Biomedical Signal Processing and Control 63 (2021) 102138

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Ensemble of kernel extreme learning machine based random forest classifiers for automatic heartbeat classification

Ping Yang, Dan Wang *, Wen-Bing Zhao, Li-Hua Fu, Jin-Lian Du, Hang Su

Faculty of Information Technology, Beijing University of Technology, 100124 Beijing, China

A R T I C L E I N F O

Keywords:
ECG
Heartbeat classification
Inter-patient
Kernel extreme learning machine
Random forest
Ensemble

A B S T R A C T

Automatic heartbeat classification technology based on the ECG plays an important role in assisting doctors with arrhythmia diagnosis. While many heartbeat classification studies can achieve good performance under the intra-patient paradigm, they still cannot offer acceptable classification results under the inter-patient paradigm. Additionally, the available ECG datasets are highly class imbalanced since normal heartbeats appear much more frequently than abnormal heartbeats, resulting in most methods having low sensitives and positive predictive values on minority class ectopic heartbeats. To solve the above problems, this study proposes an automatic ECG heartbeat classification method based on ensemble learning and multi-kernel learning. First, we use a linear combination of the radial basis function kernel and the polynomial kernel to produce a **m**ixed-**k**ernel-based **e**xtreme **l**earning **m**achine (MKELM). Then, a MKELM-based random forest binary classifier (MKELM-RF) is constructed. Finally, an ensemble multiclass classifier MKELM-RF-OVO is proposed based on *one-vs.-one* (OVO) reduction and MKELM-RF. We evaluated the proposed method on the public MIT-BIH-AR benchmarks database, under the inter-patient paradigm, classifying four types of heartbeats, namely, normal beats (N), supraventricular ectopic beats (S), ventricular ectopic beats (V) and the fusion of ventricular and normal (F). The obtained overall accuracy and the average positive predictive value are 98.1% and 93.9%, respectively, which are higher than the current studies by approximately 4% and 6%, respectively. The sensitivities for classes S and V are 1 and 94.4%, respectively, which outperforms most methods. The evaluation results show that our proposed method achieves a superior classification performance compared to the state-of-the-art methods.

1. Introduction

According to the World Health Organization, cardiovascular disease worldwide accounts for approximately 31% of the deaths from diseases [1]. In this study, we concentrate on detecting arrhythmia, which is one of the most important cardiovascular diseases. The electrocardiogram (ECG) is an important diagnostic tool for studying cardiac electrical activity, as it is non-invasive and inexpensive. However, even for professional clinicians, diagnosing various abnormal rhythmic changes from long-term ECG records is exhausting. Hence, computer-aided diagnosis plays a vital role in arrhythmia detection because of its high efficiency and robustness. Arrhythmia detection involves recognizing the class of continuous heartbeats in the given ECG record. Therefore, heartbeat classification is an important part of detecting arrhythmia.

In this paper, we propose a novel ensemble method for automatically classifying four classes of heartbeats according to the suggestion of AAMI:EC57-1998 [2] under the inter-patient evaluation paradigm. The AAMI standard classifies all heartbeat classes into the following 5 classes or super classes: Normal beats (**N**), Supraventricular ectopic beats (**S**), Ventricular ectopic (**V**), Fusion beats (**F**) and Unknown beats **(Q)**. We ignored the Q class, as other studies do [3], since it does not exist in a practical scenario. Three main challenges exist in this work, as follows:

Feature combination space selection. Heartbeat classification is a multiclass classification task, and each type of feature has different classification abilities for different classes of beats. Therefore, it is necessary to find a feature combination space suitable for distinguishing between any two classes of heartbeats.

The dataset is highly class-imbalanced. In ECG signals, the majority of signals represent normal heartbeats, while abnormal heartbeats account for only a small portion. Taking the MIT-BIH Arrhythmia (MIT-BIH-AR) database [4] as an example, the N class accounts for approximately 89% of the total number of heartbeats, whereas S class, V class, and F class heartbeats make up only 3%, 7%, and 1% of the total data, respectively. A large sample size of data may potentially dominate the

* Corresponding author.
E-mail address: neuwd@sina.com.cn (D. Wang).

https://doi.org/10.1016/j.bspc.2020.102138
Received 13 July 2019; Received in revised form 29 February 2020; Accepted 7 August 2020
Available online 1 September 2020
1746-8094/

learning representations and mask useful information of the categories with small sample sizes, and the results will be biased towards the majority class. Therefore, a special modeling strategy is needed to cope with changes in the data density.

Model generalization ability is limited by patient specificity. ECG data are individual-specific, and even for the same class of heartbeat, different individuals have different manifestations of heartbeat. In practical application scenarios, that is, in the inter-patient evaluation paradigm, the test records come from individuals who have never been trained by the model. This requirement makes the automatic classification task very challenging. Therefore, the classification model needs to learn the features that distinguish different types of heartbeats, rather than individual features.

To address the above challenges, we propose an ensemble classifier based on the Mixed-Kernel-based Extreme Learning Machine (MKELM) for automatic heartbeat classification. To learn the feature combination applicable to different heartbeat classes, we constructed a Mixed-Kernel-based Extreme Learning Machine (MKELM). Different from the traditional Kernel Extreme Learning Machine, the high-dimensional space after multi-kernel learning is mapped by multiple kernel functions, which is composed of different sub feature spaces. The combination space can map different feature components through the most suitable kernel function to ensure that the different classes of heartbeats are expressed more accurately and reasonably. To address the class-imbalance problem, we use the Synthetic Minority Oversampling Technique (SMOTE) [5] to generate synthetic data points for the minority samples to balance the number of all types of heartbeats. To improve the generalization ability of the model, we propose a MKELM-based random forest (MKELM-RF) classification algorithm. MKELM-RF employs bootstrap sampling and a random selection of features to train MKELM-based learners, thus feature selection and the model optimization operation can be avoided and the model training efficiency can be improved significantly.

Finally, to obtain robust classification performance, we decompose a multiclass classification problem into multiple binary classification problems. More specifically, we train the MKLM-RF binary classifier to distinguish between any two classes of heartbeats according to the *one-vs.-one* (OVO) strategy. In the testing phase, plurality voting and the proposed class probability are employed to combine the classification results of all MKELM-RFs to obtain the final prediction label for the tested sample.

We evaluated the proposed method on the MIT-BIH-AR database under an inter-patient paradigm. The evaluation results show that our method achieves better performance than the state-of-the-art algorithms. The obtained overall accuracy is 0.981. The sensitivities of classes N, S, V, and F are 99.07%, 1, 94.38% and 7.22%, respectively, and the corresponding positive predictive values are 99.97%, 75.67%, 1 and 1, respectively.

The remainder of this paper is organized as follows: in section 2 we review the literature in this area; section 3 describes the database and our proposed method; in section 4, we present the employed performance measurements, the experiments, and the experimental results. Finally, section 5 concludes the paper.

2. Related Works

Automatic ECG heartbeat classification usually includes the following four main steps: noise removal, heartbeat segmentation, feature extraction, and classification. The overall steps are shown in Fig. 1. First, the noise interference, such as baseline wandering and power line interference, is removed from the raw ECG signal. Second, the noise-removed ECG records are divided into segments in units of heartbeats. Third, a series of features on each heartbeat that can describe the rhythmic characteristics are extracted. Finally, the classification model is trained on the extracted features.

Noise removal. ECG signals are often contaminated with noises and artifacts. There are two main types of noise. One is baseline wandering, which is usually low frequency noise caused by respiration or the motion of the subject during the recording phase. The other type is high frequency noise, such as motion artifacts generated by the electrode-skin interface, power line interference, and electromyography interference, caused by myoelectric contraction [6].

Acharya and Oh [7] used Daubechies wavelet 6 filters to perform denoising and baseline removal. However, selecting the mother wavelet and fixing the decomposition for the non-stationary baseline correction and noise suppression adaptively is difficult with the wavelet denoising method [8]. Garcia and Moreira [9] utilized two finite impulse response (FIR) filters, i.e., a 12 tap low pass filter with −3 dB at 35 Hz and a high-pass filter with −3 dB at 1 Hz. However, this type of technique has a fixed cut-off frequency, which significantly distorts the ST segment and QRS complex. In reference [10], the raw ECG signals were processed with a 200-ms width and a 600-ms width median filter to obtain a baseline, which is subsequently subtracted from the original signals to yield the baseline-corrected ECG signals. High-frequency noise was then removed by using a 12-order FIR low-pass filter with a 35 Hz cut-off frequency. In general, most of the research works [11–13] usually remove the baseline followed by filtering high-frequency noise in the heartbeat classification task. Different from these works, to keep as much of the original ECG signal information as possible, we only perform the baseline removal and do not perform any high frequency noise filtering, as in reference [14]. More specifically, a 200-ms width median filter and a 600-ms median filter are used sequentially to obtain the baseline, which is subsequently subtracted from the original ECG signals to yield the baseline-corrected signals.

Heartbeat segmentation. Each ECG record is composed of continuous heartbeats. Generally, the heartbeat position needs to be detected according to the QRS complex. Many methods for heartbeat segmentation achieve near perfect results in public databases such as MIT-BIH-AR. The Pan-Tompkins algorithm [15] is the most widely used QRS complex detection technique; it recognizes QRS complexes based upon a digital analysis of the slope, amplitude and width. Some wavelet transform methods [16–17] have also achieved approximately 99% sensitivity and positive predictive value for detecting QRS complexes. Since the objective of our study is to develop a heartbeat classifier, we use the QRS annotation included in MIT-BIH-AR to locate the R-peak.

Feature Extraction. In the early stage of extracting ECG signal features, time domain features, such as the RR interval, amplitude, and

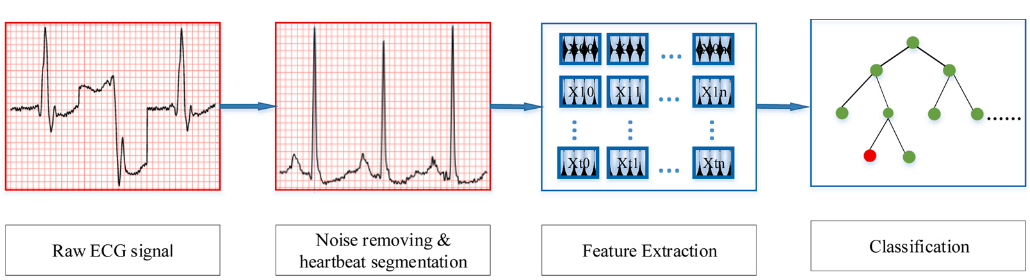

Fig. 1. Diagram of the general steps of the ECG arrhythmia classification.

QRS complex duration have received widespread attention [18]. The RR interval has especially received much attention because it can be used for heart rate variability analysis, is considered to be the most important type of time domain feature and is utilized in almost all arrhythmia detection tasks. However, extracting such features requires detecting the fiducial points of heartbeats. To avoid the exact time alignment problem of ECG waveforms, some transformation techniques have been proposed for extracting features in the frequency domain. In the frequency domain, spectral coefficients [19] and subband coefficient [20] based features are well employed. However, the lack of sufficient variations in the spectrum features in keeping with the pathological conditions caused a limitation [21]. The non-stationary and nonlinear characteristics of the ECG waveforms are the main challenges faced by the time-domain and the frequency-domain feature extraction approaches. Then, time-frequency techniques appeared as a solution, such as Short-time Fourier transform [22], Wigner-Ville distribution [23], empirical mode decomposition [24] and wavelet transform. The wavelet transform methods are especially widely explored [25–28] owing to their good performance in analyzing non-stationary signals. In addition, due to the non-linear nature of ECG signals, some non-linear features, such as higher-order statistics [29–30] and entropy measures [31] are often utilized for arrhythmia recognition.

The abovementioned time-domain, frequency-domain, time-frequency, and non-linear features have been successfully used for heartbeat classification tasks. Some studies further perform feature selection on feature sets to reduce the feature dimensions and select the most appropriate feature combination. Llamedo and Martínez [32] used the sequential forward floating search method for feature selection. Li and Yuan [33] utilized the genetic algorithm to optimize the features and parameters of neural networks. Asl and Setarehdan [34] utilized generalized discriminant analysis to reduce the feature dimension of the input heart rate variability signal. Feature selection can improve the classification performance. However, the time complexity of the feature selection process is high, thus a trade-off between the efficiency and accuracy of the model is required. Different from the above works, our proposed MKELM-RF method randomly selects the features without feature selection, which can greatly improve the training speed of the model.

Classification. There are two popular paradigms for the evaluation of the heartbeat classification task, i.e., intra-patient and inter-patient [35]. In the intra-patient paradigm, the heartbeats in the testing set and training set can come from the same individual. The model can learn the particularity of the patient's heartbeat during the training process and use it in the testing process, making the model classification score higher than in actual application scenarios. In the inter-patient evaluation paradigm, the heartbeats of the training set and the testing set are from different individuals. Since, in practice, the heartbeats to be detected cannot all come from known individuals, the inter-patient evaluation paradigm is recommended by recent studies [9]. In this paper, we follow the inter-patient scheme.

To build a classification model, many previous works have demonstrated the feasibility of machine learning algorithms for the heartbeat classification task, including methods such as the weighted Linear Discriminator [36], Support Vector Machine (SVM) [37,38], Multilayer Perceptron (MLP) [39], Convolutional Neural Network [40], Especially Extreme Learning Machine (ELM) and the Kernel Extreme Learning Machine (KELM). Cömert and Kocamaz [41] used an Artificial Neural Network (ANN) and ELM to classify fetal heart rates, and the experimental results shows that ELM achieved higher classification accuracy. Mahdiyah and Irawan [42] integrated the data selection process into ELM to overcome the data imbalance problem and compared the proposed method with ELM, SVM and BP. The results on multiple imbalanced data sets show that the proposed method has dual advantages in terms of training speed and classification accuracy. Diker [43] and Avci [44] employ a wavelet kernel ELM and use the genetic algorithm to optimize the kernel coefficients, obtaining a 5% higher accuracy than ELM. Different from the above work, we consider that in the heartbeat classification task, various features have different expression abilities for each class of heartbeat. It is necessary to select a suitable feature space for different classes of heartbeats. In this paper, multi-kernel learning is introduced into the KELM, and a MKELM classsifer is constructed.

An ensemble of classifiers can comprehensively consider the results of multiple base classifiers to obtain better classification performance than a single one. Therefore, many studies constructed ensembles of classifiers. Elhaj and Salim [44] propose a hybrid classification techniques using Bayesian and ELM for heartbeat recognition. The classification accuracy of the model is slightly higher than that of the ELM model, and the calculation time is only 1% of SVM and the neural network. Zhou and Tan [45] performed heartbeat classification by using a hybrid network CNN-ELM, using CNN to extract features and ELM as a classifier. The hybrid model obtained a higher score than CNN, and the training speed was faster than CNN. Rahhal and Bazi [46] propose an interactive ensemble learning method based on ELM and induced an ordered weighted averaging operator. In the inter-patient evaluation paradigm, the overall accuracy of the model on the MIT-BIH-AR database is 99%. Their method achieved good classification results. However, the method requires manual labeling of selected heartbeats.

Different from the above methods for building an integrated classifier, we propose a random forest algorithm based on a MKELM. Using MKELM as the base learner of random forest (RF), on the one hand, it can take advantage of the extremely fast learning speed of KELM; on the other hand, it can take advantage of the characteristics of random feature selection and sample selection to avoid extra feature selection operations. In addition, unlike the above works, which directly classify multi-class heartbeats, this paper decomposes the multiclass classification task into multiple binary classification tasks and builds a multiclass classifier integrated by multiple binary classifiers. Since the class probabilities of the test samples given by multiple binary classifiers can be comprehensively considered, the robustness of the model can be improved.

3. Materials and methodology

3.1. MIT-BIH-AR database

The MIT-BIH-AR Database contains 48 two-channel ECG records from 47 patients, one of which was a modified limb lead II (ML II lead), and the other was a modified V1 lead in most cases (occasionally V2 or V5 lead, only one V4 lead). Each record was approximately a half-hour long and was digitized at the sampling frequency of 360 per second per channel with 11-bit resolution. At least two cardiologists independently annotated each record; disagreements were resolved to obtain unanimous reference annotations for each heartbeat; approximately 110,000 heartbeats in total are included in the database. Since the Q class is the unclassifiable heartbeat marked in MIT-BIH, no such heartbeat exists in practical applications. As in other studies, this work classifies the four types of heartbeats of N, S, V, and F.

The MIT-BIH database contains 23 records numbered starting from 100, which are named the 100 series, and 25 records numbered from 200, which are named the 200 series. The "100 series" records are clinically common, while the "200 series" records are rare but clinically important. Four records (102, 104, 107 and 217) include paced beats. According to the AAMI recommendation, the above four paced beats are excluded from our experiments. To better evaluate the actual performance of the classifier and to ensure a fair competition between our methods and those from other works, we adopted the inter-patient paradigm division proposed by reference [35], which divided the database into two equal-sized training and testing datasets, each containing 22 records, as shown in Table 1.

In this study, we built the classifier utilizing the heartbeats extracted from a single lead, the ML II lead. This is mainly due to the following considerations: only the ML II lead is available for all records from the

Table 1
Records of training set DS1 and testing set DS2.

Dataset	Training set DS1	Testing set DS2
Records	101,106,108,109,112,114,115,116, 118,119,122,124,201,203,205,207, 208,209,215,220,223,230	100,103,105,111,113,117,121,123, 200,202,210,212,213,214,219,221, 222,228,231,232,233,234

MIT-BIH Arrhythmia Database; ML II lead is the most commonly used lead for cardiologists in analyzing ECG signals; an increasing amount of ECG monitoring is performed using one-lead portable devices, which need an efficient heartbeat classification algorithm based on single-lead ECG data.

3.2. Preprocessing

To remove noise, most of the previous works usually performed baseline removal and a high-frequency noise filtering in sequence. To keep the ECG signal as raw as possible, we only performed the baseline removal on the raw ECG signal. Following the methods used in the previous work [14], all the ECG signals were preprocessed utilizing a 200-ms width median filter to remove the P waves and QRS complexes followed by a median filter of 600-ms to remove the T waves. The processed signals were then treated as the baseline. Then, the baseline was subtracted from the original signals to produce the baseline-corrected ECG signals. A schematic diagram of pretreatment steps is shown in Fig. 2.

The filtered ECG time sequences were then segmented into individual heartbeats. Since the main purpose of this work is to build the arrhythmia classifier, we used the QRS annotations included in the MIT-BIH-AR database.

3.3. Feature extraction

Taking the detected R-peak as the center, we take 90 samples forward and backward; all 181 samples in this region are seen as one heartbeat. All features are extracted in that region. In view of the RR-interval, the ECG signal morphological features, wavelet transform coefficients, higher order statistics (HOS), and 1D-Local binary patterns (1D-LBP) have good performances in similar tasks. We employ the same features and the processing steps as in reference [14].

The RR-interval features are defined as the interval between R-peak points of successive heartbeats and are the most important features of the arrhythmia classification task. The morphological features of the ECG signal are calculated from the amplitude values of the ECG signal sample points. The most common method is to set a window of a certain

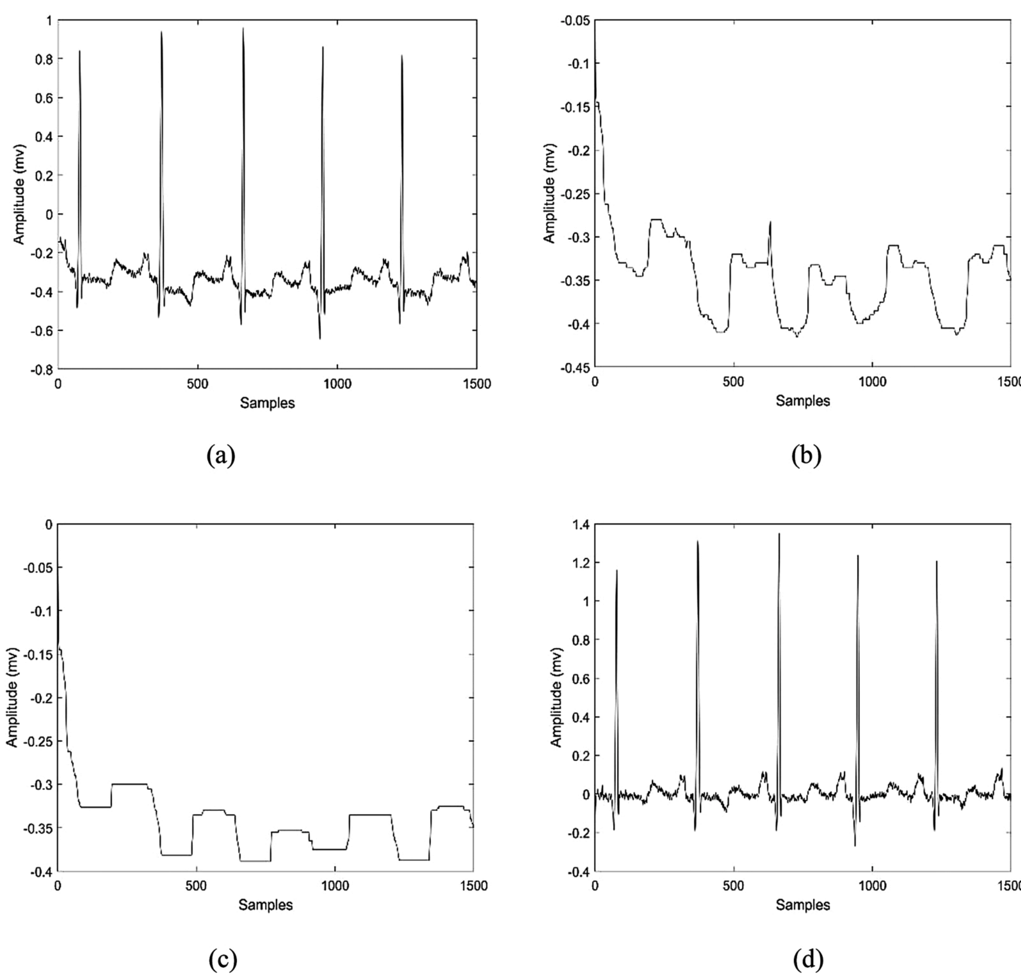

Fig. 2. Example of the ECG signal preprocessing steps. (a) Raw ECG signal. (b) Preprocessing step 1: using a 200-ms width median filter to remove the P waves and QRS complexes from (a). (c) Preprocessing step 2: using a 600-ms width median filter to remove the T waves from (b) and obtain the baseline. (d) Preprocessing step 2: subtracting baseline (c) from the raw ECG signals (a) to obtain the baseline removed ECG signal.

size in a heartbeat cycle and down sample the samples in this window; then, take the amplitude values or distance values of the sample points as the morphological features. The wavelet transform can extract the time domain and frequency domain information of ECG signal at the same time, and the obtained coefficients can be used as the wavelet transform features. The higher order statistics, which are commonly used in the ECG heartbeat classification tasks, are kurtosis and skewness. Each heartbeat interval is divided into several sub-intervals and the kurtosis and skewness are calculated in subintervals.

The local binary pattern (LBP) was proposed by Ojala [47] to describe the local texture features of images, with the advantages of rotation invariance and gray invariance. Since the ECG signal is one-dimensional, 1D-LBP will be used. A window of size w is set, each sample point P in the heartbeat interval is taken as the center, and the value of point P is compared with the value of each point in the w neighborhood of P. The size is set to 1 if the value of point P is larger; otherwise, it is set to 0, thus obtaining a binary pattern. The LBP feature is calculated using the statistical values of the frequency of occurrence of all patterns in a heartbeat. The features, feature descriptions and the number of features are summarized in Table 2.

3.4. Classifier Model

3.4.1. Kernel Extreme Learning Machine

Huang et al. [48] proposed a special single-hidden layer feedforward neural network, called the Extreme Learning Machine (ELM). The ELM only needs to preset the number of hidden layer nodes, the input weights and hidden layer biases can be randomly assigned without tuning. In addition, the output weights of ELM can be determined through a simple generalized inverse operation of the hidden layer output matrices. Therefore, the learning speed of ELM can be extremely fast. The ELM can reach the smallest norm of weights; thus, it obtains better generalization performance than gradient descent-based algorithms.

Although the ELM has the above advantages, the generalization performance of the ELM is unstable due to the random setting of the number of hidden layer nodes. Inspired by the support vector machine (SVM), Huang [49] introduced kernel learning into the ELM and proposed the Kernel Extreme Learning Machine (KELM). Compared with ELM, KELM is able to reach the least square optimization solution; thus, the KELM has better and more stable generalization performance than ELM. The smallest training error and output weights can be obtained by kernels at the same time, while the learning speed of the neural network can be raised, as shown in (1):

$$Min : ||H\beta - T||, ||\beta||, \tag{1}$$

where H is the hidden layer output matrix of the neural network, T is the output matrix, and β is the weight vector connecting the hidden nodes and the output nodes. (1) can be solved by the least squares solution in (2), as follows:

$$\beta = H^T[\frac{1}{C} + HH^T]^{-1}T, \tag{2}$$

where C is the penalty coefficient.

The output function of ELM can be defined as follows:

$$f(x) = h(\mathrm{x})\beta = h(x)H^T[\frac{1}{C} + HH^T]^{-1}T, \tag{3}$$

where $h(x)$ is the feature mapping, which does not need to be known to users. Mercer's conditions [50] are used to describe the kernel matrix of ELM, as represented as follows in (4):

$$M = HH^T : m_{ij} = H(x_j) = k(x_i, x_j). \tag{4}$$

Consequently, the ELM output function (3) can be stated as follows in (5):

$$f(x) = [k(x, x_1)...k(x, x_N)][\frac{1}{C} + M]^{-1}T. \tag{5}$$

3.4.2. Mixed-kernel-based extreme learning machine: MKELM

The traditional KELM is a single kernel-based approach. Since different kernel functions give different similarity measures for the sample points, the performance of different kernel functions may vary greatly on the same dataset. The ECG signal has the characteristics of a large volume, uneven distribution of samples caused by high-dimensional feature space, and an imbalanced class. Using a single kernel to process ECG signals cannot solve the above problems well.

The kernel function can be categorized into a local kernel function or a global kernel function depending on whether it has translation invariance or rotation invariance [51]. The local kernel function is good at extracting the local features of samples, while the global kernel function is superior at extracting global features. In multi-kernel learning, the optimal kernel is supposed to be a linear combination of a set of base kernels, and the ideal linear combination coefficients and the classifier parameters are jointly learned by maximizing the margin [52]. The radial basis function kernel (RBF kernel) [53] and the polynomial kernel [53] are a local kernel function and global kernel function with good performance, respectively. To balance both the generalization ability and classification performance, a mixed kernel-based ELM (MKELM) is produced by linearly combining the RBF kernel and polynomial kernel. The mixed kernel function is defined as follows:

$$K_{mix} = \lambda k_{rbf}(x, x_i) + (1 - \lambda)k_{poly}(x, x_i), \tag{6}$$

where $\lambda(0 < \lambda < 1)$ is the weight coefficient of the linear combination.

$$k_{rbf}(x, x_i) = \exp(\frac{-||x - x_i||^2}{2\sigma^2}), \tag{7}$$

$$k_{poly}(x, x_i) = (x \cdot x_i + 1)^d, \tag{8}$$

where d is set to 2, because the dimension of the polynomial space is n^d; when the sample size is equal to 1000 and the index is equal to 3, the dimension has reached 1 billion, and the calculation of the inner product will cause a dimensional disaster.

Finally, the output function of the MKELM is defined as follows:

$$f(x) = [K_{mix}(x, x_1)...K_{mix}(x, x_N)](\frac{1}{C} + M)^{-1}T. \tag{9}$$

3.4.3. MKELM-based random forest : MKELM-RF

MKELM-RF uses the bootstrap sampling and random attribute selection ideas to train the MKELM base learners and utilizes the majority voting method to combine theses base learners. The overall framework is as follows, where the feature number of the training sample is f and the number of base classifiers is n. First, we use the bootstrap sampling

Table 2
All classes of features used in this work.

Features	Description	Size
Pre-RR	Distance between the previous heartbeat and the current one	1
Post-RR	Distance between the next heartbeat and the current one	1
Local-RR	Average of 10 Pre-RR	1
Global-RR	Average of the last 20 min's Pre-RR	1
Normalized RR	Dividing Pre-RR, Post-RR, Local-RR, Global-RR by its mean value within the same ECG record	4
morphological	Euclidean distance between the R-peak and amplitude values over the 4 intervals: max(beat[0,40]), min(beat[75,85]), min(beat[95,105]),max(beat[150,180])	4
Wavelets	Function db1 with 3 levels of decomposition	23
HOS	Dividing each heartbeat into five intervals and computing the skewness and kurtosis with 5 intervals	10
LBP	1D uniform-LBP 8bits	59
Total	———	104

method to sample n times on the training dataset and generate n sampling subsets. Second, k features are randomly selected in each sampling subset to form n training subsets. Then, n MKELM base classifiers are trained on the generated n training subsets. Finally, the MKELM-RF classifier is produced by integrating all the base classifier with the majority voting combination strategy.

If x is a testing sample, the base classifier h_i outputs a label from the classification label set $\{c_1, c_2, ..., c_N\}$. Let the N-dimensional vector $(h_i^1(x); h_i^2(x); ...; h_i^N(x))$ and $(f_i^1(x); f_i^2(x); ...; f_i^N(x))$ denote the output label and classification confidence of h_i on sample x, respectively. The output class label of the MKELM-RF for the sample x is obtained by majority voting, as follows:

$$H(x) = \begin{cases} c_j, & \text{if} \sum_{i=1}^{T} h_i^j(x) \ > \ 0.5 \sum_{k=1}^{N} \sum_{i=1}^{T} h_i^k(x); \\ \text{reject}, & \text{otherwise}. \end{cases} \tag{10}$$

The sigmoid function is used to convert the classification confidence into the class probability as follows:

$$P(c_j \middle| f_i^j(x)) = \frac{1}{1 + \exp(-f_i^j(x))}. \tag{11}$$

After determining the class label c_j of the testing sample x by the majority voting method, the base classifier whose output label equals c_j is considered to be a correct-classifier. The class probability is randomly chosen from one of the correct-classifiers as the class probability of MKELM-RF. In the multiclass integration problem in the next section, when the number of votes for some labels is the same, the confidence of the voting is examined by class probability to determine the final winner.

3.4.4. MKELM-RF and OVO strategy-based multiclass ensemble model

The KELM classifier can be directly used for multiclass classification tasks, but when solving a large amount of data, the model has a low learning speed, is prone to over-fitting, and has a poor generalization ability. Some binary classification algorithms, such as the SVM, can be turned into classifiers that perform multiclass classification tasks through multiple strategies. These strategies reduce the multiclass classification to multiple binary classification problems. There are two most important problem transformation techniques, i.e., one-*vs.*-one (OVO) and one-*vs.*-all (OVA). In the OVO reduction, one trains $C(C-1)/2$ binary classifiers for a C-way multiclass task, and each takes the samples of a pair of classes from the original training set and has to discriminate between these two classes. The OVA technique trains a single classifier per class, with the samples of that class as positive samples and all other samples as negative samples. Therefore, the OVA strategy involves C binary classifiers.

The amount of training data for each class in the OVA technique is not balanced, and the models are susceptible to being biased. Therefore, this work utilizes the OVO strategy to construct a multiclass classifier MKELM-RF-OVO based on MKELM-RF and OVO strategy. In our heartbeat classification task, since the number of classes is four, six groups of N-S, N-V, N-F, S-V, S-F, and V-F classifiers need to be trained. Each group of classifiers is made up of MKELM-RF. As in section 3.4.3, if $\{c_1, c_2, ..., c_N\}$ is a N-dimensional label set and the output label and class probability of each MKELM-RF classifier g_i on sample x are N-dimensional vectors $(g_i^1(x); g_i^2(x); ...; g_i^N(x))$ and $(p_i^1(x); p_i^2(x); ...; p_i^N(x))$, respectively, $g_i^j(x)$ is the output class label of g_i, and $p_i^j(x)$ is the class probability of g_i on class label c_j. When classifying an unseen sample x, the plurality voting method is used to determine the class label of x. The plurality voting method is formulated as follows:

$$H_1(x) = C_{\underset{j}{\text{argmax}} \sum_{i=1}^{t} h_i^j(x)}. \tag{12}$$

If more than one class label obtains the maximum number of votes at the same time, the class probability of these class labels is calculated, and the class label with the maximum class probability is the final classification label, which is given by the following:

$$H_2(x) = C_{\underset{j}{\text{argmax}}} \sum_{i=1}^{T} p_i^j, \text{s.t.} c_j \in H_1. \tag{13}$$

The overall framework of the multiclass classifier MKELM-RF-OVO based on the MKELM-RF and OVO strategy is illustrated in Fig. 3.

4. Experimentation

4.1. Performance evaluation

In this paper, six important indicators: the overall accuracy (*ACC*), sensitivity (*SEN*), positive predictive value (*PPV*), Cohen's kappa coefficient (k), *jkindex* and *jindex* were used for measuring the classification performance. All the indicators can be calculated from the confusion matrix in Fig. 4.

The overall accuracy reflects the proportion of the correctly classified samples in the total sample, as follows:

$$ACC = (n_{NN} + n_{SS} + n_{VV} + n_{FF}) \Big/ \sum. \tag{14}$$

The sensitivity, which is also called the recall and the true positive rate, evaluates the proportion of actual positives that are correctly detected as such.

$$SEN_N = n_{NN}/T_N,$$

$$SEN_S = n_{SS}/T_S,$$

$$SEN_V = n_{VV}/T_V,$$

$$SEN_F = n_{FF}/T_F. \tag{15}$$

The positive predictive values measure the proportions of positive results in the statistics and diagnostic tests that are true positive results, and it can be interpreted as indicating the accuracy of a classification result. It should be noted that, according to the AAMI specifications, when a classifier classifies an F class as a V class, it will not be rewarded or punished. That is, when calculating the positive predictive value of the V class heartbeat, it is not necessary to consider the influence of n_{FV}. The positive predictive values are defined as follows:

$$PPV_N = n_{NN}/P_N,$$

$$PPV_S = n_{SS}/P_S,$$

$$PPV_V = n_{VV}/(n_{VV} + n_{NV} + n_{SV}),$$

$$PPV_F = n_{FF}/P_F. \tag{16}$$

Cohen's kappa coefficient (k) is a statistic that evaluates the inter-rater agreement for categorical items. The kappa coefficient can comprehensively reflect the confusion matrix and has a more robust evaluation performance on the imbalanced dataset than the overall accuracy or the average accuracy. The kappa coefficient is calculated as follows:

$$k = (p_0 - p_e)/(1 - p_e),$$

$$p_0 = (n_{NN} + n_{SS} + n_{VV} + n_{FF}) \Big/ \sum,$$

$$p_e = (T_N P_N + T_S P_S + T_V P_V + T_F P_F) \Big/ \sum^2. \tag{17}$$

Considering the high class imbalance of the heartbeat classification tasks, some studies further introduced new measurement indexes, called the *j*index [58] and *jk*index [57]. The *j*index appraises the discrimination of the most two vital arrhythmias S and V. The *jk*index is a combination

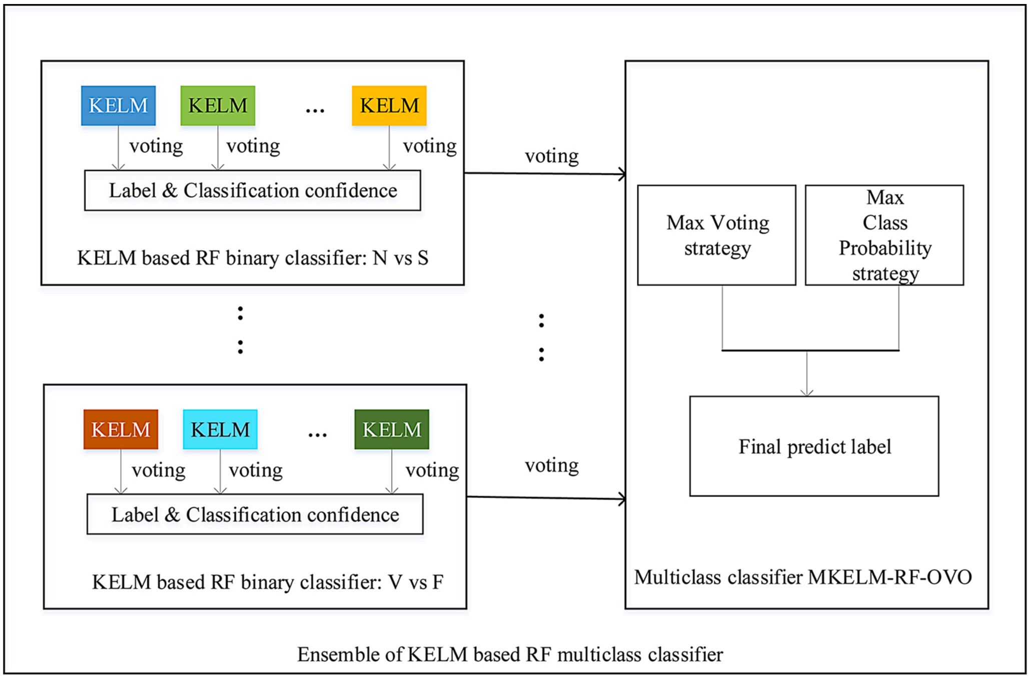

Fig. 3. Framework of the ensemble multiclass classifier MKELM-RF-OVO; RF: random forest.

Predicted labels

Reference labels	N	S	V	F	$\sum$
N	n_{NN}	n_{NS}	n_{NV}	n_{NF}	T_N
S	n_{SN}	n_{SS}	n_{SV}	n_{SF}	T_S
V	n_{VN}	n_{VS}	n_{VV}	n_{VF}	T_V
F	n_{FN}	n_{FS}	n_{FV}	n_{FF}	T_F
$\sum$	P_N	P_S	P_V	P_F	n^T

Fig. 4. Confusion matrix. n_{NS} represents the number of samples with reference label N and predicted label S. T_N indicates the total number of samples whose actual labels are N. P_N indicates the number of samples whose predicted labels are N. n^T represents the number of samples correctly classified.

of the j index and the k value. The $j$index and jkindex are formulated as follows:

$$j\ \text{index} = SEN_S + SEN_V + PPV_S + PPV_V, \tag{18}$$

$$jk\ \text{index} = w_1 k + w_2 j\ \text{index}. \tag{19}$$

where $w_1 = 1/2$ and $w_2 = 1/8$ since k takes its value in the range of [0,1] and j index in the range of [0,4].

4.2. Experimental settings

We followed the inter-patient paradigm and divided the MIT-BIH dataset into training set DS1 and testing set DS2 according to the dividing method in the literature [35] and normalized the data to [0,1]. The records that are contained in DS1 and DS2 have been described in section 3.1, and the number of heartbeats included in each class is shown in Table 3.

Since the MIT-BIH-AR database is highly imbalanced, we applied the synthetic minority oversampling technique (SMOTE) to obtain the same volume of S, V and F class heartbeats as N class heartbeats. To minimize the impact of synthetic data on the classification model, we only randomly took 20,000 heartbeats from all N class heartbeats in DS1 and constructed oversampled datasets with all other types of heartbeats in DS1. In the following, DS1 refers to the oversampled training dataset. In our work, the parameters that need to be set in each experiment are the kernel function parameters, linear combination weights and the number of features of the base learners in MEKLM-RF. The setting of parameter values is shown in Table 4. As described in section 3.3, the total number of features in our study is 104. Each base learner in MKELM-RF randomly selects a number of $\lfloor\sqrt{104}\rfloor$ features among the total features, that is, 10 features are selected. In addition, the kernel function parameters involved in each base learner are randomly selected within the corresponding ranges given in Table 4.

4.3. Experiment 1: Comparison of the proposed KELM-RF vs. random forest and KELM

The goal of this experiment is to evaluate whether the classification

Table 3
The number of different classes of heartbeats in DS1 and DS2.

Heartbeat type	Training set DS1	Testing set DS2
N	45824	44033
S	976	2050
V	3788	3220
F	414	388
total	51002	49691

Table 4
Parameter settings of the base learners in MKELM-RF.

Parameter name	Parameter type	Value
Kernel function parameter	RBF kernelδ	$(0, 40]$
	polynomial kernel$d$	2
	penalty coefficient C	$(0, 40]$
Linear combination coefficient	weight coefficient λ	$[0, 1]$
Base classifier	number of features	10

results of KELM-RF is superior to the traditional decision trees-based random forest (RF) and KELM. In addition, we further test whether the ensemble multiclass classifier KELM-RF-OVO based on the binary classifier KELM-RF outperforms the single multiclass classifier KELM-RF.

We trained the multiclass classifier KELM, traditional decision trees-based random forest (RF), KELM-based random forest KELM-RF and ensemble multiclass classifier KELM-RF-OVO based on the binary classifier KELM-RF. The KELM, KELM-RF, and KELM-RF-OVO all utilized the RBF kernel as the kernel function. The number of base classifiers in each random forest of KELM-RF and KELM-RF-OVO was 11. The algorithm proposed by Beriman [54] was used to build the RF classifier, and the optimal number of decision trees was determined by the grid search method, with Cohen's kappa coefficient (k) as the measurement. The search range of the grid search was [10, 150] with a step size of 11, and the number of decision trees was determined to be 101. The experimental results of the four types of models on the MIT-BIH DS2 testing set are shown in Table 5.

Clinically, the S class and V class are two classes of critically anomalous and acute heartbeats. Hence, the experimental results associated with V class and S class heartbeats are mainly used for comparative performance evaluations. As illustrated in Table 5, the KELM obtains the worst classification results among all other classifiers. Although the sensitivity for S class heartbeats achieves 1, the lowest positive predictive value is also obtained at the same time, which means that KELM classifies most of the other types of heartbeats into S class heartbeats, and the model has a serious over-fitting problem and a poor generalization ability.

The classification performance of RF is better than KELM. Although the sensitivity for N class and V class heartbeats increases with the number of decision trees, the sensitivity of S class heartbeats gradually decreases as the number of decision trees increase from 11 to 101, which indicates that most decision trees in random forests are biased, which cannot effectively distinguish between S class heartbeats and other classes of heartbeats. As a result, the classification result of S class heartbeats becomes worse as the number of decision trees increases.

When the number of base classifiers of KELM-RF is only 11, the KELM-RF outperforms both KELM and RF. This must be due to adding the random sampling and random selection of features to KELM, which can prevent KELM from being over-fitting and improve the generalization capability. Additionally, each parameter value of base classifiers is randomly set within a certain range, which increases the difference between the base learners. Finally, because the learning speed of KELM is extremely fast, KELM-RF is well suited for the ECG heartbeat classification task.

It can also be observed from the experimental results that although KELM-RF has better classification results than those of KELM and RF, the classification performance on some types of heartbeats is still not ideal, and a far lower score than KELM-RF-OVO is obtained. In ECG heartbeat classification, the individual multiclass classifier is prone to be biased and sometimes produces extreme results. It is easy for the classification result of some class to appear particularly poor and for the classification effect to appear unstable. Additionally, in the case of a large sample volume, it takes a much longer time to learn. An ensemble of multiple binary classifiers to form a multiclass classifier can solve above the problems well.

4.4. Experiment 2: Comparison of single kernel-based KELM-RF vs. multi-kernel-based MKELM-RF

The goal of this experiment is to evaluate whether the linear combination of the RBF kernel and polynomial kernel based classifier MKELM-RF-OVO is superior to the RBF kernel-based and poly kernel-based KELM-RF-OVO. We trained three models in turn, using the RBF kernel, polynomial kernel, and the linear combination kernel of an RBF kernel and a polynomial kernel as the kernel function of the base classifiers. The number of base classifiers of KELM-RF is 11. The classification results are described in Table 6.

As seen from Table 6, the classification result of the polynomial kernel-based classifier is better than the RBF kernel-based classifier. The former increases the sensitivity for S class heartbeats by 22% and the positive predictive value by 24%, reaching 97.9% and 83.9%, respectively. The sensitivity for V class heartbeats increases by approximately 23% to 96.1%. Therefore, for the ECG arrhythmia classification, the polynomial kernel function, who is good at capturing global features, is more effective than the local kernel function RBF.

However, we can also find that although the sensitivity for N class heartbeats and the positive predictive value for S class heartbeats of the polynomial kernel-based classifier are higher than that of the mixed kernel-based classifier, the latter has higher sensitivities for S class and V class heartbeats than the former, especially regarding the sensitivity for N class heartbeats, which increases by 7%. The decrease in the positive predictive value for S class heartbeats is mainly because the misclassified normal heartbeats are classified into S class heartbeats. Clinically, more attention is paid to the detection of supraventricular ectopic beats and ventricular. Therefore, we choose the linear combination kernel of the RBF kernel and polynomial kernel as the kernel function.

4.5. Experiment 3: The effect of the number of base classifiers on the classification results evaluation

The purpose of this experiment is to evaluate the effect of the number of base classifiers in MKELM-RF-OVO on the classification results. We increased the number of base classifiers in KELM-RF gradually. Table 7 shows the experimental results.

As shown in Table 7, when the number of base classifiers in MKELM-

Table 5
Comparison of the classification performance on MIT-BIH (DS2) among the different methods. RF-11, RF-41, and RF-101 denote random forests with 11, 41, 101 decision trees, respectively.

	N		S		V		Average		ACC	k
Method	SEN_N	PPV_N	SEN_S	PPV_S	SEN_V	PPV_V	SEN	PPV		
KELM	0.0004	1	1	0.041	0	0	0.25	0.26	0.042	0.0005
RF-11	0.402	0.947	0.370	0.039	0.883	0.272	0.414	0.315	0.429	0.096
RF-41	0.687	0.954	0.332	0.071	0.906	0.374	0.481	0.350	0.681	0.231
RF-101	0.749	0.951	0.106	0.102	**0.949**	0.251	0.451	0.326	0.730	0.257
KELM-RF	0.998	0.961	0.529	0.471	0.506	1	0.5753	0.858	0.941	0.718
KELM-RF-OAO	**0.999**	**0.987**	**0.755**	**0.595**	0.663	**1**	**0.679**	**0.896**	**0.962**	**0.808**

Table 6
Comparison of the classification results on MIT-BIH-AR (DS2) among the RBF kernel-based classifier, the polynomial kernel-based classifier and the mixed kernel-based classifier.

	N		S		V		Average		*ACC*	*k*
Kernel	SEN_N	PPV_N	SEN_S	PPV_S	SEN_V	PPV_V	*SEN*	*PPV*		
RBF	**0.999**	0.987	0.755	0.595	0.663	1	0.679	0.896	0.962	0.808
polynomial	**0.999**	0.999	0.979	**0.839**	0.896	1	**0.753**	**0.960**	**0.986**	**0.929**
mixed	0.945	**0.999**	**0.999**	0.436	**0.961**	**1**	**0.726**	0.609	0.941	0.906

Table 7
Classification performance on MIT-BIH (DS2) of the three variants of our approach. Three variants used the numbers of 11, 21 and 27 base classifiers, respectively. BaseClfNum denotes the base classifiers' number of MKELM-RF in MKELM-RF-OVO.

	N		S		V		Average		*ACC*	*k*
BaseClfNum	SEN_N	PPV_V	SEN_S	PPV_S	SEN_V	PPV_V	*SEN*	*PPV*		
11	0.945	0.999	0.999	0.436	**0.961**	**1**	0.726	0.609	0.941	0.906
21	0.987	0.999	**1**	0.712	0.942	**1**	**0.762**	0.928	0.978	0.897
27	**0.991**	**0.999**	**1**	**0.757**	0.944	**1**	0.752	**0.939**	**0.981**	**0.909**

RF increases from 11 to 21, the sensitivity of the N class heartbeats increases by approximately 4%, and the positive predictive value for S class heartbeats increased by 28%. However, the sensitivity for V class heartbeats decreases by 2%, and the *k* value also decreases slightly. Then, increasing the number of base classifiers to 27, the sensitivity for N class heartbeats and the positive predictive value for S class heartbeats increase by approximately 4%, and the *k* value also attains the highest score. However, the sensitivity for F class heartbeats decreases by 6%. Therefore, the number of base classifiers is no longer increased, that is, when the number of base classifiers in MKELM-RF is 27, the best classification result of heartbeat classification is obtained. Table 8 presents the confusion matrix of the optimal classification result.

4.6. Experiment 4: Feature performance evaluation

The features of the different categories are used as the input of the classification model to evaluate their performances individually. Table 9 shows the results obtained on the DS2 testing set for models with different categories of features as the input. Fig. 5 shows the confusion matrix of the models using a single class of features as the input. It can be seen from Table 9 that the RR feature has the best recognition effect on S class heartbeats, with a sensitivity of 99%, but its positive predictive value is low. With reference to Fig. 5(a), the reason that the *PPV* of the S class heartbeat is low is that a large number of N class heartbeats are misclassified into the S class. The Morph feature has a better ability to distinguish between N and S class heartbeats than do the RR features, so that the *PPV* values of the two types of heartbeats are higher than the RR. From Fig. 5(a) and (b), we can see that the RR feature and Morph feature have poor classification effects on the V class and F class heartbeats, and 35% of the V class heartbeats are incorrectly classified into N class and S class heartbeats. The LBP feature has the best description effect on V class heartbeats, but it lacks the ability to distinguish between N class and S class heartbeats. The HOS feature has the best effect on the classification of F. The Wavelet feature has the best classification results for three types of heartbeats except the V class heartbeats. In addition, from the evaluation index *j*index, which focuses on measuring the classification effects of S class and V class abnormal heartbeats, the wavelet feature performs best.

Table 8
Confusion matrix over MIT-BIH (DS2) of our best configuration: KELM-RF-OVO using the mixed kernel as the kernel function of ELM and the number of base classifiers in KELM-RF is 27.

	Algorithm				
	N	S	V	F	Total
Reference					
N	43622	411	0	0	44033
S	0	2050	0	0	2050
V	0	181	3039	0	3220
F	11	67	282	28	388
Total	43633	2709	3321	28	49691

We compared single class features with the random feature selection, that is, we randomly selected 10 features (i.e., the square root value of the total number of features) as the input of the base classifiers. The classification results show that the method of random feature selection outperforms any other types of features in the classification of N, S and V class heartbeats, and its average classification accuracy, sensitivity and *PPV* value are also highest. However, random feature selection has low classification sensitivity in F class heartbeats. There are two possible reasons. One is that the volume of F is small, accounting for only 1% of the total heartbeats. Although the SMOTE oversampling method synthesizes the same number of F class heartbeats as other types of heartbeats, the synthesis algorithm also brings noise to a certain extent. Second, randomly feature selection may cause the features with stronger abilities to discriminate S and V class heartbeats to have a greater impact on the model, and the contribution to the features with strong description abilities of F are weakened. The HOS and Wavelet features have better classification effects on F class heartbeats, obtaining a sensitivity of over 86%. The other three classes of features have weak classification ability for F class heartbeats, for example, the RR feature attains only a 0.5% sensitivity. After using the random feature selection method, the performance of the obtained feature combination is worse than that of the optimal classes of features and is better than that of the weak ones.

The F class is a fusion of the V class and the N class. In view of this, the AAMI standard specifies that when the classifier classifies the F class heartbeat as a V class heartbeat, no punishment or reward is given to the classifier. Fig. 5(f) shows that the classifier classifies approximately 80% of the F class heartbeats into V class heartbeats instead of mistakenly dividing them into S class heartbeats. The random feature selection method can lose a part of the classification performance of F class heartbeats and obtain better classification performance on the S and V class heartbeats, which need more attention, thus achieving a meaningful compromise.

This study uses the same feature set as in reference [14]. We compare the classification results of this method on single-class features with those in [14]. The results are shown in Fig. 6. In reference [14], the HOS has the best classification performance for S and V class heartbeats, followed by the Morph. The RR is the best descriptor considering the four classes with an emphasis on the discrimination of the S and V classes. The Wavelet presents a low score for *j*index. Conversely, the

Table 9
Results of the classifiers trained with the different features over MIT-BIH-AR(DS2). The ideal features per measurement are in bold. Random represents random feature selection.

	N		S		V		F		Average		*ACC*	
Features	SEN_N	PPV_N	SEN_S	PPV_S	SEN_V	PPV_V	SEN_F	PPV_F	*SEN*	*PPV*		*j*index
RR	0.975	0.981	**0.993**	0.569	0.614	1	0.005	1	0.647	0.887	0.945	3.177
Morph	0.995	0.985	**0.839**	0.640	0.655	1	0.041	1	0.633	0.906	0.959	3.135
LBP	0.666	0.974	0.635	0.081	**0.987**	**1**	0.790	1	0.620	0.763	0.682	2.704
HOS	**0.999**	0.968	0.418	0.336	0.409	1	**0.861**	**1**	0.672	0.826	0.936	2.165
Wavelet	0.991	**0.996**	**0.973**	**0.706**	0.832	1	**0.878**	**1**	**0.919**	**0.925**	**0.979**	**3.512**
Random	0.991	**0.999**	**1**	**0.757**	0.944	1	0.072	1	0.752	**0.939**	**0.981**	**3.701**

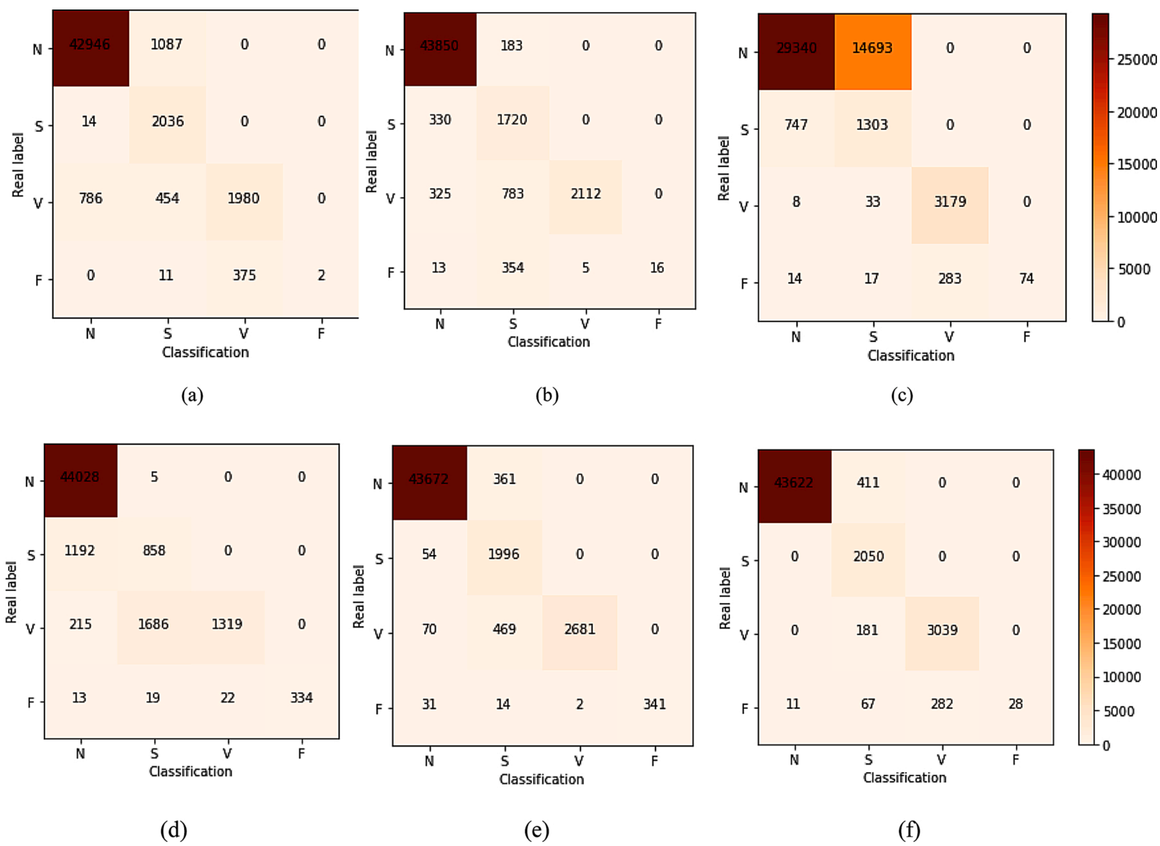

Fig. 5. Confusion Matrix of the classifiers trained with the different types of features over MIT-BIH-AR(DS2). (a) RR-feature as input. (b) Morph-feature as input. (c) LBP-feature as input. (d) HOS-feature as input. (e) Wavelet-feature as input. (f) Random feature selection, which randomly selects 10 features from all types of features.

Wavelet in our method has the optimal discrimination ability for all classes of heartbeats and obtained the highest *j*index value. Compared with the literature [14], our method has greatly improved the classification scores on the same single class feature. The highest average sensitivity, *PPV*, and accuracy achieved by our classifier are 18%, 47%, and 15% higher, respectively, than those in the literature [14]. The highest *j*index value of our method is 1.3 higher than that in the literature [14]. These finding show that our classification method can make full use of various classes of features for effective classification. In addition, it can also obtain the time-frequency description of the ECG, that is, the wavelet feature has better ability to distinguish between various types of heartbeats than do the other features.

4.7. Experiment 5: Comparison with the state-of-the-art methods

The goal of this experiment is to compare the best-configured proposed method with other classification algorithms, which also utilized the public MIT-BIH-AR database with the same inter-patient diagram. Table 10 contains the comparison results.

The results in Table 10 show that the classification results obtained by our proposed method are superior to those of the state-of-the-art methods in overall accuracy, positive predictive value and k value. In particular, the k value exceeds the best of the other methods [14] by 15%, reaching 0.909, which indicates that the classification results of our proposed method are almost identical to the reference values. From the classification results of each subclass, our method achieves the highest sensitivity for N class heartbeats among the methods listed in Table 10. In addition, the sensitivity and the positive predictive value for S class heartbeats are 1 and 0.757, respectively. The classification results of our method are superior to those of other methods.

Considering the high class imbalance of the heartbeat classification tasks, *j*index and *jk*index are also used to evaluate the proposed method and the other methods. Table 11 includes the comparison of our proposed method and the other methods in *j*index and *jk*index.

The experimental results in Table 11 show that our method achieves more than 15% and 14% improvements regarding Cohen's kappa and

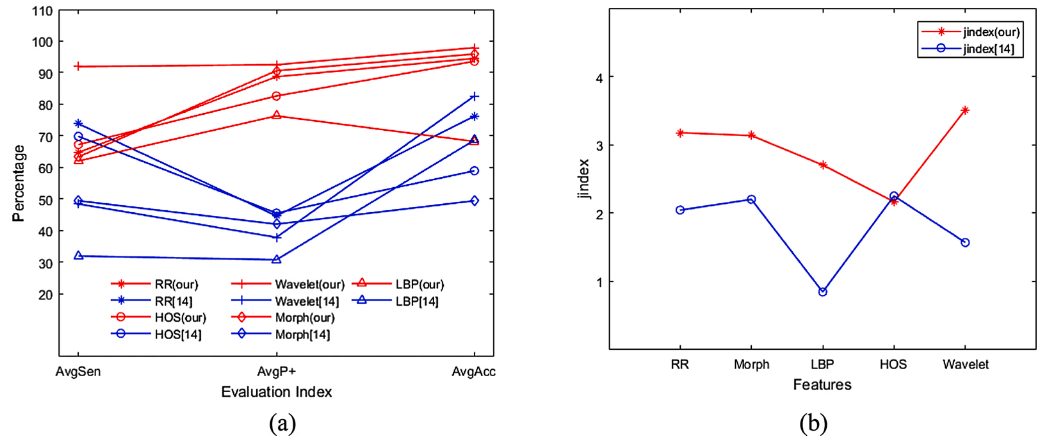

Fig. 6. Using single-category features as the input, the comparison between the classification performance of our method (Our) and the literature [14]. (a) Comparison of the classification results on average sensitivity (AvgSen), average positive predictive value (AvgP +), and average accuracy (AvgAcc). (b) Comparison of the classification results on the index *jindex*, which is a classification evaluation index of S class heartbeats and V class heartbeats.

Table 10
Classification performance on MIT-BIH (DS2) comparing our best setup against the state-of-the-art methods.

	N		S		V		F		Average		*ACC*	*k*
Method	SEN_N	PPV_N	SEN_S	PPV_S	SEN_V	PPV_V	SEN_F	PPV_F	*SEN*	*PPV*		
Our Method	**0.991**	**0.999**	**1**	**0.757**	0.944	**1**	0.072	**1**	0.752	**0.939**	**0.981**	**0.909**
Mondéjar et al.[14] 2019	0.959	0.982	0.781	0.497	**0.947**	0.939	0.124	0.236	0.703	0.664	0.945	0.755
Shi et al.[55] 2019	0.950	0.851	0.818	0.546	0.881	0.760	0.279	0.838	0.796	0.889	0.745	0.681
Li et al.[56] 2016	0.947	0.997	0.20	0.002	0.942	0.898	0.500	0.005	0.647	0.476	0.946	–
Mar et al. [57] 2011	0.896	0.991	0.832	0.335	0.868	0.759	0.611	0.166	0.802	0.564	0.899	0.599
Chazal et al. [35] 2004	0.871	0.992	0.760	0.385	0.803	0.866	0.894	0.086	**0.832**	0.570	0.862	0.532

Table 11
Comparison of classification results *j*index and *jk*index on MIT-BIH (DS2) between our best configurations and state-of-the-art methods.

Method	*k*	*j* index	*jk* index
Our method	**0.909**	**3.701**	**0.917**
Mondéjar et al. [14] 2019	0.755	3.165	0.773
Shi et al. [55] 2019	0.681	3.005	0.716
Li et al. [56] 2016	–	2.042	–
Mar et al. [57] 2011	0.599	2.794	0.649
Chazal et al. [35] 2004	0.532	2.814	0.618

the *jk*index in comparison with the reference [14], which obtains the second highest score. Furthermore, our method also attains the highest *j*index score. Therefore, our method has a good classification effect on arrhythmia, especially for the two most important abnormal classes, S and V. As far as we know, the experimental results of our proposed method are better than those of the state-of-the-art methods.

5. Conclusion

A novel ECG arrhythmia classification approach, i.e., an ensemble of kernel extreme learning machine based random forest classifiers, was proposed in this paper. To facilitate reproduction and comparison of the experiment with other methods, all the experiments were performed on the public MIT-BIH-AR database under the inter-patient paradigm. In the feature extraction stage, features commonly used in the ECG classification task, such as the RR interval feature, morphological features, wavelet transform coefficients, higher order statistical features and LBP features, were utilized. In the classification stage, a random forest classifier KELM-RF based on KELM according to random sampling and feature random selection was proposed. Second, a mixed kernel, which is the linear combination of the RBF kernel and the polynomial kernel, was employed to produce the **m**ixed **k**ernel-based **e**xtreme **l**earning **m**achine MKELM. Finally, an ensemble multiclass classifier was built with the OVO strategy to combine the MKELM-RF binary classifiers. The obtained experimental results show that (1) for ECG classification tasks, the KELM-RF outperforms KELM and the traditional decision trees-based RF, (2) the MKELM is better than the single RBF kernel or polynomial kernel-based KELM, so that the MKELM obtains higher classification scores for S and V classes, and (3) using the OVO strategy to combine binary classifiers for the multiclass classification task has better generalizability than training a single multiclass classifier.

In the future, a fusion method of multiple leads will be researched. Some mature and efficient parameter optimization methods may be utilized to optimize the model parameters.

CRediT authorship contribution statement

Ping Yang: Conceptualization, Methodology, Software, Writing - original draft, Writing - review & editing. **Dan Wang:** Conceptualization, Methodology, Writing - review & editing. **Wen-Bing Zhao:** Writing - original draft. **Li-Hua Fu:** Software. **Jin-Lian Du:** Formal analysis. **Hang Su:** Formal analysis.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant 61672505.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Cardiovascular Diseases Factsheet, WHO, 2017. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[2] ANSI/AAMI EC57, Testing and reporting performance results of cardiac rhythm and ST-segment measurement algorithms. AAMI –Association for the advancement of Medical Instrumentation. –CD-ROM AAMI standards and recommended practices, 1998.

[3] K.N.V.P.S. Rajesh, R. Dhuli, Classification of imbalanced ECG beats using re-sampling techniques and AdaBoost ensemble classifier, Biomed. Signal Process. Control 41 (2018) 242–254, https://doi.org/10.1016/j.bspc.2017.12.004.

[4] G.B. Moody, R.G. Mark, The impact of the MIT-BIH arrhythmia database, IEEE Eng. Med. Biol. Mag. 20 (3) (2001) 45–50, https://doi.org/10.1109/51.932724.

[5] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: Synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357, https://doi.org/10.1613/jair.953.

[6] M. Bahaz, R. Benzid, Efficient algorithm for baseline wander and powerline noise removal from ECG signals based on discrete Fourier series, Australas. Phys. Eng. Sci. Med. 41 (1) (2018) 143–160, https://doi.org/10.1007/s13246-018-0623-1.

[7] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, M. Adam, A. Gertych, R.S. Tan, A deep convolutional neural network model to classify heartbeats, Comput. Biol. Med. 89 (2017) 389–396, https://doi.org/10.1016/j.compbiomed.2017.08.022.

[8] Y. Sun, K.L. Chan, S.M. Krishnan, ECG signal conditioning by morphological filtering, Comput. Biol. Med. 32 (2002) 465–479, https://doi.org/10.1016/S0010-4825(02)00034-3.

[9] G. Garcia, G. Moreira, D. Menotti, E. Luz, Inter-Patient ECG Heartbeat Classification with Temporal VCG Optimized by PSO, Sci. Rep. 7 (1) (2017), 10543, https://doi.org/10.1038/s41598-017-09837-3.

[10] Z. Zhang, J. Dong, X. Luo, K.S. Choi, X. Wu, Heartbeat classification using disease-specific feature selection, Comput. Biol. Med. 46 (2014) 79–89, https://doi.org/10.1016/j.compbiomed.2013.11.019.

[11] B.M. Asl, S.K. Setarehdan, M. Mohebbi, Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal, Artif. Intell. Med. 44 (1) (2008) 51–64, https://doi.org/10.1016/j.artmed.2008.04.007.

[12] G. Sannino, G. De Pietro, A deep learning approach for ECG-based heartbeat classification for arrhythmia detection, Futur. Gener. Comput. Syst. 86 (2018) 446–455, https://doi.org/10.1016/j.future.2018.03.057.

[13] Y. Li, Y. Pang, J. Wang, X. Li, Patient-specific ECG classification by deeper CNN from generic to dedicated, Neurocomputing 314 (2018) 336–346, https://doi.org/10.1016/j.neucom.2018.06.068.

[14] V. Mondéjar-Guerra, J. Novo, J. Rouco, M.G. Penedo, M. Ortega, Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers, Biomed. Signal Process. Control 47 (2019) 41–48, https://doi.org/10.1016/j.bspc.2018.08.007.

[15] J. Pan, W. Tompkins, A real-time QRS detection algorithm, IEEE Trans. Biomed. Eng. 32 (3) (1985) 230–236, https://doi.org/10.1109/TBME.1985.325532.

[16] A. Ghaffari, M.R. Homaeinezhad, M. Akraminia, M. Atarod, M. Daevaeiha, A robust wavelet-based multi-lead electrocardiogram delineation algorithm, Med. Eng. Phys. 31 (10) (2009) 1219–1227, https://doi.org/10.1016/j.medengphy.2009.07.017.

[17] J.P. Martínez, R. Almeida, S. Olmos, A.P. Rocha, P. Laguna, A wavelet-based ECG delineator evaluation on standard databases, IEEE Trans. Biomed. Eng. 51 (4) (2004) 570–581, https://doi.org/10.1109/TBME.2003.821031.

[18] E.J. da S. Luz, W.R. Schwartz, G. Cámara-Chávez, D. Menotti, ECG-based heartbeat classification for arrhythmia detection: a survey, Comput. Methods Programs Biomed. 127 (Suppl. C) (2016) 144–164, https://doi.org/10.1016/j.cmpb.2015.12.008.

[19] K.I. Minami, H. Nakajima, T. Toyoshima, Real-time discrimination of ventricular tachyarrhythmia with fourier-transform neural network, IEEE Trans. Biomed. Eng. 46 (2) (1999) 179–185, https://doi.org/10.1109/10.740880.

[20] V.X. Afonso, W.J. Tompkins, T.Q. Nguyen, S. Luo, ECG beat detection using filter banks, IEEE Trans. Biomed. Eng. 46 (2) (1999) 192–202, https://doi.org/10.1109/10.740882.

[21] K.N.V.P.S. Rajesh, R. Dhuli, Classification of imbalanced ECG beats using re-sampling techniques and AdaBoost ensemble classifier, Biomed. Signal Process. Control. 41 (2018) 242–254, https://doi.org/10.1016/j.bspc.2017.12.004.

[22] L. Su, H.-T. Wu, Extract fetal ECG from single-lead abdominal ECG by de-shape short time Fourier transform and nonlocal median, Front. Appl. Math. Stat. 3 (2017) 2, https://doi.org/10.3389/fams.2017.00002.

[23] A. Gavrovska, V. Bogdanović, I. Reljin, B. Reljin, Automatic heart sound detection in pediatric patients without electrocardiogram reference via pseudo-affine Wigner-Ville distribution and Haar wavelet lifting, Comput. Methods Programs Biomed. 113 (2) (2014) 515–528, https://doi.org/10.1016/j.cmpb.2013.11.018.

[24] D. Xia, Q. Meng, Y. Chen, Z. Zhang, Classification of ventricular tachycardia and fibrillation based on the lempel-ziv complexity and EMD. Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Springer, 2014, pp. 322–329, https://doi.org/10.1007/978-3-319-09330-7_39.

[25] S. Pal, M. Mitra, Detection of ECG characteristic points using Multiresolution Wavelet Analysis based Selective Coefficient Method, Measurement 43 (2) (2010) 255–261, https://doi.org/10.1016/j.measurement.2009.10.004.

[26] S.Z. Mahmoodabadi, A. Ahmadian, M.D. Abolhasani, M. Eslami, J.H. Bidgoli, ECG feature extraction based on multiresolution wavelet transform, Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc., IEEE (2005) 3902–3905, https://doi.org/10.1109/iembs.2005.1615314.

[27] R. Singh, R. Mehta, N. Rajpal, Efficient wavelet families for ECG classification using neural classifiers, Procedia Comput. Sci., Elsevier B.V. (2018) 11–21, https://doi.org/10.1016/j.procs.2018.05.054.

[28] T. Tuncer, S. Dogan, P. Pławiak, U. Rajendra Acharya, Automated arrhythmia detection using novel hexadecimal local pattern and multilevel wavelet transform with ECG signals, Knowledge-Based Syst. 186 (2019) 104923, https://doi.org/10.1016/j.knosys.2019.104923.

[29] S. Osowski, T.H. Linh, ECG beat recognition using fuzzy hybrid neural network, IEEE Trans. Biomed. Eng. 48 (11) (2001) 1265–1271, https://doi.org/10.1109/10.959322.

[30] G. De Lannoy, D. François, J. Delbeke, M. Verleysen, Weighted SVMs and feature relevance assessment in supervised heart beat classification. Commun. Comput. Inf. Sci., Springer, 2011, pp. 212–223, https://doi.org/10.1007/978-3-642-18472-7_17.

[31] B. Singh, M. Singh, V.K. Banga, Sample Entropy based HRV: Effect of ECG Sampling Frequency, Biomed. Sci. Eng. 2 (3) (2014) 68–72, https://doi.org/10.12691/BSE-2-3-3.

[32] M. Llamedo, J.P. Martínez, Heartbeat classification using feature selection driven by database generalization criteria, IEEE Trans. Biomed. Eng. 58 (2011) 616–625, https://doi.org/10.1109/TBME.2010.2068048.

[33] H. Li, D. Yuan, X. Ma, D. Cui, L. Cao, Genetic algorithm for the optimization of features and neural networks in ECG signals classification, Sci. Rep. 7 (2017), 41011, https://doi.org/10.1038/srep41011.

[34] B.M. Asl, S.K. Setarehdan, M. Mohebbi, Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal, Artif. Intell. Med. 44 (1) (2008) 51–64, https://doi.org/10.1016/j.artmed.2008.04.007.

[35] P. De Chazal, M. O'Dwyer, R.B. Reilly, Automatic classification of heartbeats using ECG morphology and heartbeat interval features, IEEE Trans. Biomed. Eng. 51 (7) (2004) 1196–1206, https://doi.org/10.1109/TBME.2004.827359.

[36] M. Llamedo Soria, J.P. Martínez, Analysis of multidomain features for ECG classification, Comput. Cardiol. (2009) 561–564.

[37] R.J. Martis, U.R. Acharya, K.M. Mandana, A.K. Ray, C. Chakraborty, Application of principal component analysis to ECG signals for automated diagnosis of cardiac health, Expert Syst. Appl. 39 (14) (2012) 11792–11800, https://doi.org/10.1016/j.eswa.2012.04.072.

[38] H. Li, X. Feng, L. Cao, E. Li, H. Liang, X. Chen, A new ECG signal classification based on WPD and ApEn feature extraction, Circuits, Syst. Signal Process. 35 (1) (2016) 339–352, https://doi.org/10.1007/s00034-015-0068-7.

[39] E. Alickovic, A. Subasi, Effect of multiscale PCA de-noising in ECG beat classification for diagnosis of cardiovascular diseases, Circuits, Syst. Signal Process. 34 (2015) 513–533, https://doi.org/10.1007/s00034-014-9864-8.

[40] A. Sellami, H. Hwang, A robust deep convolutional neural network with batch-weighted loss for heartbeat classification, Expert Syst. Appl. 122 (15) (2019) 75–84, https://doi.org/10.1016/j.eswa.2018.12.037.

[41] Z. Cömert, A.F. Kocamaz, S. Güngör, Cardiotocography signals with artificial neural network and extreme learning machine, in: 2016 24th Signal Process. Commun. Appl. Conf. SIU 2016 - Proc, IEEE, 2016, pp. 1493–1496, https://doi.org/10.1109/SIU.2016.7496034.

[42] U. Mahdiyah, M.I. Irawan, E.M. Imah, Integrating data selection and extreme learning machine for imbalanced data, Procedia Comput. Sci. 59 (2015) 221–229, https://doi.org/10.1016/j.procs.2015.07.561.

[43] A. Diker, D. Avci, E. Avci, M. Gedikpinar, A new technique for ECG signal classification genetic algorithm Wavelet Kernel extreme learning machine, Optik (Stuttg) 180 (2019) 46–55, https://doi.org/10.1016/j.ijleo.2018.11.065.

[44] F.A. Elhaj, N. Salim, T. Ahmed, A.R. Harris, T.T. Swee, Hybrid classification of Bayesian and extreme learning machine for heartbeat classification of arrhythmia detection, in: 6th ICT Int. Student Proj. Conf. Elev. Community Through ICT, ICT-ISPC 2017, IEEE, 2017, pp. 1–4, https://doi.org/10.1109/ICT-ISPC.2017.8075320.

[45] S. Zhou, B. Tan, Electrocardiogram soft computing using hybrid deep learning CNN-ELM, Appl. Soft Comput. J. (2019) 105778, https://doi.org/10.1016/j.asoc.2019.105778.

[46] M.M.A. Rahhal, Y. Bazi, N. Alajlan, S. Malek, H. Al-Hichri, F. Melgani, M.A. A. Zuair, Classification of AAMI heartbeat classes with an interactive ELM ensemble learning approach, Biomed. Signal Process. Control. 19 (2015) 56–67, https://doi.org/10.1016/j.bspc.2015.03.010.

[47] T. Ojala, M. Pietikäinen, T. Mäenpää, Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns, Springer, Berlin, Heidelberg, 2000, pp. 404–420, https://doi.org/10.1007/3-540-45054-8_27. ISBN: 978-3-540-67685-0.

[48] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks, Neural Netw. 2 (2004) 985–990, https://doi.org/10.1109/ijcnn.2004.1380068.

[49] G.B. Huang, An insight into extreme learning machines: random neurons, random features and kernels, Cogn Comput. 6 (2014) 376–390, https://doi.org/10.1007/s12559-014-9255-2.

[50] J.C. Ferreira, V.A. Menegatto, Eigenvalues of integral operators defined by smooth positive definite kernels, Integr. Equations Oper. Theory. 64 (1) (2009) 61–81, https://doi.org/10.1007/s00020-009-1680-3.

[51] L.D. Wang, Research and Applications of SVM Based on the Mixed-kernel Function, M.S. Thesis, Dalian Maritime University, Dalian, China, 2016.

[52] Y. Shi, T. Falck, A. Daemen, L.C. Tranchevent, J.A.K. Suykens, B. De Moor, Y. Moreau, L2-norm multiple kernel learning and its application to biomedical data fusion, BMC Bioinformatics 11 (1) (2010) 309, https://doi.org/10.1186/1471-2105-11-309.

[53] J.P. Vert, K. Tsuda, B. Schölkopf, A primer on kernel methods, Kernel Methods Comput. Biol. 47 (2004) 35–70, https://doi.org/10.7551/mitpress/4057.003.0004.

[54] L. Breiman, Random Forests, Machine Learn. 45 (1) (2001) 5–32, https://doi.org/10.1023/A:10109334.

[55] H.T. Shi, H.R. Wang, F. Zhang, Y.X. Huang, L.Q. Zhao, C.L. liu, Inter-patient heartbeat classification based on region feature extraction and ensemble classifier, Biomed. Signal Process. Control 51 (2019) 97–105.

[56] T.Y. Li, M. Zhou, ECG classification using wavelet packet entropy and random forests, Entropy 18 (8) (2016) 285.

[57] T. Mar, S. Zaunseder, J.P. Martínez, M. Llamedo, R. Poll, Optimization of ECG classification by means of feature selection, IEEE Trans. Biomed. Eng. 58 (8) (2011) 2168–2177, https://doi.org/10.1109/TBME.2011.2113395.

[58] M.L. Soria, J.P. Martinez, An ECG classification model based on multilead wavelet transform features, IEEE, Piscataway, New Jersey, USA, 2007, pp. 105–108, https://doi.org/10.1109/cic.2007.4745432. ISBN：978-1-4244-2533-2534.