Comparing performance of Random Forest, SVM and Their Variants for ECG Quality Assessment Combined with Nonlinear Features

Yatao Zhang[1, 2*], Shoushui Wei[1], Li Zhang[3], Chengyu Liu[4]

[1] School of Control Science and Engineering, Shandong University, Jinan, China
[2] School of Mechanical, Electrical & Information Engineering, Shandong University, Weihai, China
[3] Computational Intelligence Group, Northumbria University, Newcastle, UK
[4] School of Instrument Science and Engineering, Southeast University, Nanjing, China

Abstract*—For evaluating performance of nonlinear features and iterative and non-iterative classification algorithms (i.e. kernel support vector machine (KSVM), random forest (RaF), least squares SVM (LS-SVM) and multi-surface proximal SVM based oblique RaF (ORaF) for ECG quality assessment we compared the four algorithms on 7 feature schemes yielded from 27 linear and nonlinear features including four features derived from a new encoding Lempel-Ziv complexity (ELZC) and the other 26 features. The seven feature schemes include the first scheme consisting of 7 waveform features, the second consisting of 15 waveform and frequency features, the third consisting of 19 waveform, frequency and approximate entropy (ApEn) features, the fourth consisting of 19 waveform, frequency and permutation entropy (PE) features, the fifth consisting of 19 waveform, frequency and ELZC features, the sixth consisting of 23 waveform, frequency, PE and ELZC features, and the last consisting of all 27 features. Up to 1500 mobile ECG recordings from the Physionet/Computing in Cardiology Challenge 2011 were employed in this study. Three indices i.e. sensitivity (*Se*), specificity (*Sp*) and accuracy (*Acc*), were used for evaluating performances of the four classifiers on the seven feature schemes, respectively. The experiment results indicated PE and ELZC can help to improve performance of the aforementioned four classifiers for assessing ECG quality. Using all features besides ApEn features obtained the best performances for each classifier. For this sixth scheme, the LS-SVM yielded the highest *Acc* of 92.20% on hidden test data, as well as a relatively high *Acc* of 93.60% on training data. Compared with the other classifiers, the LS-SVM classifier also demonstrated the superior generalization ability.

***Keywords*- ECG quality assessment; Nonlinear Features; Encoding Lempel-Ziv complexity; LS-SVM; Random forest**

1 INTRODUCTION

Combination of several machine learning algorithms (i.e. ensemble decision tree, neural networks and support vector machine (SVM)) and time-frequency features has been used for assessing quality of physiological signals [1-3]. For these methods, two main factors affect their assessment results, and the first factor is effectiveness of features extracted by artificial experience and another factor is performance of several machine learning algorithms on these features. So far time-frequency features are popularly used for assessing physiological signals because calculation of these features are relatively simple as well as time features are more easy to be identified. Li et al. [4] utilize time features derived from the beat to assess quality of pulsatile signals. Orphanidou et al. [5] assessed quality of heart rate variability using signal quality index based on a simple rule, the heart rate calculated from the 30-s window must fall within a physiologically probable range of 40-180 beats per minute. Langley *et al*. [6] and Johannesen [7] determined poor quality of ECG signals when waveform features of signals did not satisfy with the preset thresholds. Several frequency features generated from power spectrum of different ECG frequency bands were employed to assess quality of ECG signals [8]. Clifford et al. [3] and Zhang et al. [2] combined time domain features with power spectrum features so as to achieve relatively satisfied results. Actually, in quality assessment, the ECG recordings are not preprocessed so that waveforms of the recordings are complicated and arbitrary, so it cause poor generalization ability of wave features, furthermore power spectrum features also contain ambiguous information since the frequency range of ECG and that of noise usually overlap. So the nonlinear analysis should be

*Corresponding author: zytboy@sdu.edu.cn
This work was supported by Postdoctoral Science Foundation of China (No. 2017M612280), National Natural Science Foundation of China (No. 61473174), and Natural Science Foundation of Shandong Province, China (No. ZR2014EEM003)

considered to assess quality of physiological signal.

Actually, in [9], the nonlinear analysis i.e. recurrence properties already was used in ECG quality classification, furthermore not only the nonlinear and nonstationary characteristics within 3-lead vector cardiogram yielded from the 12-lead ECG, but also the self-organizing neural network was employed to control the quality of ECG. Ensemble decision tree combined with frequency features derived from power spectrum was reported in [8]. In the works of Kužílek *et al.* [10] and Clifford *et al.* [3, 11], Kernel function based SVM (KSVM) was used. Clifford *et al.* [3] compared three classifiers: naive Bayes, SVM and multilayer perceptron artificial neural network classifiers and verified that SVM is the most valuable method for ECG quality assessment. However, in the actual applications, contaminated ECG segments due to noises and outliers usually result in the absolutely rejection for an "acceptable" recording. In our latest experiment, we proposed a fuzzy SVM method to restrict the effect of the outliers on ECG signal quality assessment [12]. However we further found the fuzzy function could restrain the effect of the essential support vectors on classification by penalizing their contribution weights, especially for the common vague separating hyper plane used in signal quality assessment. Previous study has showed that lease squares support vector machine (LS-SVM) classifier, using the quadratic loss function as the decision function, can reduce the effects of noises and outliers on the classification accuracy, significantly improve accuracy of classification, as well as obtain good generalization ability [13]. Although SVM and its variants exhibit relatively better performance, however a non-iterative method namely, random forest (RaF) has become popular research method in solving classification and regression problems [14-18], and it builds a classification ensemble with a set of decision trees and is comparable in performance to many other non-linear learning algorithms, and it was considered as a competitive classifier among 179 classifiers when tested with 121 datasets [14]. Zhang *et al*. [15] proposed a multi-surface proximal support vector machine based oblique random forest (ORaF) and verified its robust classification performance. It is necessary to compare performance of the iterative methods i.e. SVM and LS-SVM, with that of non-iterative methods i.e. RaF and ORaF on quality assessment field.

At present, the nonlinear complexity methods are not yet reported in quality assessment of physiological signals. In fact a proper nonlinear complexity method can be utilized to classify quality of ECG since the unexpected randomness and the nonlinear chaotic within signals are different component within physiological signals. However the typical complexity methods approximate entropy (ApEn), permutation entropy (PE) and a new encoding Lempel-Ziv complexity (ELZC) have different performance for measuring complexity within signals, so it is necessary to evaluate their performance for reflecting the inherent nonlinear properties within ECG signals for quality assessment. In addition SVM and RaF are regarded as the potential machine learning methods on analysis of physiological signals, so it is necessary to explore efficiencies of the iterative methods i.e. SVM and LS-SVM, and that of the non-iterative methods i.e. RaF and ORaF on several feature schemes including linear and nonlinear features in ECG quality assessment. So in this study, for validating classification performance of the aforementioned four machine learning algorithms on ECG quality assessment and capability of waveform features, frequency features and several common nonlinear feature (i.e. ApEn, PE and ELZC) to reflect the inherent information within signals, we carried out the four algorithms on seven feature schemes respectively. Finally this study tried to find a satisfied classification algorithm and feature scheme for assessing ECG quality.

2 METHOD AND MATERIAL

2.1 Data

All experiment data are derived from the Physionet/CinC Challenge 2011 in the MIT/BIH database and are standard 12-lead ECG recordings (leads I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5 and V6) with a sample of 500 Hz and duration of 10 s [19]. These data collected by smart phone were annotated by clinical experts and technicians as "acceptable" or "unacceptable" ECGs for clinical interpretation. ECG recordings are divided into two subsets: Set A and Set B, wherein Set A includes 1,000 labeled ECG recordings and is used as the training data, and Set B includes 500 ECG recordings where the labels are not publicly available and is used as the testing data. Table 1 details the data profile.

TABEL 1. DATA PROFILE OF THE TRAINING AND TEST SET.

Database	**# recordings**			**Time length**	**Sample rate**
	# acceptable	**# unacceptable**	**# total**		
Training	775	225	1000	10 s	500 Hz
Test	unknown	unknown	500	10 s	500 Hz

2.2 Multiple features calculation

For time and frequency features, we improved the existed quality features based on our previous work [12].

a) Waveform features

The most prominent feature is one of waveform features, namely lead-fall feature. For real time ECG recordings collected by smart phone, poor electrode contact or lead movement could cause signal waveform that seems like a straight line, however, in practice, the waveforms have slight fluctuation. So in this study, lead fall was detected by calculating the difference between the maximum and the minimum amplitude in any lead, and the lead fall was found when the difference was less than 0.025 mV. Finally this study employed the number of leads containing lead-fall in a 12-leads ECG recording as the lead-fall feature F_{Lf} of the ECG recording.

Baseline wander is also relatively more common distortion of waveform and cause poor quality signals for clinic application. In this study, we calculated four features (F_{Bd1}, F_{Bd2}, F_{Bd3} and F_{Bd4}) based on baseline wander. F_{Bd1} and F_{Bd2} denoted the maximum and mean value of the maximum voltage of baseline curve of each lead within 12 leads, respectively.

Additionally, the baseline drift is also considered to occur when amplitude of the baseline is higher than 1.5 mV for lasting more than (continuous) 1.5 s. The Atl_i denoted the accumulated time length of baseline signals of the ith lead lasting larger than 1.5 mV, and i=1, …, 12. Then the two features F_{Bd3} and F_{Bd4} were defined as follows:

$$\begin{aligned} F_{Bd3} &= \max(Atl_i) \\ F_{Bd4} &= mean(Atl_i) \\ i &= 1, 2, \ ..., \ 12 \end{aligned} \quad . \tag{1}$$

Actually, few huge impulses exist in some acceptable ECG recordings, however more number of huge impulses is found in poor quality ECG recordings. This study firstly computed the number of huge amplitudes that were greater than 5.0 mV within each lead of each 12-lead recording, then they were denoted as HA_i (i=1, 2, …, 12). Finally the maximum and mean of the HA_i were calculated and denoted as the quality features FH_{a1} and FH_{a2}, respectively.

b) Power spectrum features

Normal ECG signals have a range of frequency band from 0.05 to 100 Hz. High frequency noise within ECG signals is mainly caused by muscle electricity during periods of contraction or due to a sudden body movement, and its frequency range is from 0 Hz to several kHz. It can be seen that frequency component of QRS complex overlaps with that of high frequency noise. Low frequency noise is mainly caused by baseline wander, and its frequency is usually below 1 Hz, even 0.5 Hz [20]. According to the aforementioned frequency ranges, this study employed the ratio of power spectral density (PSD) in different frequency ranges to that in the overall energy band as quality features to assessing ECG quality. Table 2 shows that several PSD features. In this study, the AR model spectrum estimation algorithm and Burg algorithm were performed to calculate PSD and estimate parameter.

TABLE 2 DEFINITIONS OF THE PSD FEATURES IN EACH LEAD OF 12-LEAD ECG RECORDING.

Features	Description
PSD^n_i	Power of the normal power in the band of 0.05-100 Hz in the *ith* lead
PSD^h_i	Power of high frequence noise in 10-1000 Hz in the *ith* lead
PSD^l_i	Power of low frequence noise in 0-1 Hz in the *ith* lead
$PSD^{h/n}_i$	Power ratio of PSD^h to PSD^n in the *ith* lead
$PSD^{l/n}_i$	Power ratio of PSD^l to PSD^n in the *ith* lead

According to Table 2, the following features (F_{Psd1} to F_{Psd8}) were derived:

F_{Psd1}, F_{Psd2}, F_{Psd3} and F_{Psd4} represented the maximum, the minimum, the mean and the standard deviation of $PSD^{h/n}_i$ respectively.

F_{Psd5}, F_{Psd6}, F_{Psd7} and F_{Psd8} represented the maximum, the minimum, the mean and the standard deviation of $PSD^{l/n}_i$ respectively.

c) Non-linear features

The non-linear features are expected to address the inherent non-linear characteristic in ECG signal. Several non-linear approaches, such as approximate entropy (ApEn) [21], permutation entropy (PE) [22] and our recently developed ELZC method, were employed to analyze the ECG signal quality.

The ELZC method can not only distinguish the chaotic and random characteristics in the ECG recording [23], but also can indicate the noise level contained in the ECG recording, especially for the signals contaminated by high frequence noise. The classical LZ complexity consists of two steps. Firstly, an original time series is transformed into a new binary symbolic sequence by comparing with the mean or median of the original series, and then the LZ value from the binary sequence is calculated. In this study, the original series was transformed into an 8-state symbolic (3-bit binary) sequence by an encoding way.

Each x_i within the original signal $X=x_1, x_2, \ldots, x_n$ is transformed into a 3-bit binary symbol $b_1(i)b_2(i)b_3(i)$, and the detailed process consists of three steps and is described as follows [23]:

Step 1, the $b_1(i)$ is determined by comparing x_i with the mean of signal X, and $b_1(i)$ is set 0 when the x_i is less than the mean, otherwise the $b_1(i)$ is 1.

Step 2, the $b_2(i)$ is 0 when the difference between x_i and x_{i-1} is less than 0, otherwise the $b_2(i)$ is set to 1. Initially, $b_2(1)$ is set to 0.

Step 3, the calculation process of the third digit $b_3(i)$ is relatively complex, where a variable Flag is first denoted as follows:

$$Flag(i)=\begin{cases}0 \ if \ \ |x_i - x_{i-1}| < dm \\ 1 \ if \ \ |x_i - x_{i-1}| \geq dm\end{cases}, i=2,3,...,n, \tag{2}$$

where dm is the mean distance between adjacent points within signal X. Subsequently, $b_3(i)$ is calculated as follows:

$$b_3(i) = NOT(b_2(i) \ XOR \ Flag(i)), \ i=2,3,...,n, \tag{3}$$

where $b_3(1)$ is 0.

After the symbolic process, the LZ value of the new symbolic sequence will be calculated, and the detailed calculation process is illustrated in detail in [23].

In this study, the ELZC was employed as the non-linear ECG signal quality feature. The value of ELZC from the 12 leads were calculated as $ELZC_i$ (i=1, 2, …, 12), and then four ELZC features derived from $ELZC_i$ were defined as follows:

$$\begin{aligned} F_{ELZC1} &= \max(ELZC_i) \\ F_{ELZC2} &= \min(ELZC_i) \\ F_{ELZC3} &= mean(ELZC_i), \\ F_{ELZC4} &= std(ELZC_i) \\ i &= 1,2,...,12 \end{aligned} \tag{4}$$

where F_{ELZC4} is the standard deviation of $ELZC_i$.

Similarly, four PE and ApEn features were also derived from the generated PE and ApEn values from 12 leads as follows:

$$\begin{aligned} F_{ApEn1} &= \max(ApEn_i) \\ F_{ApEn2} &= \min(ApEn_i) \\ F_{ApEn3} &= mean(ApEn_i) \\ F_{ApEn4} &= std(ApEn_i) \\ F_{PE1} &= \max(PE_i) \\ F_{PE2} &= \min(PE_i) \\ F_{PE3} &= mean(PE_i) \\ F_{PE4} &= std(PE_i) \\ i &= 1,2,...,12 \end{aligned} \tag{5}$$

where F_{ApEn4} and F_{PE4} is the standard deviation of $ApEn_i$ and PE_i, respectively.

f) Features normalization: In the study, the feature vector $\boldsymbol{X}$ for each ECG recording was zero-mean normalized as:

$$\boldsymbol{x} = \frac{\boldsymbol{X} - \mu_X}{\sigma_X}, \tag{6}$$

where μ_X and σ_X are the mean and standard deviation of feature vector $\boldsymbol{X}$, respectively. After zero-mean normalization, each feature vector has a mean value of 0.

2.3 Comparative classifiers

2.3.1 Kernel support vector machine (KSVM)

For the classical KSVM, eventually, a nonlinear classification problem can be transform into a dual optimization problem, and it is described as [12, 24]:

$$\begin{aligned} &\max_{\alpha}\left\{\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(\boldsymbol{X}_i, \boldsymbol{X}_j)\right\}, \\ &s.t.\ 0 \le \alpha_i \le C,\ \sum_{i=1}^{n}\alpha_i y_i = 0 \end{aligned} \tag{7}$$

where each $\boldsymbol{X}_i \in R^d$, is a training sample (herein, a feature set of an ECG recording), α_i is the Lagrange multiplier; y_i is the known category of $\boldsymbol{X}_i$; C is the penalty parameter, and K is the nonlinear kernel function. The training procedure of KSVM classifier is essentially a constrained quadratic optimization problem. The decision function is defined as:

$$f(\boldsymbol{X}) = sign\left(\sum_{\boldsymbol{X}_i \in SV}\alpha_i y_i K(\boldsymbol{X}_i, \boldsymbol{X}) + b\right), \tag{8}$$

where b is the bias parameter.

2.3.2 Least squares vector machine (LS-SVM)

According to [13], LS-SVM with Gaussian radial basis function (GRBF) function obtains a more competitive performance than that of the conventional SVM. Let R^d denote the feature space. $\boldsymbol{X}_i \in R^d$ (i=1,2, …, n) is a set of feature vectors and can be treated as the sequence of training points, and $y_i \in$ {-1,1} is the corresponding class label of the $\boldsymbol{X}_i$ (y_i=1 for positive class and y_i=-1 for negative class)

For binary classification problems, the LS-SVM classifier aims to obtain the parameters $\boldsymbol{w}$ and b within a decision function by solving the following optimization problem:

$$\min J(\boldsymbol{w}, b, e) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \frac{C}{2}\sum_{i=1}^{N}e_i^2, \tag{9}$$

Subject to the equality constraints

$$y_i[\boldsymbol{w}^T\varphi(\boldsymbol{X}) + b] = 1 - e_i,\ i = 1,\ ...,\ N, \tag{10}$$

where $\boldsymbol{w}$ is the weight vector; C is a trade-off parameter indicating the relative importance of the model complexity when compared to the training error, namely the penalty parameter; e_i is the training error associated with the i-th sample, and is used to realize soft margins; and $\varphi(\cdot)$ is a nonlinear function which maps the input space into a higher dimensional space. This formulation consists of equality instead of inequality constraints and takes into account a squared error with regularization term similar to ridge regression.

The solution is obtained after constructing the Lagrangian:

$$L(\boldsymbol{w}, b, e; \alpha) = J(\boldsymbol{w}, b, e) - \sum_{i=1}^{N}\alpha_i\{y_i[\boldsymbol{w}^T\varphi(\boldsymbol{X}_i) + b] - 1 + e_i\}, \tag{11}$$

where α_i are the Lagrange multipliers that can be positive or negative in the LS-SVM formulation, and the α_i obeys the equality constraints as follows from the Karush-Kuhn-Tucker (KKT) conditions:

$$\begin{cases} \dfrac{\partial L}{\partial w} = 0 \rightarrow \boldsymbol{w} = \sum_{i=1}^{N} \alpha_i y_i \varphi(\boldsymbol{X}_i) \\ \dfrac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i y_i \\ \dfrac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma e_i,\ i = 1,\ ...,\ N \\ \dfrac{\partial L}{\partial \alpha_i} = 0 \rightarrow y_i[\boldsymbol{w}^T \varphi(\boldsymbol{X}_i) + b] - 1 + e_i,\ i = 1,\ ...,\ N \end{cases} . \quad (12)$$

The linear KKT system can also be written as the following set of linear equations:

$$\begin{bmatrix} 0 & -Y^T \\ Y & \Omega + C^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 1_v \end{bmatrix}, \quad (13)$$

where $Y=[y_1,\ldots,y_N]$, $1_v=[1,\ldots,1]$, $e=[e_1,\ldots,e_N]$, $\alpha=[\alpha_1, \ldots, \alpha_N]$, and the matrix $\Omega(\boldsymbol{X}_i, \boldsymbol{X}_j)=y_i, y_j K(\boldsymbol{X}_i, \boldsymbol{X}_j)$. The Mercer's condition is used for the Ω, and so the kernel function responsible for the nonlinear mapping is provided by $K(\boldsymbol{X}_i, \boldsymbol{X}_j)= \varphi(\boldsymbol{X}_i)^T \varphi(\boldsymbol{X}_j)$, $i, j=1, \ldots, N$.

The decision function of LS-SVM model for classification is provided in the following equations:

$$f(\boldsymbol{X}) = sign(\sum_{\boldsymbol{X}_i \in SV} \alpha_i y_i K(\boldsymbol{X}_i, \boldsymbol{X}) + b), \quad (14)$$

where *SV* denotes the support vector set; $\boldsymbol{X}_i$ is the *i*th support vector; y_i is the known category of $\boldsymbol{X}_i$, α_i is the Lagrange multiplier with $0<\alpha_i<C$; and $\boldsymbol{X}$ is the data to be classified. Parameters α_i^* and b are obtained during the training process.

The kernel function could affect the classification performance of the SVM classifier [25, 26]. The Gaussian radial basis function (GRBF) is a popular kernel function and is employed in this study:

$$K(\boldsymbol{X}_i, \boldsymbol{X}) = \exp(\frac{-\|\boldsymbol{X} - \boldsymbol{X}_i\|^2}{\sigma^2}), \quad (15)$$

where σ is the parameter of Gaussian kernel function.

The selection of GRBF parameter σ in Eq. (15) and error penalty factor C in Eq. (13) affects the precision of the LS-SVM classifier significantly. In fact, there is not a unified theory for the selection of parameter σ and C. The parameters (σ, C) can be set by searching a parameter space for the best evaluating estimator performance score.

In this study, we tested each pair of parameters (σ, C) in the KSVM and LS-SVM classifier and its corresponding classification accuracy using the grid search (GS) method.

2.3.3 RaF

RaF is an ensemble machine learning technique widely used in classification. The basic principle is that a group of "weak learners" is combined to form a "strong learner". It consists of a collection of decision tree classifiers defined as $\{h(\boldsymbol{x}, \theta_k), k=1, \ldots\}$ where θ_k represents identically distributed random vectors. Each tree is grown using training set and random vector θ_k, and casts a unit vote for the most popular class at input vector $\boldsymbol{x}$.

An ensemble of classifiers $h_1(\boldsymbol{x})$, $h_2(\boldsymbol{x})$, …, $h_k(\boldsymbol{x})$ was given, and the training set was generated by random sampling from the distribution of the random vector $\boldsymbol{X}$ and $\boldsymbol{Y}$. The margin function is defined as

$$mg(\boldsymbol{X}, \boldsymbol{Y}) = av_k I(h_k(\boldsymbol{X}) = \boldsymbol{Y}) - \max_{j \neq Y} av_k I(h_k(\boldsymbol{X}) = j) \quad (16)$$

where $I(\cdot)$ is the indicator function [27, 28].

In RaF algorithm, the generalization error is given by:

$$PE^* = P_{\boldsymbol{X}, \boldsymbol{Y}}(mg(\boldsymbol{X}, \boldsymbol{Y}) < 0) \quad (17)$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are random vectors that indicate the probability is over the $\boldsymbol{X}$, $\boldsymbol{Y}$ space and *mg* represents the margin function measure, the extent to which the average number of votes at random vectors for the right output exceeds the average vote for any other output.

2.3.4 ORaF

Zhang and Suganthan [29] proposed an ORaF method using a set of oblique decisions tree based on multi-surface proximal support vector machine (MPSVM) wherein MPSVM was employed to split. The MPSVM is proposed for binary classification problem and seeks two planes in R^n [29, 30]

$$\begin{aligned} \boldsymbol{X} * \boldsymbol{W}_1 - \lambda_1 = 0 \\ \boldsymbol{X} * \boldsymbol{W}_2 - \lambda_2 = 0 \end{aligned} \tag{18}$$

where the first plane ($\boldsymbol{W}_1$, λ_1) is closest to the samples of class1 and furthest form the samples in class 2, while the second plane ($\boldsymbol{W}_2$, λ_2) is closest to the samples in class 2 and furthest from the samples in class 1. Eventually this leads to the following optimization problem:

$$\min_{z \neq 0} \frac{z'Gz}{z'Hz}, \tag{19}$$

where G and H are symmetric matrices in $R^{(n+1)\times(n+1)}$. Finally the two clustering hyperplanes can be found by the eigenvectors corresponding to the smallest eigenvalues of the following two generalized eigenvalue problems:

$$\begin{aligned} Gz = \lambda Hz, \ z \neq 0, \\ Lz = \lambda Mz, \ z \neq 0. \end{aligned} \tag{20}$$

The ORaF is based on oblique decision tree ensemble where the decision tree is growing using heterogeneous test functions. Actually for the oblique decision tree, each decision hyperplane in the internal node of tree classifier is not always orthogonal to a feature axis. According to geometric properties of a randomly selected feature subset from the training set, each internal node is divided into two hyper-classes. Then MPSVM is used for obtaining two clustering hyperplanes, and each hyperplane is closest to one group of the data, and in the meanwhile remains as far as possible from the other group. Finally the test hyperplane for this internal node uses one of the bisectors of the two hyperplanes. Regularization methods, i.e. Tikhonov, axis-parallel and null space approaches, are used for handling the small sample size problem as the tree grows. In this study, the RaF based on MPSVM with Tikhonov approach was employed.

2.4 Evaluation procedure

In this study, three statistics indices i.e. sensitivity (*Se*), specificity (*Sp*) and assessment accuracy (*Acc*) were utilized to evaluate performance of the aforementioned four machine learning classifiers i.e. KSVM, RaF, ORaF and the new LS-SVM. The *Acc* is the ratio of the number of correctly identified recordings (including acceptable and unacceptable) to all training recordings. The *Se* denotes the percentage of unacceptable recordings that are correctly recognized as unacceptable in training data, and it represents the capability of a classifier to correctly identify unacceptable recordings. The *Sp* denotes the percentage of acceptable recordings that are correctly identified as acceptable recordings in training data, and it exhibits the capability of a classifier to correctly identify acceptable recordings.

Fig 1 shows that the flowchart of evaluation procedure, which consists of three steps. In step 1, the multiple features were extracted by analyzing the lead-fall, baseline drift, extreme amplitude, power spectrum and non-linear characteristics of the ECG signals. Feature selection methods can affect results of data analysis [31]. For comparing performance of the ELZC and selecting optimal features, 7 feature schemes were designed (as shown in Table 3) to compare the classification performances of the four classifiers when using different number of signal quality features. In step 2, the zero-mean normalization was employed for the data preprocessing procedure. Then the parameters of the KSVM and LS-SVM classifiers were optimized by the GS method using the 1,000 training ECG recordings, for each of the 7 feature schemes. Thus, the optimized parameters were determined. In step 3, we compared the performances of the KSVM, LS-SVM, RaF and ORaF classifiers, for classifying the ECG recordings as one of two types, i.e., acceptable or unacceptable, by using *K*-fold cross validation (*K*-CV) method. *K*-CV was a relatively effective method to avoid over-fitting because the training sample is independent of the validation sample. For the 1,000 training ECG recordings, a *K*-fold partition of the dataset was created. For each of the *K* experiments, *K*-1 folds were used for training and the remaining one was used for testing. For the evaluation of each feature scheme, the average results from the *K*-fold results were reported, and the results on test dataset were used as the final classifier performance evaluation. In this study, *K* was set to 6.

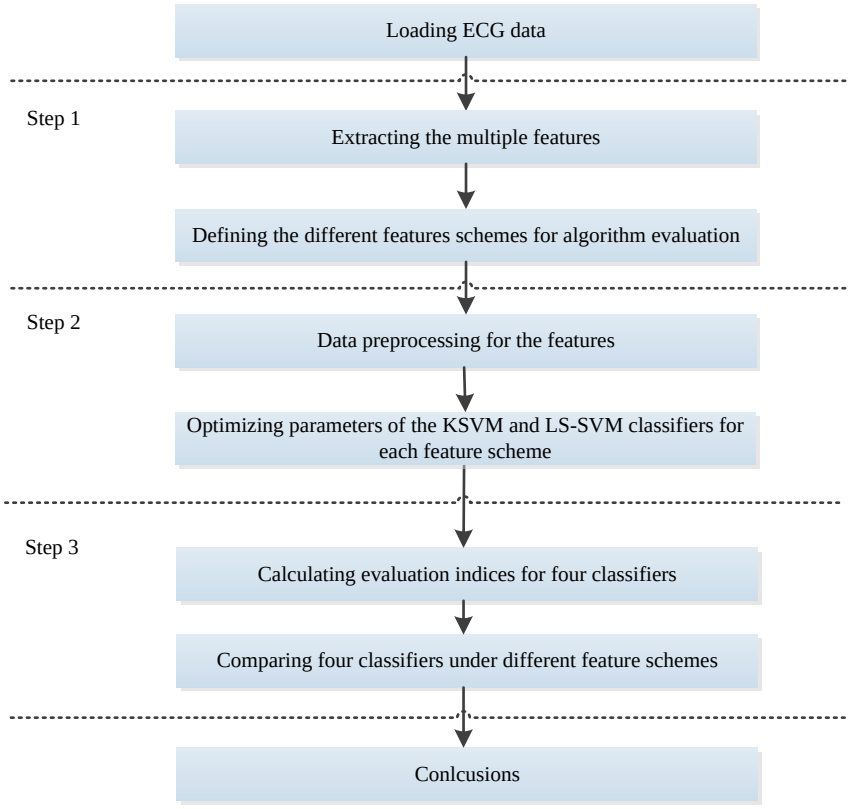

Figure 1. The flowchart of evaluation scheme

TABLE 3 DEFINITION OF THE SEVEN FEATURE SCHEMES.

Scheme number	Features in scheme	Number of features	Description
1	Waveform features: lead-fall, baseline drift and amplitude	7	Time-domain features
2	Waveform features and PSD features	15	Time-domain+ frequency-domain features
3	Waveform features, PSD features and ApEn	19	Time-domain+ frequency-domain + ApEn features
4	Waveform features, PSD features and PE	19	Time-domain+ frequency-domain + PE features
5	Waveform features, PSD features and ELZC	19	Time-domain+ frequency-domain + ELZC features
6	All features except ApEn	23	Time-domain+ frequency-domain + PE + ELZC features
7	All features	27	Time-domain+ frequency-domain + ApEn + PE + ELZC features

3 EXPERIMENT RESULTS

3.1 Parameter optimization for the aformentioned four classifiers

In this section, the GS method was used to search the optimal parameters (C, σ) for the KSVM and LS-SVM classifiers for each of the seven feature schemes. The search ranges were set as 1×10^{-6}~1×10^{5} for both parameters C and σ. Table 4 shows seven feature schemes and the corresponding optimal parameter combinations (C, σ) for the KSVM and LS-SVM classifiers on each scheme, respectively. In addition, in this study, both RaF and ORaF classifiers consisted of 300 trees, each constructed by random feature selected from features, and the number of random features is the square root of the number of all features.

Table 4 shows that the KSVM classifier obtains the highest mean *Acc* of 0.9300, tested upon the training data set using 6-fold cross validation when the parameters C and σ are set as 28526.2000 and 0.0010, respectively. Similarly, for the LS-SVM, the highest mean *Acc*, 0.9390, of cross validation is yielded when the parameters C and σ are 24.5553 and 3.7322, respectively.

TABEL 4 THE RESULTS OF OPTIMIZED PARAMETER COMBINATIONS FOR THE KSVM AND LS-SVM CLASSIFIERS FOR EACH OF THE SEVEN FEATURE SCHEMES USING GS METHOD ON THE TRAIN DATA SET.

Feature schemes	LS-SVM		KSVM	
	C	***σ***	***C***	***σ***
1	1265.4840	1.7361	16384.0000	0.2500
2	362.3092	1.4292	16384.0000	0.2500
3	3.9855	1.2855	4.0000	0.1436
4	0.1769	0.8153	0.2500	2.2974
5	0.7064	1.8205	36.7583	0.0090
6	24.5554	3.7322	28526.2000	0.0010
7	41.5410	11.9837	111.4300	0.0017

3.2 ECG signal quality assessment results using the training data set

This section reports the results of ECG signal quality assessment for the aforementioned seven feature schemes defined in Section 2.4, using the four classifiers, i.e. the KSVM, LS-SVM, RaF and ORaF. The parameter settings for KSVM and LS-SVM classifiers are provide in Table 4 for each of the seven feature schemes. The parameters of the RaF and ORaF classifiers are described in Section 3.1. Table 5 details the classification results for the training data set (i.e. set A) using 6-fold cross validation.

Table 5 shows that the KSVM classifier yields the lowest classification accuracies of 90.00%, 89.80%, 92.30%, 92.80%, 92.40% and 92.60% for feature schemes 1-5 and 7, respectively, using the training set under the 6-fold cross validation, except that it outperforms ORaF for the 6th scheme. The LS-SVM classifier shows steady performance and yields the second highest accuracies of 93.60%, 93.66%, 93.90%, 93.89%, 93.60%, 93.70% for schemes 2, 3, 4, 5, 6 and 7 respectively, and obtains the highest accuracy on scheme 1. The RaF classifier achieves the best performance for nearly all the schemes. It not only yields the highest accuracies on feature schemes 2, 3, 4, 5, 6 and 7, respectively, but also it yields the highest accuracy of 95.41% among all seven feature schemes. The classification accuracies from the ORaF on each single scheme except for the 6th scheme are higher than those from KSVM but lower than those of LS-SVM.

TALBLE 5 CLASSIFICATION RESULTS OF CROSS VALIDATION USING FOUR CLISSIFIERS ON FEATURE SCHEMES 1-7.

Scheme	Method	Results on training set			Results on test set
		Se **(%)**	***Sp*** **(%)**	***Acc*** **(%)**	***Acc*** **(%)**
1	KSVM	57.89±4.85	99.09±1.07	90.00±2.26	89.00
	LS-SVM	79.90±6.23	97.55±0.69	93.60±0.68	91.40
	RaF	80.64±4.92	95.86±1.17	92.50±1.21	89.80
	ORaF	80.35±6.99	97.02±0.58	93.22±2.24	90.80
2	KSVM	57.29±6.08	99.36±1.34	89.80±2.23	89.20
	LS-SVM	80.08±3.55	97.28±1.23	93.60±1.02	91.40
	RaF	90.51±7.56	96.80±2.29	**95.41±1.89**	89.40
	ORaF	68.17±7.66	98.73±1.44	91.90±1.78	91.00
3	KSVM	72.75±11.51	98.07±1.18	92.30±1.42	91.20
	LS-SVM	79.77±5.79	97.54±0.85	93.66±1.73	91.60
	RaF	89.40±3.74	96.63±1.29	95.10±1.06	90.20
	ORaF	79.02±5.34	97.53±1.42	93.29±2.64	91.00
4	KSVM	72.98±5.48	98.57±1.30	92.80±2.09	91.60
	LS-SVM	80.66±7.95	97.67±1.10	93.90±1.58	91.80
	RaF	87.13±4.74	97.26±1.98	94.99±1.60	91.20
	ORaF	78.22±5.61	98.21±0.69	93.60±1.36	91.60

TALBLE 5 (CONTINUED) CLASSIFICATION RESULTS OF CROSS VALIDATION USING FOUR CLISSIFIERS ON FEATURE SCHEMES 1-7.

Scheme	Method	Results on training set			Results on test set
		***Se* (%)**	***Sp* (%)**	***Acc* (%)**	***Acc* (%)**
5	KSVM	75.18±3.05	98.11±1.19	92.40±2.17	91.80
	LS-SVM	81.47±6.88	97.64±1.46	93.89±2.64	92.00
	RaF	88.38±5.93	96.35±2.06	94.59±1.53	91.40
	ORaF	78.55±11.19	98.03±1.15	93.70±1.78	91.40
6	KSVM	73.84±8.31	98.29±1.49	93.00±2.18	92.00
	LS-SVM	77.94±6.46	98.09±2.00	93.60±1.69	**92.20**
	RaF	86.54±4.62	97.04±1.26	94.70±1.22	91.00
	ORaF	76.76±5.24	97.43±1.33	92.90±1.25	92.00
7	KSVM	71.46±4.95	98.44±0.94	92.60±2.80	90.20
	LS-SVM	76.35±6.47	98.59±1.22	93.70±1.66	91.60
	RaF	87.08±7.60	96.86±1.06	94.90±1.70	91.60
	ORaF	75.84±6.62	97.66±0.82	93.00±1.21	92.00

Fig. 2 shows that comparision of classfication accuracies of the four classifiers on training and test data sets, respectively. For each feature scheme on test data, the LS-SVM can yield the optimal accuracy rates except for scheme 7, and the highest accuracy of 92.20% among those of all feature schemes is obtained on scheme 6. Althought the RaF classifier has a good performance on training set, the accuracies from RaF on unseen test data set are relatively lower than those of the other three classifiers, especially for the schemes 3, 4 and 6. Conversely, the KSVM classifier has a poor performance on the training set in comparison with those of other classifers, however it obtains better classification performance, i.e., 91.20%, 91.60%, 91.80% and 92.00%, on schemes 3, 4, 5 and 6, respectively. Classification performance of the ORaF classifier is realtively stable when tested with unseen test data set, and it yields the second highest accuracies of 90.80%, 91.00%, 91.60%, 92.00% and 92.00% on the corresponding schemes 1, 2, 4, 6 and 7.

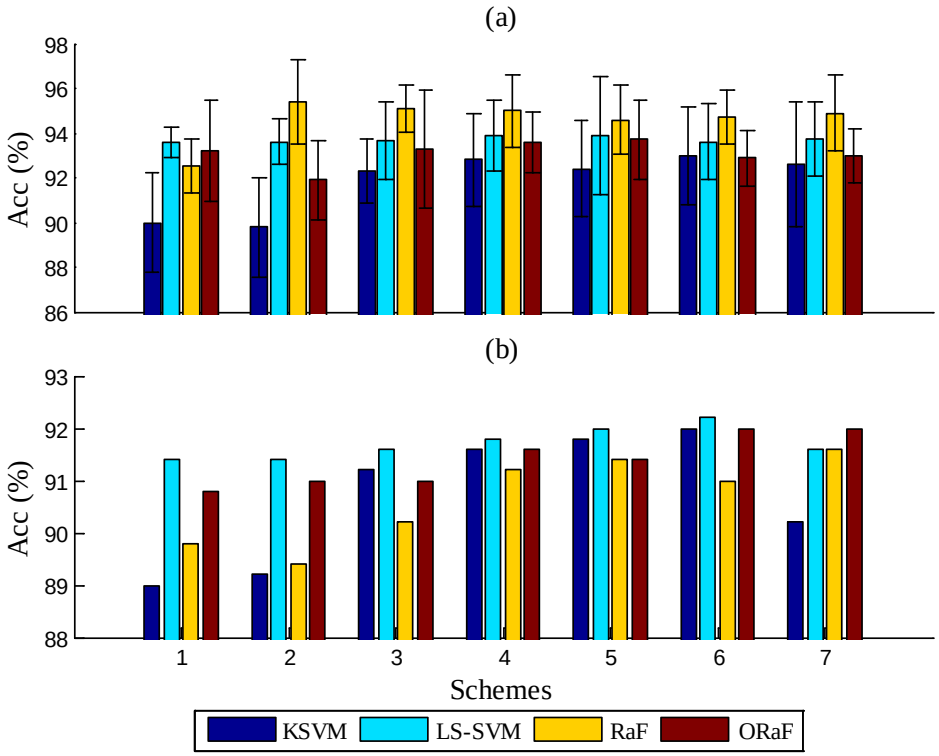

Figure 2. Classification accuracy of the four classifiers. (a) Acc obtained for the training data set(i.e., set A), (b) Acc obtained for the test data set (i.e., set B), respectively.

Fig. 3 shows the difference of classification accuracies between the training and test data sets, aiming to compare generation capabilities of the four classifiers on different feature schemes. Fig. 3 indicates the KSVM algorithm has the least performance differences than those of other classifiers on all feature schemes, except for schemes 6 and 7, and the RaF yields the highest differences on all schemes. In contrast, the corresponding differences of the ORaF on schemes 2, 4, 6 and 7 are comparatively smaller than those of LS-SVM, especially on the schemes 6 and 7.

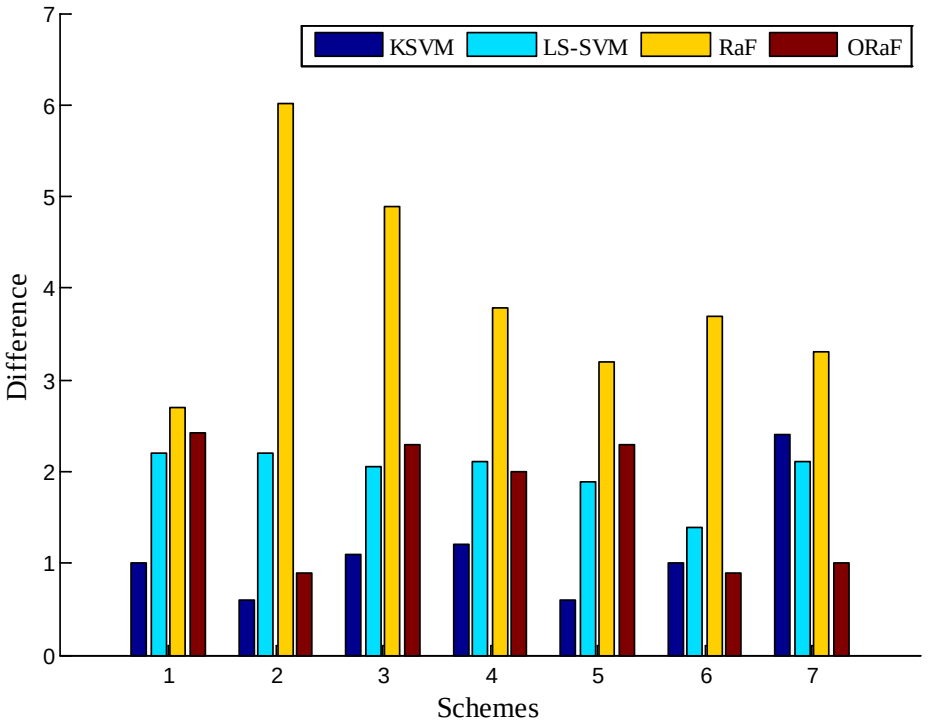

Figure 3. The Acc difference between the training and test data sets on each scheme for each classifier

4 Discussions

In this study, we performed the aforementioned four classifiers on 7 feature schemes aiming to not only compare performance of the classifiers but also evaluate the effectiveness of nonlinear features, especially ELZC feature, on ECG quality assessment.

The results indicate that two types of nonlinear feature, i.e., PE and ELZC, can help to improve classification performance of the aforementioned four classifiers because classification accuracies of nearly all the four classifiers on schemes 4 and 5 are comparatively higher than those of schemes 1, 2 and 3 on both training and test data sets. Feature scheme 1 purely relies on the waveform feature therefore it usually leads to vague and inaccurate conclusions. This is due to the fact that the waveform of considerable amount poor quality ECG signals is difficult to be distinguished from that of high quality ECG signals that can be used for clinical purpose. Similarly, frequency bandwidths of ECG signals and noise tend to overlap, thus the frequency features cannot reflect the inherent properties of ECG signals. This in turn results in comparatively lower classification accuracies or poor generalization ability of the second feature scheme. The nonlinear features can reflect inherent properties within ECG signals since ECG is also nonlinear time series. In contrast, PE and ELZC can discern randomness and nonlinear complexity within ECG signals more explicitly than ApEn, so the classification accuracies from schemes 4 and 5 are relatively higher than those generated by scheme 3. The results also indicate that the ELZC and PE have similar abilities to reflect the nonlinear properties within ECG signals, owing to the fact that accuracies on scheme 4 and 5 exhibit different trends for different classifiers.

The combinations of nonlinear features also do not always guarantee the performance improvement of signal quality assessment, thus the accuracies on schemes 6 and 7 exhibit fluctuation instead of increase. In fact, PE features of some ECG recordings and their ELZC features yield information redundancy or overlap, which in turn causes a decrease in classification performance.

The results indicate that the classical KSVM yields the worst classification results on training data set using 6-fold cross validation, and the results from LS-SVM are better than those of KSVM. In the KSVM method, the constraints are inequality, and support vectors near a separating hyperplane can dramatically impact the calculation of the hyperplane, whereas the vectors that are far away from the hyperplane have little influence to determination of the hyperplane. So classification performance of the classical KSVM method is easy to be weakened when the classifying boundary between two classes is not obviously, or even vague. Therefore, the support vectors are hard to be determined. In fact, the data of ECG quality assessment are usually the raw ECG signals that have not been processed and contain a lot of random components and noises, and so the boundary between acceptable and unacceptable ECG recordings is not obvious. This could be the reason why classification performance of the classical KSVM is relatively weak than that of the LS-SVM method. For the LS-SVM, its constraints are equality, and whether near or far from the separating hyperplane. As mentioned above, the vectors impact the calculation of the hyperplane. So performance of the LS-SVM method could be weakened when the boundary of two classes is clear, whereas its performance is improved when the boundary is vague. This is owing to the fact that the vectors that are the more far away from the hyperplane have more obvious class labels so that they can more accurately calculate the separating hyperplane. It is why the performance of the LS-SVM method is higher than that of the classical KSVM.

The RaF method yields the highest accuracy in each feature scheme for the training data set using cross validation except for scheme 1, but the accuracies are relatively lower for the unseen test set. The results indicate that generalization ability of the RaF is limited. In fact, Fig. 3 exhibits that the RaF method has the worst generalization performance. A possible cause of such poor generalization ability is that each tree is constructed by randomly selected features in the RaF method, and the random selection cannot ensure the construction of an effective RaF. Moreover, Fig. 3 indicates the generalization ability of SVM, i.e., KSVM and LS_SVM is higher than that of RaF. In fact the KSVM has the best generalization performance among the four classifiers.

Furthermore, in comparison to RaF, the ORaF uses SVM to optimize decisions for constructing trees instead of random selection, so the ORaF method improves the classification performance for the test set and exhibits a better generalization ability than that of RaF. However, the accuracy rates of ORaF for the evaluation of the test set are comparatively lower than those of LS-SVM, which further strengthens the robustness of LS-SVM.

5 Conclusions

The aforementioned four classifiers are not able to yield the satisfied classification results when they are performed on waveform and frequency features. However the proposed nonlinear complexity feature ELZC exhibits the same better capability as PE than waveform and frequency features to enhance the performance of assessing the quality of mobile ECG recordings of the four classifiers because classification accuracies of nearly all the four classifiers on schemes 4 and 5 are comparatively higher than those of schemes 1, 2, 3 on both training and test data sets. Similarly, the feature ELZC also exhibits a satisfied performance on enhancing generation capacities of most of the four classifiers except the ORaF classifier than the features PE because the difference of classification accuracies between the training and test data sets on scheme 5 is lower than that on scheme 4. So the proposed nonlinear complexity feature ELZC is more help to improving classification accuracies of the classifiers especially the LS-SVM classifier than the feature PE. Actually the feature ELZC and the LS-SVM classifier can yield the highest classification accuracy 92.00% on test data set among the features schemes 1, 2, 3, 4, and 5 with keeping a relatively lower difference of classification accuracies between the training and test data sets.

A total of 27 features are derived from waveform, power spectrum and non-linear characters of ECG signals, providing comprehensive information for signal quality. For the evaluation of the test data set containing 500 10 s mobile ECG recordings, results showed that LS-SVM classifier achieves the best classification accuracy rate of 92.20% and outperforms other classifiers, i.e. KSVM, RaF and ORaF, consistently.

Acknowledgment

We thank the MIT-BIH database for providing the invaluable data used in this study. We also thank Dr. Le Zhang and Prof. Ponnuthurai N. Suganthan, because they provide the public code of the ORaF based on multi-surface proximal support vector machine.

Conflict of interest statement

The authors declare that there are no conflicts of interest.

REFERENCES

[1] Q. Li and G. D. Clifford, Signal quality and data fusion for false alarm reduction in the intensive care unit, Journal of Electrocardiology. 45(2012) 596-603.

[2] Y. T. Zhang, C. Y. Liu, S. S. Wei, C. Z. Wei and F. F. Liu, ECG quality assessment based on a kernel support vector machine and genetic algorithm with a feature matrix, Journal of Zhejiang University SCIENCE C. 15(2014) 564-57.

[3] G. D. Clifford, J. Behar, Q. Li and I. Rezek, Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms, Physiological Measurement. 33 (2012) 1419-1433.

[4] Q. Li and G. D. Clifford, Dynamic time warping and machine learning for signal quality assessment of pulsatile signals, Physiological measurement, 33(2012) 1491-1501.

[5] C. Orphanidou and I. Drobnjak, Quality assessment of ambulatory ECG using wavelet entropy of the HRV signal, IEEE journal of biomedical and health informatics, 21(2017) 1216-1223.

[6] P. Langley, L. Y. Di Marco, S. King, D. Duncan, C. Di Maria, W. Duan, M. Bojarnejad, D. Zheng, J. Allen and A. Murray, An algorithm for assessment of quality of ECGs acquired via mobile telephones, Computing in Cardiology, IEEE. Hangzhou, 38(2011) 281-284.

[7] L. Johannesen, assessment of ECG quality on an Android platform, Computing in Cardiology, IEEE. Hangzhou, 38(2011) 433-436.

[8] S. Zaunseder, R. Huhle and H. Malberg, "Assessing the usability of ECG by ensemble decision trees," Computing in Cardiology, IEEE. Hangzhou, vol 38, pp. 277-280, 2011.

[9] Y. Chen, H. Yang, Self-organized neural network for the quality control of 12-lead ECG signals, Physiological Measurement. 33(2012) 1399-1418.

[10] J. Kužílek, M. Huptych, V. Chudáček, J. Spilka1, L. Lhotská, Data driven approach to ECG signal quality assessment using multistep SVM classification, Computing in Cardiology, IEEE. Hangzhou, vol 38(2011) 453-455.

[11] Q. Li, C. Rajagopalan, G. D. Clifford, A machine learning approach to multi-level ECG signal quality classification, Computer methods and programs in biomedicine. 117(2014) 435-447.

[12] Y. Zhang, S. Wei, L. Zhang and C. Liu, A signal quality assessment method for mobile ECG using multiple features and fuzzy support vector machine, In Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 12th International Conference on. IEEE. 2016, 966-971.

[13] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle, Least Squares Support Vector Machines, World Scientific, Singapore, 2002.

[14] L. Zhang and P. N. Suganthan, Random forests with ensemble of feature spaces, Pattern Recognition. 47(2014) 3429-3437.

[15] L. Zhang and P. N. Suganthan, Oblique decision tree ensemble via multisurface proximal support vector machine, IEEE transactions on cybernetics. 45(2015) 2165-2176.

[16] L. Zhang and P. N. Suganthan, A Survey of Randomized Algorithms for Training Neural Networks, Information Sciences. 364(2016) 146-155.

[17] L. Zhang and P. N. Suganthan, A comprehensive evaluation of random vector functional link networks, Information Sciences. 367(2015) 1094-1105.

[18] P. Mohapatra, S. Chakravarty, P. K. Dash, Microarray medical data classification using kernel ridge regression and modified cat swarm optimization based gene selection system, Swarm and Evolutionary Computation. 28(2016) 144-160.

[19] G. B. Moody, Physionet/computing in cardiology challenge 2011, July 2011. URL: http://physionet.org/challenge/2011.

[20] P. Shetty and S. Bhat, Analysis of various filter configurations on noise reduction in ECG waveform, International Journal of Computing, Communications & Instrumentation Engg, 1(2014) 88-91.

[21] J. S. Richman, J. R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, American Journal of Physiology-Heart and Circulatory Physiology. 278(2000) H2039-H2049.

[22] C. Bandt, B. Pompe, Permutation entropy: a natural complexity measure for time series. Physical review letters. 88(2002) 174102.

[23] Y. T. Zhang, S. S. Wei, H. Liu, L. N. Zhao, C. Y. Liu, A novel encoding Lempel–Ziv complexity algorithm for quantifying the irregularity of physiological time series, computer methods and programs in biomedicine. 133(2016) 7-15.

[24] C. S. Zhang, Optimization of Kernel function parameters SVM based on the GA, Advanced Materials Research. 433(2000) 4124-4128.

[25] C. S. Zhang, Optimization of Kernel function parameters SVM based on the GA, Advanced Materials Research. 433(2000) 4124-4128.

[26] X. Wang, C. Yang, B. Qin and W. Gui, Parameter selection of support vector regression based on hybrid optimization algorithm and its application, Journal of Control Theory and Applications, 3(2005) 371-376.

[27] Z. Masetic, A. Subasi, Congestive heart failure detection using random forest classifier, Computer methods and programs in biomedicine. 130(2016) 54-64.

[28] L. Breiman, Random forests, Machine learning. 45(2001) 5-32.

[29] A. Khazaee and A. Ebrahimzadeh, Classification of electrocardiogram signals with support vector machines and genetic algorithms using power spectral features, Biomedical Signal Processing and Control. 5(2010) 252-263.

[30] O. L. Mangasarian, E. W. Wild, Multisurface proximal support vector machine classification via generalized eigenvalues, IEEE Transactions on Pattern Analysis and Machine Intelligence. 28(2006) 69-74.

[31] L. Wang, Y. Wang, Q. Chang, Feature selection methods for big data bioinformatics: A survey from the search perspective, Methods. 111(2016) 21-31.