Computers in Biology and Medicine 46 (2014) 79–89

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Heartbeat classification using disease-specific feature selection

Zhancheng Zhang [a,*], Jun Dong [a], Xiaoqing Luo [b], Kup-Sze Choi [c], Xiaojun Wu [b]

[a] Suzhou Institute of Nano-Tech and Nano-Bionics, Chinese Academy of Science, Suzhou 215123, China
[b] School of Internet of Things, Jiangnan University, Wuxi 214122, China
[c] Centre for Smart Health, School of Nursing, The Hong Kong Polytechnic University, Hong Kong

ARTICLE INFO

Article history:
Received 14 June 2013
Accepted 23 November 2013

Keywords:
Feature selection
Disease specific
Heartbeat classification
Support vector machine

ABSTRACT

Automatic heartbeat classification is an important technique to assist doctors to identify ectopic heartbeats in long-term Holter recording. In this paper, we introduce a novel disease-specific feature selection method which consists of a one-versus-one (OvO) features ranking stage and a feature search stage wrapped in the same OvO-rule support vector machine (SVM) binary classifier. The proposed method differs from traditional approaches in that it focuses on the selection of effective feature subsets for distinguishing a class from others by making OvO comparison. The electrocardiograms (ECG) from the MIT-BIH arrhythmia database (MIT-BIH-AR) are used to evaluate the proposed feature selection method. The ECG features adopted include inter-beat and intra-beat intervals, amplitude morphology, area morphology and morphological distance. Following the recommendation of the Advancement of Medical Instrumentation (AAMI), all the heartbeat samples of MIT-BIH-AR are grouped into four classes, namely, normal or bundle branch block (N), supraventricular ectopic (S), ventricular ectopic (V) and fusion of ventricular and normal (F). The division of training and testing data complies with the inter-patient schema. Experimental results show that the average classification accuracy of the proposed feature selection method is 86.66%, outperforming those methods without feature selection. The sensitivities for the classes N, S, V and F are 88.94%, 79.06%, 85.48% and 93.81% respectively, and the corresponding positive predictive values are 98.98%, 35.98%, 92.75% and 13.74% respectively. In terms of geometric means of sensitivity and positive predictivity, the proposed method also demonstrates better performance than other state-of-the-art feature selection methods.

1. Introduction

The electrocardiogram (ECG) is a noninvasive, inexpensive and well-established diagnostic tool. It contains essential physiological signals that are widely used to analyze heart function. However, for the analysis of long-term ECG recording, beat-by-beat manual examination is tedious and time-consuming, especially for bedside monitoring or wearable online health care monitoring, where real-time diagnosis is a difficult task for junior doctors. Therefore, clinicians usually employ computer-assisted methods to analyze and interpret the ECG signals.

At critical levels, cardiac arrhythmias can be categorized into two types, life-threatening and non-life-threatening. Life-threatening arrhythmias may trigger cardiac arrest and sudden death, such as ventricular fibrillation and tachycardia. Patients under these situations require emergency treatment. While non-life-threatening arrhythmias may not lead to imminent heart failure, timely therapy is still needed to avoid further deterioration of heart function. In some cases, arrhythmias may only occur sporadically in patient's daily living. To capture these infrequent events, a Holter device is usually employed to record long-term ECG data. Hence, automatic recognition of abnormal heartbeats from a large amount of ECG data is an important and essential task. The first type of arrhythmias has been well studied and some detection algorithms have been developed and implemented in automatic external defibrillators (AED) [1–3]. To recognize the second, detecting ectopic beat is a critical step. In this study, we only focus on the investigation of non-life-threatening arrhythmias and the related ectopic beat detection.

In the past few decades, a considerable amount of research has been dedicated to automatic heartbeat classification [4–9]. The work can be categorized into two classification paradigms, namely, “intra-patient” and “inter-patient”, or also known as “class-oriented” and “subject-oriented” respectively [8]. Intra-patient paradigm [4,5] partitions the whole data set into training and testing subsets based only on the beat label, where an ECG recording may partly appear in both the data subsets. By this scheme, the classifiers usually produce over optimistic results. In clinical practice, the classification performance declines due to the inter-individual variation.

In order to adapt to practical situations, de Chazal et al. [6] proposed the inter-patient paradigm where the training

* Corresponding author. College of Electronic and Information Engineering, Suzhou University of Science and Technology, Suzhou, 215009, China
E-mail address: cimszhang@163.com (Z. Zhang).

0010-4825/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiomed.2013.11.019

Table 1
Size of data sets DS1 and DS2 and the mapping between AAMI and MIT-BIH-AR labels.

Data set	N			S						V		F	Q	
	N	L	R	e	j	A	a	J	S	V	E	F	f	Q
DS1[a]	38 052 45 777	3946	3779	16 973	16	807	100	32	2	3664 3769	105	414 414	0 8	8
DS2[b]	36 413 44 011	4123	3475	0 2049	213	1735	50	51	0	3215 3216	1	388 388	0 7	7

[a] Recordings in DS1: 101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, 230.
[b] Recordings in DS2: 100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234.

and testing subsets were constructed from different ECG recordings so that inter-individual variation was taken into account and the classifier would exhibit better generalization ability. This paradigm has been adopted [7–9] to evaluate the classification performance on MIT-BIH arrhythmia (MIT-BIH-AR) data set [10,11].

Besides, a hybrid paradigm called "patient-specific" was also proposed by de Chazal et al. [12], in which a global classifier was first trained and a local classifier was then employed to tune the global classifier. It was reported that this approach outperformed those achieved by pure inter-patient classifiers [6]. Inspired by this paradigm, Ince et al. [13] proposed a "patient-specific" neural network (NN) where the training data was consisted of two parts: (i) common representative beats randomly selected from the training recordings and (ii) patient-specific beats segmented from the first 5 min of each recording. Moreover, Wiens et al. [14] employed an active learning method to detect ventricular ectopic beats and supraventricular ectopic beats. Furthermore, Llamedo et al. [15] studied how much the performance would be improved when experts efficiently assisted in the active learning iteration steps. Most of the studies followed the Advancement of Medical Instrumentation (AAMI) recommendation [16] which specifies heartbeats using five labels, namely, normal or bundle branch block beat (N), supraventricular ectopic beat (S), ventricular ectopic beat (V), fusion of ventricular and normal beat (F), and heartbeats that cannot be classified (Q). This recommendation makes possible a fair comparison among various heartbeat classifiers.

Regarding the ECG features commonly employed for classification, features surrounding RR intervals are most widely used, such as pre-RR, post-RR, local-RR, average-RR [6] and other RR-based features [7]. Other time domain features including PP interval, P duration, QRS duration, PR interval, T duration and QT interval are also considered. Moreover, "morphology" features of ECG samples in P wave, QRS complex and T wave as well as the morphological distances [14] between the beats and the median beat have also been used. These features have been clinically studied and related diagnostic standards have also been stipulated [17]. Although vectorcardiogram (VCG) based features [7,18] can provide comprehensive information about heart conditions, reconstruction of VCG requires more leads and thus the applicability of these features is rather limited. Besides, frequency domain feature analysis can also provide deep insight into ECG signals. Signal processing methods include wavelet decomposition (WT) [4,5], principal component analysis (PCA) and independent component analysis (ICA) [8]. Although these features are associated with clear mathematical interpretations, they do not have physiological meaning that allows doctors to comprehend in an intuitive way.

Hence, the selection of the most relevant features from the mass of features to improve the classification performance is a great challenge [19]. Llamedo et al. [7] employed a sequential floating feature selection (SFFS) algorithm to enhance the Bayesian classifier. They tried to find an optimal feature subset for all subclassifiers. However, in clinical practice, different diseases are characterized by different features (symptoms). By analyzing the features, clinicians are able to differentiate one disease from another, or distinguish diseased state from a healthy state. Based on this notion, we attempt to develop a novel "disease-specific" feature selection method and utilize it to enhance the performance of heartbeat classification. To evaluate this method, we use the inter-patient classification paradigm to ensure better generalization ability and focus on the four clinically well-studied classes of N, S, V and F. The AAMI recommendation is followed to present testing results.

2. ECG data

In this study, the MIT-BIH-AR database is used for training and testing. The database consists of 48 two-lead recordings obtained from 47 subjects (recordings 201 and 202 were obtained from the same subject), each recording was measured for approximately 30 min and sampled at 360 Hz. Of the 48 recordings, 23 of them (the "100 series") were collected from routine ambulatory practice and the remaining 25 (the "200 series") were selected to include examples of uncommon but clinically important arrhythmias cases that were not well represented in the 23 100-series recordings. The ECG leads varied among the subjects due to physical limitation of electrode placement [11]. For most of the recordings, the first channel was a modified limb lead II (MLII) (only 114 recordings used V5 as the first lead and MLII as the second lead; the leads were then swapped in the study). The second channel was usually V1 (sometimes V2, V4 or V5, depending on subjects). The database contains annotations for both QRS position and beat class information, verified by at least two independent experts.

To ensure a fair comparison with the results in related literatures [6,7] and to comply with the AAMI recommendation, the four recordings with paced beats, namely, 102, 104, 107 and 217, are discarded in the study. Also, all original heartbeat labels are mapped to the AAMI labels with the mapping rules listed in Table 1. We also follow the training set (DS1) and testing set (DS2) division schemes adopted in [6]. Note that the AAMI Q class (unclassified and paced heartbeats) is discarded since this class is marginally represented in MIT-BIH-AR. The numbers of heartbeats in DS1 and DS2 data sets are listed in Table 1 by class.

3. Methodologies

A four-stage classification strategy is adopted in this study to analyze the ECG data. As shown in Fig. 1, the whole classification strategy consists of four stages: signal preprocessing, feature extraction, feature selection and classifier tuning. Compared with previous work [6,7], in addition to morphology and interval

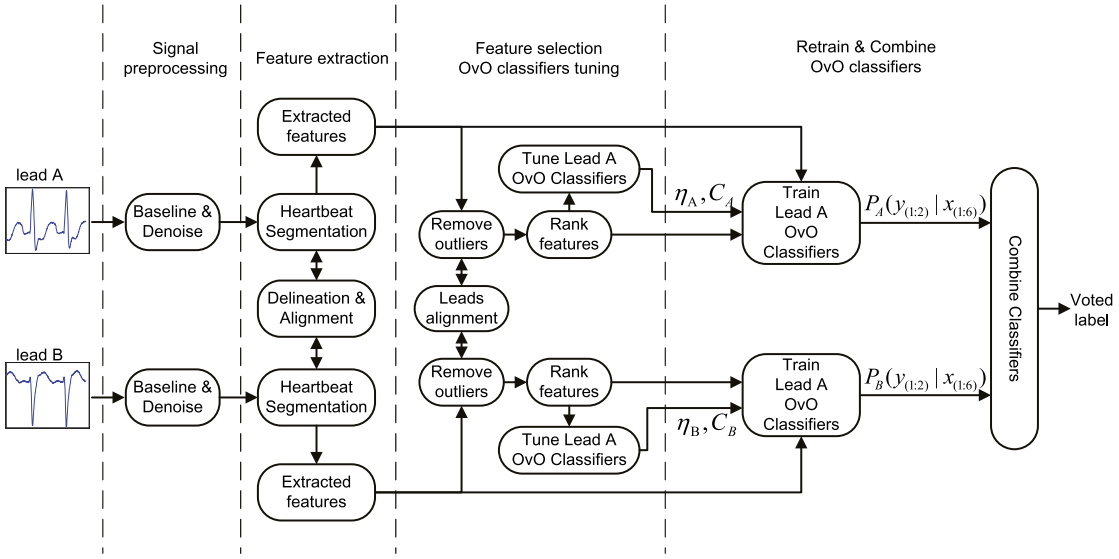

Fig. 1. The fully automatic heartbeat classification procedure.

features, areas under P wave, QRS complex and T wave are also considered as candidate features in the study. The feature selection stage is additionally introduced to determine the appropriate features. Moreover, by combining several OvO SVM binary classifiers trained independently with data acquired from the two leads, a fused classifier with enhanced generalization ability is developed. In contrast to the work of Llamedo et al. [7], where all the classes utilized the same feature subset, the proposed OvO based feature ranking approach is a more reasonable and realistic feature selection method, i.e., the "optimal" feature subset obtained is disease-specific and restricted to a binary classification task only. This advantage will be discussed in detail in Section 3.3.

3.1. ECG preprocessing

Since the ECG signals in the MIT-BIH-AR database were collected by Holter devices, the signals were contaminated by baseline wandering and noises from power-line as well as high-frequency electromyography disturbance. Following the approaches adopted in previous work [12,6], all the ECG signals are first preprocessed using a 200-ms width median filter to remove P wave and QRS complex, then a 600-ms width median filter to remove T wave. The resulted signals are then regarded as the baseline which is subsequently subtracted from the original signals to yield the baseline-corrected ECG signals. A 12-order FIR low-pass filter with a 35 Hz cut-off frequency is then used to remove power-line and high-frequency noise. Finally, the filtered two-lead ECG signals are fed into the next stage for further processing.

3.2. Heartbeat delineation and segmentation

Before extracting the heartbeat features, the ECG time sequence must be segmented into individual heartbeats based on the positions of the QRS waves. While many methods can be used to locate QRS waves and detect the fiducial points, such as the techniques used in our previous work [20,21] and wavelet-based methods [22,23], the objective of the present study is to develop a heartbeat classifier by feature selection. For convenience and to operate well with the WaveForm DataBase (WFDB), we employ the QRS annotation included in MIT-BIH-AR and use the tool "ecgpuwave" (a QRS detector and waveform limit locator available from the *physionet* at http://www.physionet.org/physiotools/ecgpuwave/) to detect the fiducial points, which include P-onset, P-peaks, P-offset, QRS-onset, QRS-offset, T-onset, T-peaks and T-offset. When P-wave is absent from a heartbeat, the onset and offset of the heartbeat are assigned to the next QRS-onset point. As a result, the heartbeat would have zero P-duration, a truncated PR interval and a truncated post-PP interval. Similarly, when T-wave is absent from a heartbeat, the onset and offset are assigned to the previous QRS-offset, resulting in zero T-duration and a truncated QT-interval.

3.3. Feature extraction

Based on the detected fiducial points, a total of 46 features per-beat and per-lead are considered in the study, which are listed in Table 2 and illustrated in Fig. 2.

3.3.1. Inter-beat intervals

Five inter-beat intervals are defined as the interval between successive heartbeat fiducial points. Based on the QRS-peak points, there are four RR related intervals. The *pre-RR interval* is the RR-interval between a given heartbeat and the previous heartbeat. The *post-RR interval* is the RR-interval between a given heartbeat and the following heartbeat. The *local*-10 *average RR-interval* is determined by averaging the 10 RR-intervals surrounding a heartbeat. The *average RR-interval* is the mean of the RR-intervals for a recording and has the same value for all heartbeats in a recording. Finally, based on P-peak points, the *post-RR interval* is defined as the interval between a given heartbeat's R-peak and the following heartbeat's R-peak. The post-RR intervals characterize ventricular periods and the post-PP intervals characterize atrial period. As illustrated in Fig. 2, post-RR interval, post-PP interval and intra-beat intervals are interdependent.

3.3.2. Intra-beat intervals

An intra-beat interval is defined as the interval between a posterior fiducial point and an anterior fiducial point in a heartbeat. The five intra-beat intervals used in this study are depicted in Fig. 2 and their related fiducial points are also listed in the corresponding rows and the final column of Fig. 2. We use these intervals as candidate features because they are well-known and widely used in clinical practice.

Table 2
Features used in this study.

Feature group	Indices	Feature description
Inter-beat intervals	1	Pre-RR interval
	2	Post-RR interval
	3	Local-10 Average RR-interval
	4	Average RR-interval
	5	Post-PP interval
Intra-beat intervals	6	P duration, P offset − P onset
	7	QRS duration, QRS offset − QRS onset
	8	T duration, T offset − T onset
	9	PR interval, QRS onset − P onset
	10	QT interval, T offset − QRS onset
Morphological amplitudes	11–20	P morphology, 10 samples between P onset and P offset
	21–30	QRS morphology, 10 samples between QRS onset and QRS offset
	31–39	ST morphology, 9 samples between QRS offset and T offset
Morphological areas	40,41	Positive and negative areas of P wave
	42,43	Positive and negative areas of QRS complex
	44,45	Positive and negative areas of T wave
Morphological distance	46	DTW distance between a beat and the median beat of a recording

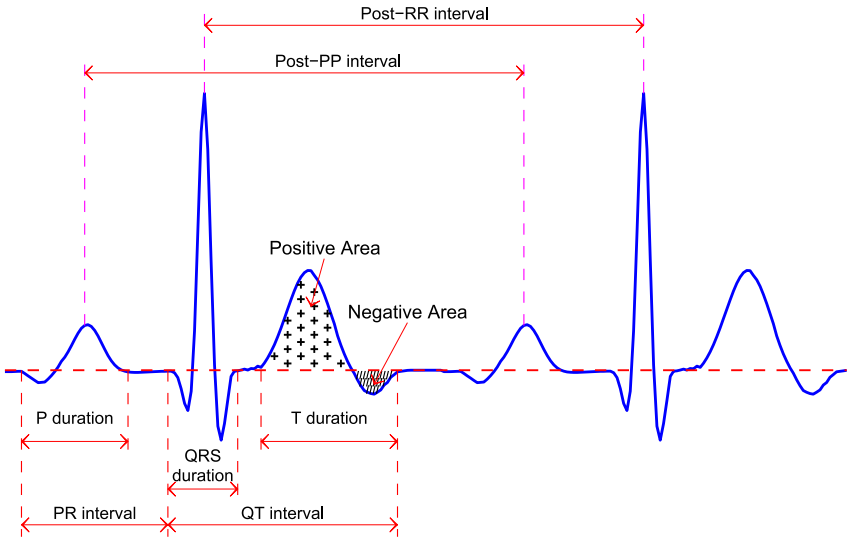

Fig. 2. Illustration of area-based and interval-based feature in ECG: regions filled with + and / sign to represent positive and negative areas.

3.3.3. *Morphological amplitudes*

The morphological amplitude features presented in [6] are also adopted in this study. To depict ECG morphology, a group of values are derived by down-sampling the signal amplitude in a specific window. The *P morphology* is defined as physical amplitudes of 10 samples between P onset and P offset. Similarly, the *QRS morphology* and *ST morphology* are defined as physical amplitudes of 10 samples between QRS onset and QRS offset and physical amplitudes of 9 samples between QRS offset and T offset respectively.

3.3.4. *Morphological areas*

Given the importance of morphological and interval features in automatic heartbeat classification [6,12], we introduce wave-area-based features in this study to take heartbeat morphology into account for evaluating the effect of feature selection and improving the classification performance. Specifically, as illustrated in Fig. 2, it is clear that some regions are enclosed by the wave intervals, characteristic ECG waves and the baseline. The sum of the area of the regions above baseline is thus defined as a positive area (AreaPos), whereas the area below the baseline is defined as a negative area (AreaNeg). With these enclosed areas as measures, 6 area-based features are obtained, i.e., the AreaPos and AreaNeg of P wave, QRS complex and T wave respectively, which collectively represent the combined characteristics of wave interval and sample amplitude.

3.3.5. *Morphological distance*

Despite individual variations, normal ECG usually possesses a similar morphology between heartbeats of a lead in a recording. To characterize this similarity between heartbeats, Wiens et al. [14] present a novel feature based on the dynamic time warping (DTW) distance between a given beat and the median beat of a recording. This feature has excellent discrimination power for ventricular ectopic beats and supraventricular ectopic beats. In this study, ECG signal started at 250 ms left of QRS-peak and finished at 380 ms right of QRS-peak within each beat are segmented and down-sampled with 50 Hz. The DTW distance from each segment to the median segment is then considered as the morphological distance.

Although ecgpuwave is a robust tool for detecting QRS duration and locating waveform limit, it may still fail to detect fiducial

Table 3
Lower and upper bounds of features for identifying a heartbeat.

Feature name	Lower bound	Upper bound
P duration (ms)	0	0.2
QRS duration (ms)	0	0.5
T duration (ms)	0	0.8
PR interval (ms)	0	0.4
QT interval (ms)	0	1.2
Inter-beat intervals (ms)	0.02	4
Morphology amplitude (mv)	−3.5	4

points in extremely noisy segments and lead to faulty measures and abnormal features. It is therefore necessary to remove these outliers to compare the feature scores. In this study, we employe a rule-based procedure to identify and filter the outliers. Here, the interval features are restricted to a reasonable range. Any beats out of the range are considered as outliers.

With reference to the physiological function of heart [17], the lower and upper bounds of the related features listed in Table 3 are used to qualify a beat. Note that when an outlier beat of a lead is removed, in order to align with the beats of different leads, the corresponding beats in the other leads are removed manually. Note that the outlier removal procedure is only applied to DS1 in the preprocessing stage. The outlier-removed DS1 is used only for feature selection and classifier tuning. In the final training and testing procedures to be presented in Section 4, both DS1 and DS2 are the original beats and no outlier is removed.

Furthermore, normalization is required in order to rank the features. In this study, the tangent sigmoid function (*tansig* in Matlab) is used to normalize all the features into the range $[-1, 1]$.

3.4. SVM with feature selection

Because of excellent generalization ability, L_2 regularized SVM is employed in this study as the heartbeat classifier. This classifier attempts to identify an optimal hyperplane that maximizes the separation margin between two different classes. In this task, suppose a training set consists of N samples $\{(y_i, \mathbf{x}_i),\ i = 1, \ldots, N\}$, where $\mathbf{x}_i \in \mathbb{R}^d$ denotes the d-dimensional feature vector of the ith example and $y_i \in \mathbb{R}$ denotes the corresponding class label and $y_i \in \{\pm 1\}$. The optimal hyperplane is represented by a decision function $f(\mathbf{x})$ learned from the training set to predict the class label in the subsequent tests. By using kernel trick, the decision function can be formulated as

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i \in SVs} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right), \tag{1}$$

where $K(\cdot,\cdot)$ is the kernel function which maps the data into a higher dimensional space, and α_i is the Lagrange multiplier for each training data example. The linear kernel $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$ is adopted for its ability in linear interpretation when being wrapped in a feature selection procedure. In (1), a few of α_i's are usually nonzero and the corresponding training data examples are termed as support vectors (SVs).

Compared with L_2 regularized SVM, another regularized SVM known as L_1 regularized SVM whose objective optimization tends to learn more α equal to zero. This classifier can be consider as a natural feature selection method as well as an embedded feature selection method. In this paper, it serves as a baseline for comparison in Section 4.

A number of software packages can be used to learn the decision function, the *liblinear* software package [24] is selected as the binary classifier solver in the study due to its high efficiency and MATLAB-friendly interface.

As SVM is intrinsically a binary classifier, to extend it for multi-class classification task, specifically for a 4-class heartbeat classification, the one-versus-rest (OvR) and the OvO-based methods are commonly utilized. As discussed previously, in heartbeat classification task, comparison of ECG features in one disease with that in another is a more realistic practice and the decomposition is of fine granularity. Thus, with the OvO schema, 6 (i.e. $\binom{4}{2}$) binary classifiers are built to vote for a predicted label. Moreover, we will show that the OvO schema is also suitable for multi-class feature selection.

It is well known that the incorporation of feature selection can improve the generalization ability of classifiers. To provide a faster and more cost-effective predictor and to gain a better understanding of the processed data [19], a number of methods can be used to select the optimal feature subsets, such as wrapper-based methods, filter-based methods and embedded methods. In wrapper-based methods, classifier is utilized as a black box to score subsets of features according to their predictive power, whereas the selection of subsets of features in filter-based methods is only a preprocessing step which is independent of the chosen predictor. In embedded methods, variable selection is performed in the process of training and is usually classifier specific. For easy integration with SVM, the wrapper method is adopted in this study to combine with SVM.

However, two problems are associated with the use of wrapper methods, namely, the long search time required and the assessment of SVM predictor. To tackle the first problem, we accelerate the search process by employing the *F*-score criterion [25,26] to rank the features and using the backward elimination search strategy. For the latter, all the features are first ranked in descending order of the importance of the features. The least promising features are then progressively eliminated. Specifically, in a binary classification, given the training vectors $\mathbf{x}_i,\ i = 1, \ldots, N$, and the number of positive and negative instances are n^+ and n^- respectively, the *F*-score of the kth feature is defined as

$$F(k) \equiv \frac{(\overline{\mathbf{x}}_k^{(+)} - \overline{\mathbf{x}}_k)^2 + (\overline{\mathbf{x}}_k^{(-)} - \overline{\mathbf{x}}_k)^2}{\frac{1}{n^+ - 1}\sum_{i=1}^{n^+}(x_{i,k}^{(+)} - \overline{\mathbf{x}}_k^{(+)})^2 + \frac{1}{n^- - 1}\sum_{j=1}^{n^-}(x_{j,k}^{(-)} - \overline{\mathbf{x}}_k^{(-)})^2} \tag{2}$$

where $\overline{\mathbf{x}}_k$, $\overline{\mathbf{x}}_k^{(+)}$, $\overline{\mathbf{x}}_k^{(-)}$ are respectively the average of the kth feature of the whole, positive and negative data samples, $x_{i,k}^{(+)}$ and $x_{j,k}^{(-)}$ are the kth feature of the ith positive instance and jth negative instance. The numerator and denominator in the equation represent the inter-class discrimination and intra-class compactness respectively. Therefore, the larger the *F*-score of a feature, the more likely this feature is a discriminative and important feature.

We also note from Table 1 that the number of training instances is imbalanced for the four classes. To balance the sensitivities in a binary classifier, the geometric mean (*g*-means) of the two predicted sensitivities are used to assess the predictor of SVM, which is given by

$$g = \sqrt{\text{Se}^+ \cdot \text{Se}^-},$$

where Se^+ and Se^- are the predicted sensitivity of the positive and negative classes respectively. This metric has been widely used to deal with imbalanced data sets [27] to take into consideration the classification results on both positive and negative classes.

In each binary classifier, it is necessary to adjust the penalty parameter C. An additional feature selection stage is introduced in the proposed method to tune the top η features in the backward feature elimination procedure. The optimal C and η are tuned by a 22-fold (leave-one) cross-validation strategy on DS1, where C is searched over the grid $\{2^0, 2^2, \ldots, 2^{20}\}$. With these parameters, the OvO training procedures are re-executed on the training set to obtain the corresponding SVM bin-models. These procedures are individually executed on each lead. Thus, there are 6 binary

classifiers on each lead and a total of 12 binary classifiers for the two-lead MIT-BIH-AR. In the following subsection, we will show how to fuse all these binary classifiers and vote for the final prediction class.

3.5. Combining classifiers

ECG data usually involve signals from multiple leads, which can be regarded as making observations of the same cardiac activity from different positions (perspectives). Integrating the signals from different leads can enhance the classification accuracy. In this study, we employ the product rule to fuse the same pair of binary classifiers on the two-lead of MIT-BIH-AR. It is based on the Bayesian theory and robust performance has been reported [28]. As discussed previously, since two independent training procedures are executed for the data acquired from each lead, we simply assume the equal prior probability.

Given M classes and L classifiers, the final predicted class y_m can be decided by the average rule and is given by

$$\arg\max_m \frac{1}{L}\sum_{l=1}^{L} P(y_m|\varkappa_l),$$

where $P(y_m|\varkappa_l)$ is the posteriori probability predicted on the class y_m by the classifier $\varkappa_l$. The fusion procedure is executed over two individual binary classifiers trained on the two leads, i.e., $M=2$ and $L=2$.

In a binary classifier, $y_1=+1$ and $y_2=-1$ denotes positive and negative classes respectively. Given a future observation $\mathbf{x}$, the pairwise posteriori probability $P(y_m,x)$ of $\mathbf{x}$ being positive or negative class can be estimated by the sigmoid function [29]:

$$P(y_m|\varkappa_l)=\frac{1}{1+\exp(-y_m f(\mathbf{x}))},$$

where $f(\mathbf{x})$ is the decision value given by (1) which is the distance from $\mathbf{x}$ to the hyperplane.

After a series of fused decisions are made, the bin-label $\{+1,-1\}$ can be obtained to make a prediction on the multi-class problem. Then the majority vote rule below is applied

$$\arg\max_c \sum_r^6 \delta_{cr},$$

where $\sum_r^6 \delta_{cr}$ simply counts the number of votes received for class c from each binary classifier, $c=1,\ldots,4$ (i.e., class N, S, V and F), and $r=1,\ldots,6$.

4. Results

In this paper, we propose a disease-specific feature selection method to improve the heartbeat classification performance on ECG data by identifying an optimal feature subset. An experiment involving three steps is conducted to evaluate this method. First, all the features are ranked in descending order of the OvO F-scores as calculated by (2). Second, these ranked features are fed into the validation model and the parameter search procedures are conducted on DS1. To demonstrate the contribution of feature selection, the L_2 regularized SVM, L_1 regularized SVM without feature selection and with the same OvO rule are considered as baseline classifiers. The SVM penalty parameters in each OvO rule of all methods are independently searched over the grid $\{2^0,2^2,\ldots,2^{20}\}$ and tuned with the g-mean criterion and the 22-fold (leave-one) cross-validation strategy on DS1. The best parameters for the proposed method are listed in Table 4

With these parameters, we set up two evaluation experiments. In the first experiment, a series of optimal OvO binary classifiers are re-trained on DS1, then the prediction and combination are executed on DS2. Using the same training and testing set division scheme, the reference classifiers and our proposed method are compared and the detailed confusion matrix and the recording-by-recording performance comparison are reported in Tables 5 and Table 6. In the second experiment, instead of the fixed division scheme, we run 10 times of random division to evaluate the generalization capability of the proposed method and the results are given in Table 7.

Table 4
Parameters searched with g-mean criterion.

Lead	OvO rule[a]	Se_{pos} (%)	Se_{neg} (%)	g-mean (%)	η	C
Lead A	N vs. S	93.42	87.98	90.66	8	2^2
	N vs. V	94.49	98.73	96.59	8	2^6
	N vs. F	91.13	93.00	92.06	5	2^0
	S vs. V	89.52	95.28	92.35	2	2^6
	S vs. F	97.95	99.03	98.49	35	2^{10}
	V vs. F	94.83	96.38	95.60	34	2^{16}
Lead B	N vs. S	94.64	89.52	92.04	31	2^{14}
	N vs. V	97.12	97.24	97.18	37	2^8
	N vs. F	92.27	87.20	89.70	5	2^0
	S vs. V	95.68	93.42	94.54	35	2^{14}
	S vs. F	95.79	88.65	92.15	10	2^2
	V vs. F	92.07	93.24	92.65	28	2^0

[a] In a binary classifier with "Class1 vs. Class2" rule, we treat Class1 as positive class and Class2 as negative class.

4.1. Feature evaluation

To facilitate visual comparison, we scale the minimum feature score to 0 and the maximum to 1, so that all feature scores are normalized into the range of $[0,1]$. The normalization does not affect the ranks of the features. Fig. 3 shows that the top 20 F-scores at lead A and their corresponding feature indices over each OvO rule. We also stack all the F-scores obtained by different rules along each feature. Fig. 4 plots the top 20 total F-scores. It can be clearly seen that different feature ranks are obtained with different rules, which exactly explains that the features are specific to the diseases of concern and can shift to other features depending on the comparison rules. Moreover, note that the rank of the total F-score in Fig. 4 is also different from that in Fig. 3. This means that for multi-class classification tasks, F-score can only provide a coarse assessment on the importance of a feature. It is more appropriate to decompose the monolithic comparison into granular rules.

It is further observed from Fig. 4 that the rank at lead A is different from that at lead B and all F-scores at lead A are higher than that at lead B except the index of 46. The 46th feature is based on the morphological distance between a beat and the median beat of a recording. This feature has excellent discrimination power for ventricular ectopic beats and supraventricular ectopic beats [14]. In MIT-BIH-AR, lead B is a chest lead placed close to ventricle, so this feature contributes more to recognizing ventricular ectopic beats, and its scores in Fig. 3(h) (N vs. V) and Fig. 3(j) (S vs. F) are higher than other features. The other lower F-scores at lead B may be due to signal variation in this channel, usually V1, and could be V2, V4 or V5 sometimes. Hereafter, we mainly focus on the analysis of feature rank at lead A since the F-scores at lead B are relatively low.

Among all the features listed in Table 2, the RR interval related features (with indices 1–4) are important features which are often ranked top 20, or even top 10 with the exception of the N vs. F rule at lead A. This is in line with the feature selection results in [7], where 4 RR interval related features are selected from 39 features.

Table 5
Performance on DS2 of MIT-BIH-AR with reference classifiers and the proposed method.

Llamedo *et al.* [7]

		Predicted					
		N	S	V	F	Total	$Se(\%)$
True	N	34270	1807	80	8031	44188	77.55
	S	124	1403	280	28	1835	76.46
	V	46	182	2669	321	3218	82.94
	F	11	2	5	370	388	95.36
Total		34451	3394	3034	8750	49629	83.08
$P^+(\%)$		99.47	41.34	87.97	4.23	58.25	78.00

L_1 regularized SVM

		Predicted					
		N	S	V	F	Total	$Se(\%)$
True	N	39043	3111	226	1631	44011	88.71
	S	871	1121	37	20	2049	54.71
	V	165	181	2730	140	3216	84.89
	F	107	1	11	269	388	69.33
Total		40186	4414	3004	2060	49664	74.41
$P^+(\%)$		97.16	25.40	90.88	13.06	56.62	86.91

L_2 regularized SVM

		Predicted					
		N	S	V	F	Total	$Se(\%)$
True	N	39009	3133	219	1650	44011	88.63
	S	848	1144	38	19	2049	55.83
	V	287	193	2600	136	3216	80.85
	F	90	1	12	285	388	73.45
Total		40234	4471	2869	2090	49664	74.69
$P^+(\%)$		96.96	25.59	90.62	13.64	56.70	86.66

Our method

		Predicted					
		N	S	V	F	Total	$Se(\%)$
True	N	39142	2631	152	2086	44011	88.94
	S	359	1620	56	14	2049	79.06
	V	31	249	2749	187	3216	85.48
	F	15	2	7	364	388	93.81
Total		39547	4502	2964	2651	49664	86.82
$P^+(\%)$		98.98	35.98	92.75	13.73	60.36	88.34

	N		S		V		F		Average			g-mean	
Classifier[1]	Se	P^+	Se	P^+	Se	P^+	Se	P^+	Se	P^+	Acc	Se	P^+
Llamedo *et al.*[7]	77.55	**99.47**	76.46	**41.34**	82.94	87.97	**95.36**	4.23	83.08	58.25	78.00	82.75	35.17
L_1 regularized SVM	88.63	96.96	55.83	25.59	80.85	90.62	73.45	13.64	74.69	56.70	86.66	73.63	41.84
L_2 regularized SVM	88.71	97.16	54.71	25.40	84.89	90.88	69.33	13.06	74.41	56.62	86.91	73.11	41.37
Our method	**88.94**	98.98	**79.06**	35.98	**85.48**	**92.75**	93.81	**13.73**	**86.82**	**60.36**	**88.34**	**86.66**	**46.15**

[1] In each metric column, the bold metric denotes the best one among the four methods.

Besides, as shown in Fig. 3(a), the P wave related features (positive area of P wave, P duration, PR interval, PP interval and P morphology, with the respective indices of 40, 6, 9, 5 and 15–18) and ST related features (QT interval, T duration, positive and negative areas of T wave, with indices of 10, 8, 44 and 45 respectively) are important for distinguishing supraventricular ectopic beats from normal or branch block beats (N vs. S rule).

It is worth noting that the positive areas of P wave (40) in Fig. 3 (a), (b) and (f) are associated with the P duration (6) and PR interval (9), and that the ranks of these features are close. This phenomenon is attributed to the physiological fact that a wide P wave would result in short PR interval or vice versa. Similarly, in Fig. 3(c)–(e), the QRS morphology of the ninth sample (29) is associated with the negative area of QRS complex (43). This is because the negative area of the QRS complex is mainly attributed to the S wave hollow (tail part of QRS morphology, under the baseline), as illustrated in Fig. 2.

Redundancies existing in these concomitant features are not desirable from the view point of dimension reduction. The redundant features can be removed by making use of the mutual information between the features. However, in this paper, our goal is to find the optimal features to improve heartbeat classification rather than achieving the lowest possible dimension. In fact, redundant features may compensate the errors introduced in the delineation stage and improve the robustness of classification.

4.2. Evaluation on DS1/DS2 division

In this subsection, the recording division used in [6] is adopted to compare the proposed method with other reference methods. Both L_1 regularized SVM and L_2 regularized SVM are executed on full features. Their penalty parameter C are tuned on DS1 by the leave-one cross validation strategy and with the same grid search schema as described in Section 3.4. The results reported previously by Llamedo et al. [7] are also included in Table 5 to serve as baselines for comparison.

Following the AAMI recommendation, the multi-class classification result is evaluated with a confusion matrix to illustrate the performance of the classifiers and the detailed distribution of the misclassified samples. Based on the confusion matrix, the classification performance is measured in terms of class sensitivity Se_c and class positive predictive value (e.g. P_c^+ for class c), which has been formulated in detail in [7] and used to measure inter-patient classification performance across various databases and classifiers. Besides, the recording-by-recording classification performance on DS2 is also given in Table 6 for reference.

The results in Table 5 show that our method outperformed in 10 out of the 13 metrics while the method of Llamedo et al. [7] produced better results in the other three metrics. Compared with L_1 and L_2 SVM methods without feature selection, our method significantly improved the classification performance and outperformed in all the metrics. These findings indicate that the proposed feature selection method can make positive contributions to the classification performance.

On the other hand, the results show that the proposed method loses positive predictivity of S and sensitivity of F, as evident from the confusion matrices obtained and the comparison with the method of Llamedo et al. [7] which also employed feature selection. First, the misclassified number of "S→N" is 359, which is greater than that of the finding in the work of Llamedo et al.

Table 6
Performance of the proposed method on the DS2 data set of MIT-BIH-AR.

Rec	Number of beats				N		S		V		F		Average		
	N	S	V	F	Se	P^+	Se	P^+	Se	P^+	Se	P^+	Se	P^+	Acc
100	2237	33	1	0	100	99.42	60.61	100	100	100	–	–	86.87	99.81	99.43
103	2081	2	0	0	99.09	99.95	50.00	5.26	–	–	–	–	74.54	52.61	99.04
105	2525	0	41	0	91.56	100	–	–	92.68	29.92	–	–	92.12	64.96	91.58
111	2122	0	1	0	99.81	100	–	–	100	50.00	–	–	99.91	75.00	99.81
113	1788	6	0	0	98.83	100	0	0	–	–	–	–	49.41	50.00	98.49
117	1533	1	0	0	100	100	100	100	–	–	–	–	100	100	100
121	1860	1	1	0	98.55	100	100	3.57	100	100	–	–	99.52	67.86	98.55
123	1513	0	3	0	99.80	100	–	–	100	100	–	–	99.90	100	99.80
200	1740	30	826	2	91.03	98.63	70.00	5.88	73.61	98.70	0	0	58.66	50.80	85.18
202	2060	55	19	1	66.84	99.93	69.09	5.47	100	45.24	100	5.00	83.98	38.91	67.21
210	2420	22	193	10	83.35	98.82	72.73	4.15	79.79	92.22	0	0	58.97	48.80	82.68
212	2747	0	0	0	66.36	100	–	–	–	–	–	–	66.36	100	66.36
213	2639	28	220	362	92.76	99.71	7.14	66.67	28.64	73.26	97.51	50.07	56.51	72.43	88.21
214	2001	0	256	1	94.30	100	–	–	99.22	96.21	0	0	64.51	65.40	94.82
219	2080	7	64	1	84.57	99.72	28.57	0.63	82.81	98.15	0	0	48.99	49.62	84.29
221	2030	0	396	0	79.21	100	–	–	99.75	100	–	–	89.48	100	82.56
222	2059	421	0	0	70.03	87.34	48.46	25.47	–	–	–	–	59.25	56.40	66.37
228	1687	3	362	0	97.04	99.94	100	5.00	97.24	99.72	–	–	98.09	68.22	97.08
231	1566	1	2	0	99.87	99.94	0	0	100	100	–	–	66.62	66.65	99.81
232	397	1382	0	0	99.75	80.00	92.76	99.92	–	–	–	–	96.26	89.96	94.32
233	2228	7	828	11	68.81	99.93	0	0	96.86	98.77	90.91	1.39	64.14	50.02	76.28
234	2698	50	3	0	99.89	99.23	58.00	96.67	100	100	–	–	85.96	98.63	99.13

Table 7
Comparison of classification performance among L_1 regularized SVM, L_2 regularized SVM and our method.

Classifier	*g-Mean* (%)	*p*-Value	Accuracy (%)	*p*-Value
L_1 Regularized SVM	74.99 ± 4.96	*0.0009*	85.92 ± 3.09	*0.0482*
L_2 Regularized SVM	75.66 ± 5.04	*0.0015*	85.84 ± 3.24	*0.0452*
Our method	$\mathbf{85.18 \pm 5.40}$	–	$\mathbf{87.17 \pm 3.41}$	–

(i.e., 124). This could be due to the poor performance of the N vs. S binary classifiers for lead A. As shown in Table 4, the sensitivity of class S is 87.98% lower than other sensitivities. The imbalanced training data could also be a reason for the loss of sensitivity. Moreover, since class S is indeed a combination of 6 classes of MIT-BIH labels, further decomposition of class S into several subclasses might be needed to train the related classifiers. As for class F, there are only 414 heartbeats distributed over 10 recordings of DS1. The amount of data is not sufficient to train an inter-patient classifier.

In terms of positive predictive values of class N, the results show that our method is only 0.49% lower than that reported in Ref. [7] and this is caused by the "S→N" misclassified beats. In clinical practice, the positive predictive value of class N is a more important metric than the sensitivity of class N, since misdiagnosing a patient as a healthy person may delay therapy and aggravate the illness. Therefore, doctors usually attempt to achieve a trade off between these two metrics. We usually expect to achieve higher sensitivities on high-risk ectopic heartbeats (class S and V), however this may lead to a low positive predictive on class N. Thus we must maintain a reasonable positive predictivity of class N when pursuing high sensitivities of class S and V. From this point of view, the proposed method is more appropriate for computer aided ectopic heartbeat discrimination.

Table 6 shows the detailed performance by recording of DS2, following the recommendations of the AAMI for results presentation. This table provides insight into every recording loses on every class. As shown in Table 6, recordings 232 and 222 concentrate the majority of the examples for the supraventricular class. Although the results on 232 are high, the results on 222 are poor and these failures drop the overall performance of the S class. The difference between recording 232 and 222 also indicates that the inter-individual variations mainly impact the heartbeat classification performance, thus the patient-specific classification methods should be considered to improve the performance.

4.3. Evaluation on random division

For evaluating the generalization capability, we randomly split all 44 recordings into two groups, one for training and the other for testing. We repeat this division 10 times and summarize the averages and standard deviations of *g*-mean of sensitivity and accuracy in Table 7. Paired *t*-test with 0.05 significance level is also performed to evaluate whether the difference in *g*-mean and accuracy achieved by L_1 regularized SVM, L_2 regularized SVM and the proposed method is statistically significant. Here, a *p*-value less than 0.05 indicates that the difference is significant, and the smaller the *p*-value, the more significant the difference.

It is observed from the results summarized in Table 7 that comparing L_1 regularized SVM and L_2 regularized SVM with the proposed method respectively, both with *g*-mean of sensitivity and with classification accuracy, the proposed method significantly outperforms the other two methods. It is reported [30] that L_1 regularized SVM works well on data with many irrelevant features and can learn a more sparse model when compared with L_2 regularized SVM. While this makes L_1 regularized SVM a natural candidate for feature selection, it only takes effect in the presence of high dimensional inputs. In our experiments, there are only 46 features which are not large enough to enjoy the advantage of L_1 regularized SVM, and the proposed method is therefore more appropriate for the situation presented.

5. Conclusion

We have proposed a novel disease-specific feature selection method for heartbeat classification with ECG data by introducing the OvO combination method using a series of SVM binary classifiers. This method is suitable for heartbeat classification

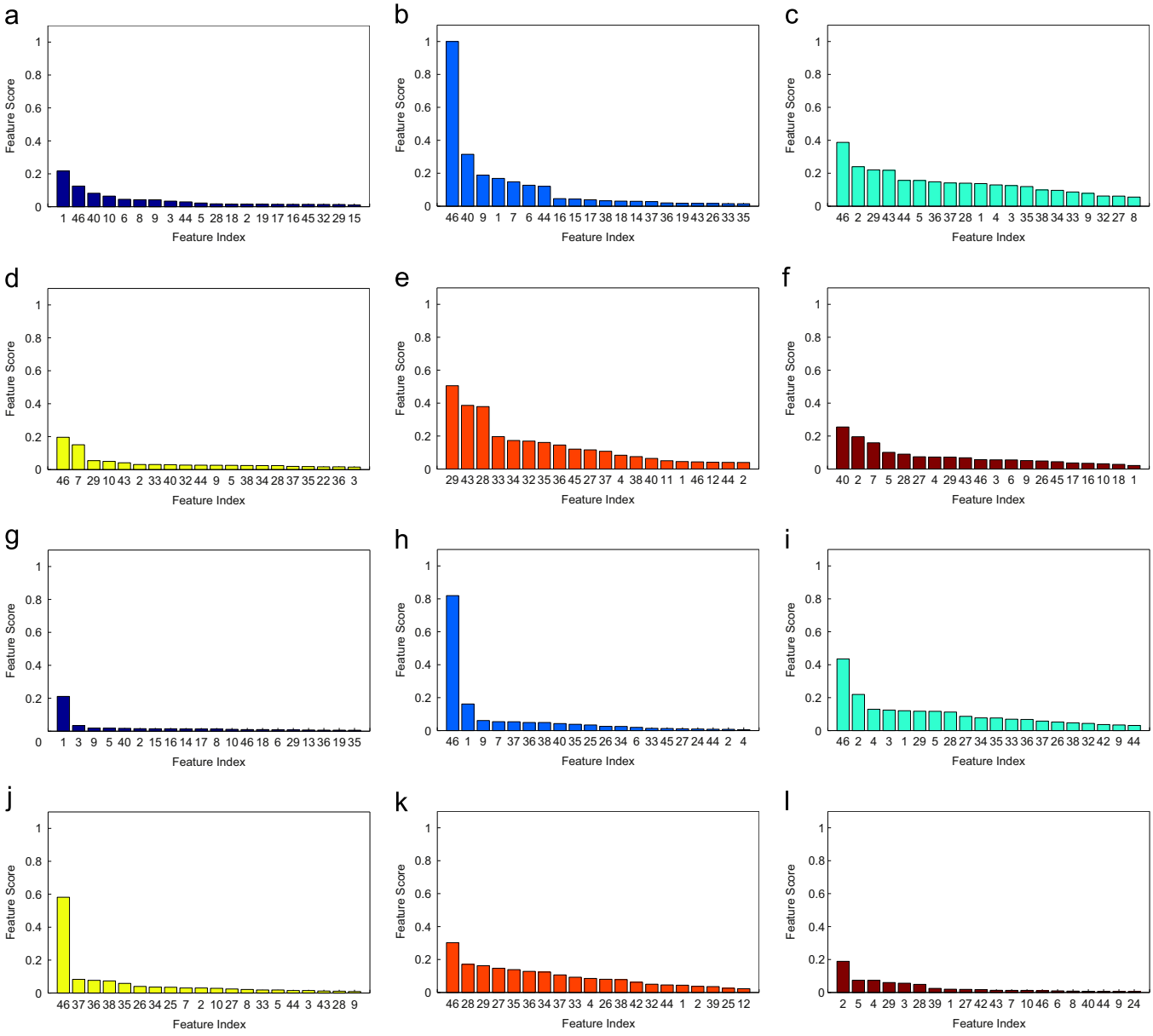

Fig. 3. The Top 20 *F*-scores at lead A ((a)–(f)) and lead B ((g)–(l)) in all OvO rules, feature indices refer to indices listed in Table 2. (a, g) N vs. S; (b, h) N vs. V; (c, i) N vs. F; (d, j) S vs. V; (e, k) S vs. F; (f, l) V vs. F.

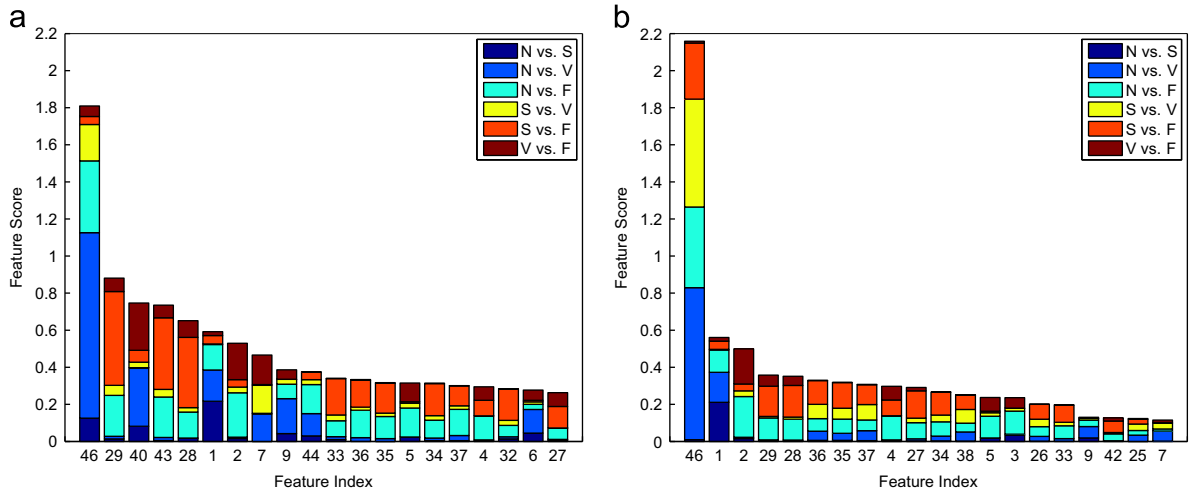

Fig. 4. The top 20 total *F*-scores for all OvO rules: (a) lead A, (b) lead B. The feature indices are listed in Table 2.

because of its pairwise feature ranking characteristic. For integrity, the training procedure is presented by wrapping it with the proposed feature selection method. The testing procedure is conducted with searched feature subsets. We evaluate the method using the MIT-BIH-AR database with the inter-patient schema, based on the inter-beat intervals, intra-beat intervals, morphological amplitudes and morphological areas features.

The classification diagram fulfils the inter-patient schema and complies with the AAMI recommendation. Using the selected features, the sensitivities of the classification system are found to be 88.94%, 79.06%, 85.48% and 93.81% respectively for the class N, S, V and F, and the corresponding values of positive predictivities are 98.98%, 35.98%, 92.75% and 13.73%. In summary, the average sensitivity, positive predictivity and accuracy are 86.82%, 60.36% and 88.34% respectively for the four classes, the geometric means of sensitivities and positive predictivities are 86.66% and 46.15%. Thus, in terms of average or *g*-mean metrics, the proposed method has demonstrated significant improvement in the classification performance over the methods previously developed for fully automatic heartbeat classification on MIT-BIH-AR. Two conclusions can be drawn from the study. First, feature ranking is lead and disease specific. Second, utilizing this property, feature selection can improve the performance of SVM and is a necessary step for heartbeat classification.

Regarding future work, one important research direction is to develop patient specific classifier by making use of heartbeats in the first 5 min as auxiliary information to improve the classification performance of the long-term recording that follows. Recently, the manual expert-assisted approach [12,13] and unsupervised machine learning methods [15] have been employed for research in this direction. Further investigations will be carried out by studying other full-automatic clustering or semi-supervised methods. Another interesting research direction is the study of the sequence of OvO classifiers in the framework of multi-task classification [31] is another interesting direction. Heartbeat classification is indeed a typical imbalanced problem of sensitivities. The adaptation of traditional machine learning algorithms for this task remains a question. Moreover, it should be borne in mind that the existing interval features and morphology features are insufficient to distinguish one type of heartbeat from another. Extractions of more effective disease specific features for accurate heartbeat classification deserve further investigations.

Conflict of interest statement

None declared.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61103128), the China Postdoctoral Science Foundation (2013M541601), the Jiangsu Planned Projects for Postdoctoral Research Funds (1201043C, 1301079C) and the General Research Fund of the Hong Kong Research Grants Council (PolyU 5134/12E).

References

[1] A. Amann, R. Tratnig, K. Unterkofler, Detecting ventricular fibrillation by time-delay methods, IEEE Trans. Biomed. Eng. 54 (1) (2007) 174–177.
[2] U. Irusta, J. Ruiz, S. de Gauna, T. Eftestol, J. Kramer-Johansen, A least mean-square filter for the estimation of the cardiopulmonary resuscitation artifact based on the frequency of the compressions, IEEE Trans. Biomed. Eng. 56 (4) (2009) 1052–1062.
[3] E. Aramendi, U. Irusta, E. Pastor, A. Bodegas, F. Benito, Ecg spectral and morphological parameters reviewed and updated to detect adult and paediatric life-threatening arrhythmia, Physiol. Meas. 31 (6) (2010) 749.
[4] R.J. Martis, C. Chakraborty, A.K. Ray, A two-stage mechanism for registration and classification of ECG using gaussian mixture model, Pattern Recognit. 42 (11) (2009) 2979–2988.
[5] F. ul Amir Afsar Minhas, M. Arif, Robust electrocardiogram beat classification using discrete wavelet transform, Physiol. Meas. 29 (5) (2008) 555–570.
[6] P. de Chazal, M. O'Dwyer, R. Reilly, Automatic classification of heartbeats using ecg morphology and heartbeat interval features, IEEE Trans. Biomed. Eng. 51 (7) (2004) 1196–1206.
[7] M. Llamedo, J. Martínez, Heartbeat classification using feature selection driven by database generalization criteria, IEEE Trans. Biomed. Eng. 58 (3) (2011) 616–625.
[8] C. Ye, B. Kumar, M. Coimbra, Heartbeat classification using morphological and dynamic features of ECG signals, IEEE Trans. Biomed. Eng. 59 (10) (2012) 2930–2941.
[9] G. de Lannoy, D. Francois, J. Delbeke, M. Verleysen, Weighted conditional random fields for supervised interpatient heartbeat classification, IEEE Trans. Biomed. Eng. 59 (1) (2012) 241–247.
[10] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) e215–e220.
[11] G. Moody, R. Mark, The impact of the MIT-BIH arrhythmia database, IEEE Eng. Med. Biol. Mag. 20 (3) (2001) 45–50.
[12] P. de Chazal, R. Reilly, A patient-adapting heartbeat classifier using ecg morphology and heartbeat interval features, IEEE Trans. Biomed. Eng. 53 (12) (2006) 2535–2543.
[13] T. Ince, S. Kiranyaz, M. Gabbouj, A generic and robust system for automated patient-specific classification of ECG signals, IEEE Trans. Biomed. Eng. 56 (5) (2009) 1415–1426.
[14] J. Wiens, J.V. Guttag, Active learning applied to patient-adaptive heartbeat classification, in: Advances in Neural Information Processing Systems, 2010, pp. 2442–2450.
[15] M. Llamedo, J. Martínez, An automatic patient-adapted ECG heartbeat classifier allowing expert assistance, IEEE Trans. Biomed. Eng. 59 (8) (2012) 2312–2320.
[16] Testing and Reporting Performance Results of Cardiac Rhythm and ST-Segment Measurement Algorithms, ANSI/AAMI Std. EC57:1998, Rev. 2008.
[17] ECGpedia, A Free Electrocardiography (ECG) Tutorial and Textbook. URL ⟨http://en.ecgpedia.org/wiki/Main_Page⟩.
[18] I. Christov, G. Gómez-Herrero, V. Krasteva, I. Jekova, A. Gotchev, K. Egiazarian, Comparative study of morphological and time-frequency ECG descriptors for heartbeat classification, Med. Eng. Phys. 28 (9) (2006) 876–887.
[19] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.
[20] H. Zhu, X. Hu, J. Dong, An improved QRS detection approach based on two-lead bi-referring, in: The IASTED 2012 African Conferences, 2012.
[21] H. Zhu, J. Dong, An R-peak detection method based on peaks of Shannon energy envelope, Biomed. Signal Process. Control 8 (5) (2013) 466–474.
[22] A. Ghaffari, M. Homaeinezhad, M. Akraminia, M. Atarod, M. Daevaeiha, A robust wavelet-based multi-lead electrocardiogram delineation algorithm, Med. Eng. Phys. 31 (10) (2009) 1219–1227.
[23] J. Martínez, R. Almeida, S. Olmos, A. Rocha, P. Laguna, A wavelet-based ECG delineator: evaluation on standard databases, IEEE Trans. Biomed. Eng. 51 (4) (2004) 570–581.
[24] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, C.-J. Lin, LIBLINEAR: a library for large linear classification, J. Mach. Learn. Res. 9 (2008) 1871–1874.
[25] Y.-W. Chen, C.-J. Lin, Combining SVMS with various feature selection strategies, in: Feature Extraction, Springer, 2006, pp. 315–324.
[26] B. Sarojini, N. Ramaraj, S. Nickolas, Enhancing the performance of LIBSVM classifier by kernel f-score feature selection, in: Communications in Computer and Information Science, vol. 40, Noida, India, 2009, pp. 533–543.
[27] M. Wu, J. Ye, A small sphere and large margin approach for novelty detection using training data with outliers, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2009) 2088–2092.
[28] J. Kittler, M. Hatef, R.P.W. Duin, J. Matas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 20 (3) (1998) 226–239.
[29] T.-F. Wu, C.-J. Lin, R.C. Weng, Probability estimates for multi-class classification by pairwise coupling, J. Mach. Learn. Res. 5 (2004) 975–1005.
[30] A.Y. Ng, Feature selection, L_1 vs. L_2 regularization, and rotational invariance, in: Proceedings of the 21st International Conference on Machine Learning, ACM, 2004, pp. 78–85.
[31] T. Evgeniou, M. Pontil, Regularized multi-task learning, in: Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2004, pp. 109–117.

Zhancheng Zhang received the B.S. degree in Material Science from Jingdezhe Ceramic Institute, Jingdezhen, China, in 2000, the M.S. degree in Computer Science from Huazhong University of Science and Technology, Wuhan, China, in 2006, and the Ph.D. degrees from the School of Information Technology, Jiangnan University, in 2012. Currently, he is a Post-Doctor with the Suzhou Institute of Nano-Tech and Nano-Bionics, Chinese Academy of Sciences. His current research interests include pattern recognition, AI in biomedical engineering and Health Informatics.

Jun Dong received the Ph.D. degree from Zhejiang University in 1997. He is currently a Professor in the Suzhou Institute of Nano-Tech and Nano-Bionics, Chinese Academy of Sciences. His current research interests include wearable wireless mobile medical terminals, body characteristic signals recognition and **computer-aided diagnosis.**

Xiaoqing Luo was born in Nanchang, Jiangxi, in 1980. She received the B.S. degree in Computer Science and Technology from Jingdezhen Ceramic Institute, Jingdezhen, China, in 2001, the M.S. degree in Computer Science and Technology and the Ph.D. degree in Pattern Recognition and Artificial Intelligence from School of Information Engineering, Jiangnan University, Wuxi, China in 2007 and in 2010, respectively. She joined the school of Internet of things, Jiangnan University since 2010, where she is an Associate Professor of Computer Science and Technology. She has published more than 25 papers. Her current research interests are image fusion, pattern recognition, and other problems in image processing.

Kup-Sze Choi received his Ph.D. degree in Computer Science and Engineering from the Chinese University of Hong Kong in 2004. He is an Associate Professor at the School of Nursing and an Affiliate Faculty Member at the Interdisciplinary Division of Biomedical Engineering, The Hong Kong Polytechnic University. He is also the Director for the Centre for Smart Health and the PolyU-Henry G. Leong Mobile Integrative Health Centre. His research interests include computing applications in medicine and healthcare, health informatics, simulation and modeling, computer graphics, virtual reality and computational intelligence.

Xiaojun Wu received his B.S. degree in Mathematics from Nanjing Normal University, Nanjing, PR China, in 1991 and his M.S. degree in 1996, and Ph.D. degree in Pattern Recognition and Intelligent System in 2002, both from Nanjing University of Science and Technology, Nanjing, PR China. He was a Fellow of United Nations University, International Institute for Software Technology (UNU/IIST) from 1999 to 2000. From 1996 to 2006, he taught in the School of Electronics and Information, Jiangsu University of Science and Technology where he was an exceptionally promoted professor. He joined the School of Information Engineering, Jiangnan University, in 2006 where he is a Professor. He won the most outstanding postgraduate award by Nanjing University of Science and Technology. He has published more than 100 papers in his fields of research. He was a visiting researcher in the Centre for Vision, Speech, and Signal Processing (CVSSP), University of Surrey, UK, from 2003 to 2004. His current research interests are pattern recognition, computer vision, fuzzy systems, neural networks and intelligent systems.