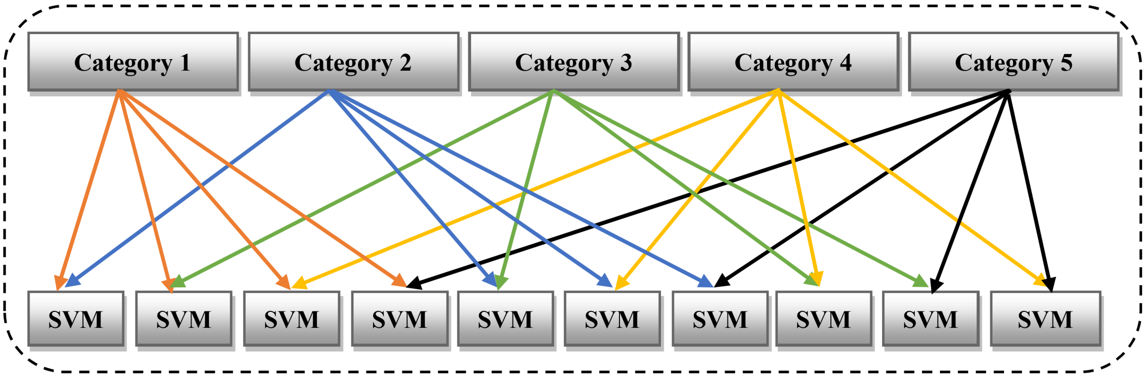

Fig. 5 *One versus one SVM classification strategy for first stage*

projected features (reduced) are then concatenated with the RR features for the classification step.

3.5 Classification

SVM is a binary classifier proposed by Vapnik [31]. However, recently it is applied for the discrimination of more than two classes. SVM separates the classes by building a hyper plan between them. Consider that there are N data samples, each of them consists of a feature vector and its ground truth $\{(x_i, y_i), i = 1, ..., N\}$. A decision function can be constructed after the training process by using the already existing data with their ground truth as mentioned in [2]

$$f(x) = \text{sign}\left(\sum \alpha i y_i K(x_i, x) + b\right) \tag{2}$$

where αi represents the Lagrange multiplier for each data sample in the training set, K (.) represents the kernel function.

In this study, a two-stage hierarchical classification process is introduced. In the first stage, the heartbeat is assigned to one of the five main categories as mentioned in Table 2. Subsequently, the heartbeat is further classified into one of the classes that belong to its category known from the previous stage as shown in Fig. 4. SVM has been utilised in both stages for classification purposes using two different strategies: namely one versus all and one versus one which can be defined as follows.

3.5.1 One versus all: One versus all is the strategy applied in most of the previous work [2, 10–12, 16]. In this work, there is only one SVM to classify the five categories in the first stage and then in the second classification stage there is one SVM to classify the classes of each category. In this strategy, the Gaussian function is used as a decision function, since it results in the best accuracies in the literature [2, 3].

3.5.2 One versus one: One versus one is another suggested strategy for classification. An SVM classifier is considered for every two classes. For example, regarding the first proposed stage, the classification process is done in five main categories. Thus, the total possible permutations for all categories yield ten classifiers. Each of the ten classifiers is trained separately using the data corresponding to its two considered categories as shown in Fig. 5. Then the testing process is applied by projecting all the testing data to ten classifiers to get ten different responses. The final decision is the dominant class that has the largest number of votes. For the second stage, according to the number of classes in each category, a number of binary SVMs has been utilised in the same way as discussed for the first stage.

3.6 Lead 1 and lead 2 fusion

In this paper, a new fusion step is proposed for each classification stage in order to take advantage of the information given by the two leads. The next sub-sections provide a detailed discussion about the accomplished fusion.

3.6.1 First stage fusion: Regarding the first stage, a stacked generalisation based method [32] is introduced to benefit as much as possible from the information derived from the two leads. The suggested fusion step is examined and compared to the rejection method utilised before in the literature [2, 16]. More details about the two fusion methods are explained as follows:

i. *Rejection method:* If the decision of the heartbeat from lead 1 and lead 2 is the same, the rejection method takes this beat into consideration. However, if the decision is different, the rejection method neglects this beat for a further manual classification process as shown in Fig. 6 which is considered as the main drawback for this method as the system will not be fully automated.
ii. *Stacked generalisation method:* Stacked generalisation [32] is used mainly to combine different models to enhance the predictive results. Its main idea is that the training data is divided into two disjoint sets.

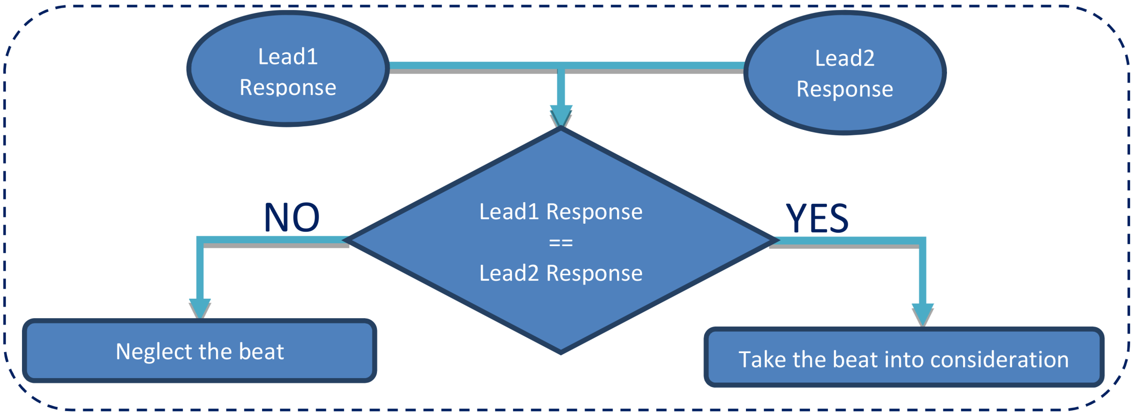

Fig. 6 *Rejection method*

Table 3 Number of beats used in testing the second level classification experiments

Heartbeat type	Testing beats number
normal 'NOR'	56,163
left bundle branch block 'LBBB'	2904
right bundle branch block 'RBBB'	2601
atrial premature contraction 'APC'	915
premature ventricular contraction 'PVC'	4037
aberrated atrial premature 'AP'	37
ventricular flutter wave 'VF'	118
fusion of ventricular and normal 'VFN'	200
blocked atrial premature 'BAP'	48
nodal (junctional) escape 'NE'	57
fusion of paced and normal 'FPN'	245
ventricular escape 'VE'	26
nodal (junctional) premature 'NP'	20
atrial escape 'AE'	4
unclassifiable 'UN'	4

The first part is used to train many base learners then the second part is used to test the trained learners. Subsequently, the results achieved from testing the trained learners are used as the inputs and the correct responses as the outputs to train a higher level learner and finally, the learners of the two stages are tested using the testing data. It seems like cross-validation but instead of using a winner-takes-all approach, the base learners are linearly or non-linearly combined [32].

In this study, the networks of the first stage are trained using the same training data portion used for the rejection method for comparison, while the data portion needed for training the second stage is taken from the testing data portion. Two SVM learners for the two leads are considered in the first stage. In addition, different network classifiers have been investigated for the learner of the second layer such as: SVM, PNN and decision trees.

i. *PNN:* Before projecting the testing data to the classification process, PNN network is built by projecting the input data to the input layer which consists of a number of nodes according to the number of features used in the classification. Each node is connected to all the nodes in the pattern layer, and then each node in the pattern layer is connected to only one node in the category layer which has a node for each class to get the weights that will be used in the testing process [9, 33].
ii. *Decision trees:* Decision trees are considered from the most used classifiers due to their structure as perfect classification can be achieved by a minimal number of decisions [34]. In this study, decision trees are applied to a different number of trees in order to get the optimum number that can achieve the best results. The decision trees are trained using the introduction to classification & regression trees (CART) algorithm [34].

3.6.2 Second stage fusion: Unfortunately, due to lack of data for some classes, the stacked generalisation method cannot be applied at this stage. Some classes as shown in Table 3 have only four beats which are not sufficient to be divided into two sets in order to apply stack generalisation. Hence, another fusion strategy should be considered. From the previous stage accomplished with stacked generalisation and our initial experiments regarding classification of classes, it has been found that if the category is correctly known, each lead is superior in providing enough information to classify all or most of the classes of some categories efficiently. Thus, lead 1 is considered for classifying classes of some categories, while lead 2 is considered for the others.

4 Experimental results

The aim of this proposed method is to increase the accuracy of each category/class separately besides the overall accuracy. For reliable results, tenfold cross-validation strategy is applied and the proposed method is evaluated by four main measures.

The first three measures are mentioned in literature and they are namely overall accuracy, sensitivity (Se) and positive predictivity (+P). On the other hand, the fourth measure is the average accuracy considered by this study to give the same significance to all the classes (any of five categories or the 15 class). The four measures are computed as follows:

$$\text{Overall accuracy} = \frac{\text{Number of correctly classified beats in all classes}}{\text{Total number of all beats for all classes}} \tag{3}$$

$$\text{Sensitivity (Se)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

$$\text{Positive predictivity (+P)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

$$\text{Average accuracy} = \frac{\sum_{1}^{\text{NC}} \text{Accuracy of each category}}{\text{NC}} \tag{6}$$

where the TP (true positive) is the instances from a given class that are correctly classified as from that class; FN (false negative) is the instances from the given class that are incorrectly classified to another class; FP (false positive) is the instances of the other classes that are incorrectly classified to the given class; NC is the number of considered classes (five categories, classes of a given category or all the 15 classes).

The proposed method is applied separately on both lead 1 and lead 2. Four-level wavelet decomposition is applied on the segmented beats using db8 and Bior 2.4 mother wavelets (as discussed in Section 3.3.1) for lead 1 and lead 2, respectively. Initially, the resulted coefficient structure is reduced by considering the coefficients that represent only the ECG spectra (0.5–40). Although all the ECG spectra can be found in the approximation of

Table 4 Lead 1 and lead 2 classification of ten trials average results using one versus all SVM classifier

	Without RR features				With RR features			
	Lead 1		Lead 2		Lead 1		Lead 2	
	Se, %	+P, %	Se, %	+P, %	Se, %	+P, %	Se, %	+P, %
N	97.7	83.7	97.0	82.1	98.7	86.03	98.5	89.6
S	84.3	93.9	81.2	95.97	91.2	98.39	89.5	98.2
V	95.5	84.1	90.2	81.97	95.3	87.1	93.8	86.3
F	74.6	98.04	74.7	99.5	76.1	99.6	77.0	99.7
Q	91.4	92.5	94.6	91.6	90.7	94.3	93.3	91.8
average	88.7		87.5		90.4		91.0	
overall	97.1		96.1		98.2		97.9	

Table 5 Lead 1 and lead 2 classification of ten trials average results using one versus one SVM classifier

	Without RR features				With RR features			
	Lead 1		Lead 2		Lead 1		Lead 2	
	Se, %	+P, %	Se, %	+P, %	Se, %	+P, %	Se, %	+P, %
N	96.7	86.7	96.3	82.5	98.6	86.1	98.7	88.1
S	89.4	96.2	86.3	94.98	93.2	98.1	91.7	97.5
V	96.1	91.1	91.3	89.9	95.4	92.7	94.8	92.1
F	88.7	97.7	87.5	97.5	88.7	98.1	88.0	98.2
Q	93.7	98.5	95.0	99.4	92.2	99.9	95.4	99.96
average	92.9		91.3		93.6		93.7	
overall	96.4		95.8		98.2		98.3	

the second level A2, two more levels are considered as mentioned above in order to get better frequency resolution. Our experiments have revealed that considering coefficients of D3 (details of level 3), D4 (details of level 4) and A4 (approximation of level 4) for the ECG spectra instead of only A2 provides the best accuracies for all classes. Thus, a vector of 114 (i.e. 32 from A4, 32 from D4 and 50 from D3) coefficients is considered as a feature vector for lead 1, while a vector of 103 (i.e. 28 from A4, 28 from D4 and 47 from D3) coefficients is considered as a feature vector for lead 2.

Subsequently, the PCA algorithm is applied to the feature set and the first m principle components have been empirically chosen. The optimum value of m equals to 18 for both leads. The feature vectors are projected using the chosen 18 PCA components resulting in 18 final features and are considered for classification with and without the RR features. The two different SVM strategies have been applied as mentioned in the previous section. Table 4 shows the average first stage results of ten trials using one versus all SVM before and after adding the features to examine their significance for both lead 1 and lead 2 separately.

It is clear that the overall accuracy is above 95% but the average accuracy is low since some classes (S and F) have low classification results. The addition of RR features significantly improves the average accuracy of category S.

This is medically justified by the fact that category S causes changes in the rhythm of the heartbeats which are characterised by the dynamic information provided by the RR features. Table 5 shows the average results of ten trials for lead 1 and lead 2 after applying the one versus one strategy on the projected features with and without the RR features.

Generally speaking, by comparing the results of Table 4 with Table 5, one versus one strategy is superior (above 4% improvement in the average accuracy), especially for abnormal categories.

Furthermore, by analysing the results of Table 5, it is also clear that adding the RR features improves the accuracies especially for the S category. Moreover, lead 1 provides better results for S and V categories. On the other side, lead 2 provides better results for the Q category.

Hence, these results encourage a fusion step between lead 1 and lead 2 in order to gather the benefits of both. Approaching this target, two fusion methods have been investigated by this study (discussed in Section 3.6.1). For the stack generalisation based method, the decision generated by each lead will be the input for a learner in an advanced layer that is non-linearly trained to give the final decision without rejecting any beats. Regarding the learner of the advanced layer, three architectures have been investigated: one versus all SVM, PNN and decision trees. Moreover, rejection method has been applied for comparison.

Table 6 shows the classification results by both methods. It is worth mentioning that ∼1.8% of the testing beats are rejected (excluded) using the rejection method. Moreover, although the overall accuracy is still the same, the average accuracy has dropped by more than 6%. Thus, these results reveal the deficiency of this method for fusion since it does not provide a mean to benefit from the fact that each lead is superior in providing information that characterises some categories more than others.

On the other hand, the results have revealed the superiority of decision trees for stack generalisation in providing the best accuracies.

The PNN has achieved the best overall accuracy. However, this result is misleading since the average accuracy is low (only the dominant category N has high accuracy). The average accuracy reaches its maximum value 97.5% with decision trees since the added fusion step has the ability to combine the best of utilising each of the two leads resulting in high classification for all the categories.

Table 7 shows a comparison between the proposed method and a related study that considered the same five main categories. The proposed method provides comparable overall accuracy with features invariant to heart rate variability and average accuracy for all categories, the measure that was not considered before by the existing studies. Moreover, in this study the overall accuracy 99.3% achieved by Martis *et al.* [9] is also achieved by using PNN classifier in the fusion step but the average was lower than that achieved using the decision trees classifier.

Regarding the second stage, the one versus one strategy is also superior in providing the best classification results for classes of each category for each lead separately. As discussed in Section 3.6.2, it has been observed that each lead provides better results for most of the classes in some categories. Hence, as shown in Table 8, lead 2 is considered for classes of categories N, S and Q, respectively. On the other hand, lead 1 is considered only for classes of category V. Category F is not considered in the second classification stage since it is only one class as mentioned in Fig. 4. Finally, Fig. 7 shows the final hierarchical two-stage classification scheme.

Finally, the second stage fusion step yields a 94.94% overall accuracy and 93.19% average accuracy. To the best knowledge of

Table 6 Fusion classification of ten trials average results using rejection method and stacked generalisation with different classification methodologies in the second layer

	Rejection fusion method		Stacked generalisation fusion method					
			SVM		PNN		Decision trees	
	Se, %	+P, %	Se, %	+P, %	Se, %	+P, %	Se, %	+P, %
N	99.0	80.4	99.7	90.1	99.9	60.7	98.5	94.9
S	92.0	99.4	83.3	99.5	87.1	100.0	96.8	99.4
V	93.5	93.6	91.4	97.3	89.9	99.9	96.6	98.04
F	92.25	99.7	66.3	97.4	75.2	99.89	98.0	97.4
Q	96.7	99.9	98.7	99.99	89.3	100.0	97.8	99.98
average	94.7		87.5		88.3		97.5	
overall	98.6		98.6		99.31		98.4	

Table 7 Comparison with the existing studies

	Martis *et al.* [9]	Proposed method
number of classes	15 classes mapped to 5 categories	15 classes mapped to 5 categories
dataset used	MIT-BIH	MIT-BIH
beats segmentation	100 samples before R-peak and 100 samples after	dynamic segmentation method
features extracted	QRS detection using pan Tompkins [35]wavelet transform	discrete wavelet transformRR features
reduction method	independent component analysis	principle component analysis
classifier	probabilistic neural network	support vector machine(one versus one)
fusion method	—	stacked generalisation
overall accuracy	99.28%	99.3% (fusion with PNN)98.4% (fusion with decision trees)
average accuracy	—	88.3% (fusion with PNN)97.5% (fusion with decision trees)

Table 8 Classification of ten trials average results for category N, S and Q for lead 2 and category V for lead 1 using one versus one SVM methodology

Leads	Categories	Classes	Se, %	+ P, %	Average, %	Overall, %
lead 2	category N	NOR	93.5	75.7	90.4	94.3
		LBBB	98.8	98.6		
		RBBB	99.0	91.7		
		AE	67.5	98.98		
		NE	92.9	97.8		
	category S	APC	96.4	99.7	93.9	95.9
		AP	91.6	91.04		
		BAP	92.3	100.0		
		NP	95.1	98.5		
	category Q	FPN	99.7	80.0	86.7	99.3
		UN	76.8	100.0		
lead 1	category F	VFN	98.0	98.0	98.0	98.0
		PVC	99.8	97.5		
	category V	VE	97.9	99.96	98.5	99.7
		VF	97.8	99.9		
					93.19	94.94

the authors as mentioned in Section 2, except for the proposed work by Ye *et al.* [2], none of the existing studies have considered all the classes. Ye *et al.* [2] achieved an overall accuracy ∼99% (4% more than this study). However, by computing the accuracy achieved by each class in [2] (total number of correctly classified beats divided by total number of class beats), it was found that the overall accuracy is biased towards the normal classes (category N with the huge number of testing beats). Since, the average accuracy is ∼88% which means 5% less than this study. Moreover, it is worth mentioning that Ye *et al.* [2] have utilised more features (Section 2). Finally, regarding the time needed for testing, each beat approximately needs only 0.0039 s to be fully classified by both stages. The time was computed on a machine with 8 GB RAM, 2.20 GHz CPU and Core i7.

5 Conclusions and future directions

To sum up, an automatic hybrid hierarchy method of two stages has been developed. The first stage classifies a heartbeat into one of five main categories. Thereafter, the heartbeat moves to the second stage to know which class in this category it belongs. A DWT is applied on dynamically segmented heartbeats. The resulted wavelet coefficient structures have been projected to sub-space of 18 PCA components. After concatenating the four RR features, the final feature vectors are fed into the SVM NN for classification.

Two strategies for classification, one versus one and one versus all have been examined. The proposed method has been applied to ECG data from both lead 1 and lead 2. A new fusion step based on stacked generalisation in the first stage is suggested to get the advantage of considering the information derived from both leads 1 and 2. The new step has been examined and compared with the rejection method utilised by the existing studies. All experiments have been conducted using an MIT-BIH database.

The results reveal the significance of RR interval features for characterising some categories. In addition, simplifying the classification process to a one versus one strategy achieved the best accuracies for both lead 1 and lead 2. Furthermore, not only the

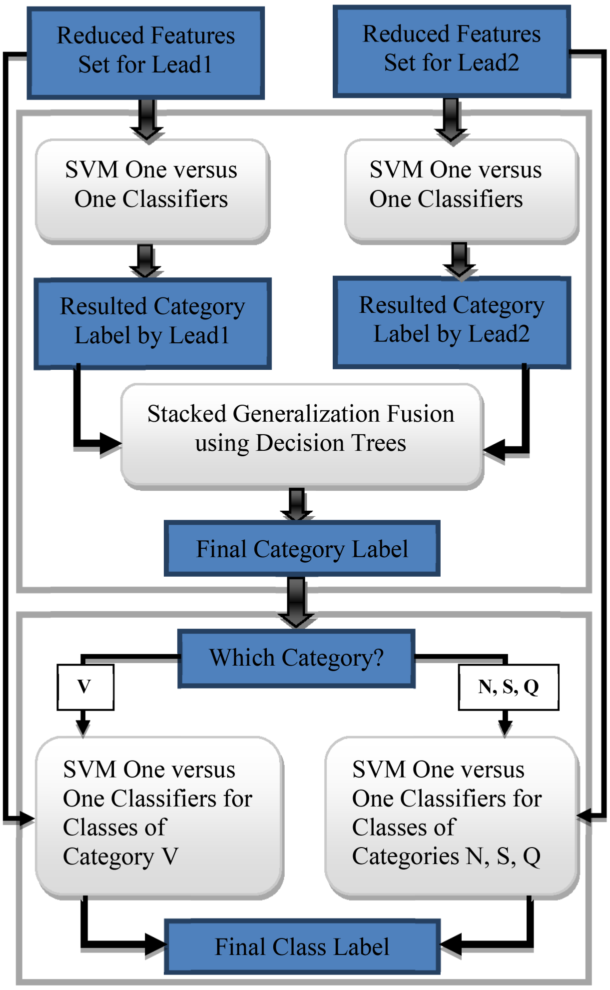

Fig. 7 *Final two-stage hierarchical method*

overall accuracy has been considered by this study (as all studies in the literature) but also the average accuracy. Moreover, the proposed fusion step based on the stacked generalisation method with decision trees, overwhelms the rejection method and provides the best benefit of both leads information regarding the first stage.

In addition, to the best knowledge of the authors, none of the existing studies provides information on both category and class in the hierarchical approach. If only the category is needed the first stage will be enough. On the other hand, if more specific information on the nature of the heart beat is needed, it will be the role of the second stage. Finally, the proposed method has achieved in the first stage the best result of 98.4% overall accuracy and the best average accuracy of 97.5%. Regarding the second stage, 94.94% overall accuracy and 93.19% average accuracy have been achieved using one versus one classification strategy. The average accuracy achieved for each class provides evidence on the robustness of the proposed method to classify all classes (accuracy of most classes is above 90%) and not only some classes at the expense of the others. In the future, we are looking forward to having enough data to gain the advantage of applying stack generalisation in both classification stages.

6 References

[1] Tantawi, M., Revett, K., Salem, A., *et al.*: 'Electrocardiogram (ECG): a new burgeoning utility for biometric recognition', in Hassanien, A., Kim, TH., Kacprzyk, J., Awad, A. (Eds.): '*Bio-inspiring cyber security and cloud services: trends and innovations*' (Intelligent Systems Reference Library, Springer-Verlag, Berlin Heidelberg, 2014), pp. 349–382

[2] Ye, C., Kumar, B.V.K.V., Coimbra, M.T.: 'Heartbeat classification using morphological and dynamic features of ECG signals', *IEEE Trans. Biomed. Eng.*, 2012, **59**, (10), pp. 2930–2941

[3] El-Saadawy, H., Tantawi, M., Shedeed, H.A., *et al.*: 'Electrocardiogram (ECG) classification based on dynamic beats segmentation'. INFOS '16 Proc. the 10th Int. Conf. Informatics and Systems, Giza, Egypt, NY, USA, 2016, pp. 75–80

[4] Huikuri, H.V., Castellanos, A., Myerburg, R.J.: 'Sudden death due to cardiac arrhythmias', *New England J. Med.*, 2001, **345**, (20), pp. 1473–1482

[5] Alfonso, V.X., Tompkins, J.: 'Detecting ventricular fibrillation', *IEEE Trans. Biomed. Eng.*, 2007, **54**, (1), pp. 174–177

[6] Barro, S., Ruiz, R., Cabello, D., *et al.*: 'Algorithmic sequential decision-making in the frequency domain for life threatening ventricular arrhythmias and imitative artefacts: a diagnostic system', *IEEE Trans. Biomed. Eng.*, 1989, **11**, (4), pp. 320–328

[7] Minami, K., Nakajima, H., Toyoshima, T.: 'Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network', *IEEE Trans. Biomed. Eng.*, 1999, **46**, (2), pp. 179–185

[8] Kastor, J.A.: '*Arrhythmias*' (W.B. Saunders, London, 1994, 2nd edn.)

[9] Martis, R.J., Acharya, U.R., Min, L.C.: 'ECG beat classification using PCA, LDA, ICA and discrete wavelet transform', *Biomed. Signal Proc. Control*, 2013, **8**, (5), pp. 437–448

[10] Yazdanian, H., Nomani, A., Yazdchi, M.R.: 'Autonomous detection of heartbeats and categorizing them by using support vector machines'. 20th Iranian Conf. Biomedical Engineering (ICBME), Tehran, Iran, 2013, pp. 223–227

[11] Martis, R.J., Acharya, U.R., Mandana, K.M., *et al.*: 'Application of principal component analysis to ECG signals for automated diagnosis of cardiac health', *Expert Syst. Appl.*, 2012, **39**, (14), pp. 11792–11800

[12] Bulusu, S.C., Faezipour, M., Ng, V., *et al.*: 'Transient ST-segment episode detection for ECG beat classification'. Life Science Systems and Applications Workshop (LiSSA), 2011 IEEE/NIH, Stock United States, Bethesda, MD, 2011, pp. 121–124

[13] Daamouche, A., Hamami, L., Alajlanc, N., *et al.*: 'A wavelet optimization approach for ECG signal classification', *Biomed. Signal Proc. Control*, 2011, **7**, (2012), pp. 342–349

[14] Melgani, F., Bazi, Y.: 'Classification of electrocardiogram signals with support vector machines and particle swarm optimization', *IEEE Trans. Inf. Technol. Biomed.*, 2008, **12**, (5), pp. 667–677

[15] Khazaee, A.: 'Combining SVM and PSO for PVC Detection', *Int. J. Adv. Eng. Sci.*, 2013, **3**, (4), pp. 1–5

[16] Joshi, N.P., Tambe, S.A., Topannavar, P.S.: 'Morphological & dynamic feature based heartbeat classification', *Int. J. Eng. Res. Gen. Sci.*, 2014, **2**, (3), pp. 115–121

[17] Oster, J., Behar, J., Sayadi, O., *et al.*: 'Semisupervised ECG ventricular beat classification with novelty detection based on switching Kalman filters', *IEEE Trans. Biomed. Eng.*, 2015, **62**, (9), pp. 2125–2134

[18] Chen, X., Ji, J., Loparo, K., *et al.*: 'Real-time personalized cardiac arrhythmia detection and diagnosis: A cloud computing architecture'. 2017 IEEE EMBS Int. Conf. Biomedical & Health Informatics (BHI), Orlando, FL, USA, 2017, pp. 201–204

[19] Alickovic, E., Subasi, A.: 'Medical decision support system for diagnosis of heart arrhythmia using DWT and random forests classifier', *J. Med. Syst.*, 2016, **4**, (40), pp. 1–12, https://doi.org/10.1007/s10916-016-0467-8

[20] 'MIT-BIH Arrhythmias Database', http://www.physionet.org/physiobank/database/mitdb/, accessed October 12 2017

[21] Rodriguez, J., Goni, A., Illarramendi, A.: 'Real-time classification of ECGs on a PDA', *IEEE Trans. Inf. Technol. Biomed.*, 2005, **9**, (1), pp. 23–34

[22] De Lannoy, G., Francois, D., Delbeke, J., *et al.*: 'Weighted conditional random fields for supervised interpatient heartbeat classification', *IEEE Trans. Biomed. Eng.*, 2012, **59**, (1), pp. 241–247

[23] Lagerholm, M., Peterson, C., Braccini, G., *et al.*: 'Clustering ECG complexes using hermite functions and self-organizing maps', *IEEE Trans. Biomed. Eng.*, 2000, **47**, (7), pp. 838–848

[24] Jiang, W., Kong, S.G.: 'Block-based neural networks for personalized ECG signal classification', *IEEE Trans. Neural Netw.*, 2007, **18**, (6), pp. 1750–1761

[25] De Chazal, P., Dwyer, M.O., Reilly, R.B.: 'Automatic classification of heartbeats using ECG morphology and heartbeat interval features', *IEEE Trans. Biomed. Eng.*, 2004, **51**, (7), pp. 1196–1206

[26] De Chazal, P., Reilly, R.B.: 'A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features', *IEEE Trans. Biomed. Eng.*, 2006, **53**, (12), pp. 2535–2543

[27] Llamedo, M., Martinez, J.P.: 'Heartbeat classification using feature selection driven by database generalization criteria', *IEEE Trans. Biomed. Eng.*, 2011, **58**, (3), pp. 616–625

[28] Ince, T., Kiranyaz, S., Gabbouj, M.: 'A generic and robust system for automated patient-specific classification of ECG signals', *IEEE Trans. Biomed. Eng.*, 2009, **56**, (5), pp. 1415–1426

[29] 'Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms', ANSI/AAMI EC57:1998 standard, Association for the Advancement of Medical Instrumentation, 1998

[30] 'Dealing with border distortion', https://www.mathworks.com/help/wavelet/ug/dealing-with-border-distortion.html, accessed October 14 2017

[31] Vapnik, V.N.: '*The nature of statistical learning theory*' (Springer-Verlag, New York, 1995)

[32] Wolpert, D.H.: 'Stacked Generalization', *Neural Netw.*, 1992, **5**, (2), pp. 241–259

[33] Duda, R.O., Hart, P.E., Stork, D.G.: '*Pattern classification*' (Wiley, New York, 2000, 2nd edn.)

[34] Rokach, L., Maimon, O.: 'Decision trees', in Maimon, O., Rokach, L. (Eds): '*Data mining and knowledge discovery handbook*' (Springer, Boston, MA, 2005), pp. 165–192

[35] Pan, J., Tompkins, W.J.: 'A real-time QRS detection algorithm', *IEEE Trans. Biomed. Eng.*, 1985, **32**, (3), pp. 230–236