Biomedical Signal Processing and Control 47 (2019) 41–48

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers

V. Mondéjar-Guerra[a,b,*], J. Novo[a,b], J. Rouco[a,b], M.G. Penedo[a,b], M. Ortega[a,b]

[a] *Department of Computing, University of A Coruña, A Coruña, Spain*
[b] *CITIC-Research Center of Information and Communication Technologies, University of A Coruña, A Coruña, Spain*

ARTICLE INFO

Article history:
Received 6 April 2018
Received in revised form 12 June 2018
Accepted 8 August 2018

Keywords:
Electrocardiogram (ECG)
Heartbeat classification
Support vector machine (SVM)
Combining classifiers
Ensemble of classifiers

ABSTRACT

A method for the automatic classification of electrocardiograms (ECG) based on the combination of multiple Support Vector Machines (SVMs) is presented in this work. The method relies on the time intervals between consequent beats and their morphology for the ECG characterisation. Different descriptors based on wavelets, local binary patterns (LBP), higher order statistics (HOS) and several amplitude values were employed. Instead of concatenating all these features to feed a single SVM model, we propose to train specific SVM models for each type of feature. In order to obtain the final prediction, the decisions of the different models are combined with the product, sum, and majority rules. The designed methodology approaches are tested on the public MIT-BIH arrhythmia database, classifying four kinds of abnormal and normal beats. Our approach based on an ensemble of SVMs offered a satisfactory performance, improving the results when compared to a single SVM model using the same features. Additionally, our approach also showed better results in comparison with previous machine learning approaches of the state-of-the-art.

1. Introduction

Disturbances in the heart rate, popularly known as arrhythmias, may be life-threatening, requiring immediate care and often an intervention with defibrillator [1]. Nevertheless, most of arrhythmias are harmless; but even then, they may require therapy to prevent further severe problems [2]. Arrhythmias are often associated with other forms of heart disease. According to the World Health Organization (WHO), “Cardiovascular diseases (CVDs) are the number 1 cause of death globally: more people die annually from CVDs than from any other cause. An estimated 17.7 million people died from CVDs in 2015, representing 31% of all global deaths”. Electrocardiograms (ECG) are a noninvasive and inexpensive technique commonly employed by cardiologist in their clinical practice routine. They are frequently used to detect cardiac rhythm abnormalities, measuring the electrical activity of the heart over a period of time. For a routine analysis of the heart’s electrical activity, an ECG recorded from 12 separate leads is typically used. The 12-lead ECG consists of three bipolar limb leads (I, II, and III), the unipolar limb leads (AVR, AVL, and AVF), and six unipolar chest leads, also called precordial or V leads (V1, V2, V3, V4, V5 and V6). Each lead is a view of the electrical activity of the heart from a particular angle across the body. The record contains approximately 2.5 s of duration for each lead. Additionally, to accurately assess the cardiac rhythm, a prolonged recording from one lead is used to provide a rhythm strip of 10 s. Lead II is the most commonly used for the rhythm strip [3], since it usually gives a good view of the most important waves: P, Q, R, S and T (see Fig. 1). Each beat of the heart contains a series of deflections away from the baseline on the ECG, or waves, that reflect the time evolution of electrical activity in the heart. P-wave is a small defection caused by atrial depolarisation, Q, R, and S waves are usually considered as a single event known as the QRS-complex, which is the largest-amplitude portion of the ECG, being caused by ventral depolarisation. T wave is caused by ventral repolarisation. Finally, in some cases, an additional U wave may follow the T wave.

The different types of arrhythmias can be detected through the analysis of the changes that appeared on these waves. The development of a fully automatic system that is able to classify the ECG heartbeats has been a research topic of high interest throughout the last decades. Fig. 2 shows a diagram of a general automatic system for arrhythmia classification. First, the signals that were captured through the device are preprocessed. This step usually includes the baseline removal and the cleaning of high-frequency noise. Next, a

* Corresponding author at: Department of Computing, University of A Coruña, A Coruña, Spain.
E-mail addresses: v.mondejar@udc.es (V. Mondéjar-Guerra), jnovo@udc.es (J. Novo), jrouco@udc.es (J. Rouco), mgpenedo@udc.es (M.G. Penedo), mortega@udc.es (M. Ortega).

https://doi.org/10.1016/j.bspc.2018.08.007
1746-8094/© 2018 Published by Elsevier Ltd.

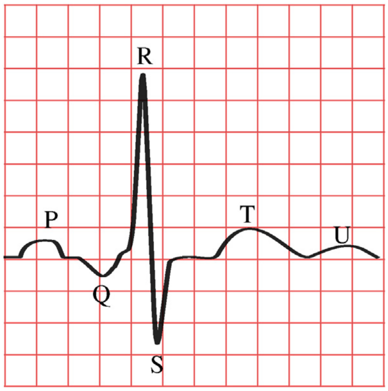

Fig. 1. Waves of a lead II ECG.

heartbeat segmentation algorithm is applied to split the signal at beat level. This is usually done detecting the QRS-complex. Then, several descriptors are applied to each beat in order to extract the features to feed a classifier, which finally determines the type of heartbeat. Many algorithms were proposed in the literature for the heartbeat segmentation [4–7], reaching up to near optimal results in well-known databases like MIT-BIH [8]. In this work, we focus on the two last steps, feature extraction and classification. Many features were proposed to describe the ECG heartbeats, highlighting the use of wavelets [9,10], higher order statistics (HOS) [11,12], and heartbeat intervals, popularly known as R-R intervals [13,14]. To built the classification model, numerous previous works reported the feasibility of machine learning algorithms for this task [15]; including methods such as Linear Discriminant (LD) [2], AdaBoost [16], Multilayer Perceptron (MLP) [9,17,18], Genetic Algorithm-Back Propagation Neural Network (GA-BPNN) [19], Convolutional Neural Networks (CNN) [20], and, especially, Support Vector Machine (SVM) [17,18,21–24].

An ensemble of classifiers combines the decisions of the individual classifiers that compose it, in order to improve the final prediction. There are many techniques in the literature to create an ensemble of classifiers [25]. Some methods train each classifier with a different subset of the training examples like Bagging [26], or AdaBoost [27]. Dietterich and Bakiri [28] deal with a problem that requires a large number of classes, partitioning the number of outputs in different sets, generating an ensemble of classifier. Other works train each classifier in a different subset of the input features. Duin and Tax [29] performed a large experimentation of this alternative and concluded that the combination of classifiers trained on different feature sets was very useful, especially when the single classifiers offered a good performance. Waske and Benediktsson [30] employed an ensemble of SVMs in a multi-source land cover classification problem using a balanced dataset. Their ensemble of SVMs, training each SVM with a different data source, significantly improved the results in comparison to a single SVM that was trained with the whole data sources.

The main goal of this work is to evaluate the benefits of using an ensemble of SVMs for the arrhythmia classification problem, i.e., combining several SVM models each one trained with a different feature. Several feature descriptors based on R-R intervals, wavelets, HOS, LBP, and several amplitude values were employed. Besides, the suitability of each single feature is also evaluated in this work. Our approach is similar to the work of Zhang et al. [22], which also used an ensemble of SVMs for the automatic arrhythmia classification. However, they extracted features from the leads II and V1 and posteriorly, they trained a separated model from the features of each lead. Finally, they combined the decisions of both models with the product rule. In our approach, a SVM model is created for each type of feature being all the features extracted from lead II. Additionally, an extensive experimentation was made evaluating all the possible combinations of the selected features. Finally, we tested several combination rules, including the commonly employed sum, product, and majority rules [31].

In the literature, we can distinguish two popular paradigms for evaluation of arrhythmia classification task, known as intra-patient and inter-patient. In the first paradigm, the whole database can be employed to generate and test the classification models without any restriction. This paradigm presents a main drawback regarding the generalisation of the classifier. Since the model can learn the particularities of the patients during the training, the score achieved in the evaluation step may not be highly reliable. Ideally, an automatic arrhythmia classifier must give an accuracy diagnosis for any patient, even if the system does not contain any previous information about it. Therefore, a method with high generalisation is desirable, since a trained database with records from all the possible patients would be unviable. In order to employ a more realistic scenario, Chazal et al. [2] proposed the inter-patient paradigm. They performed a division of the MIT-BIH database records in two different sets: one for training and other testing. These sets were carefully designed avoiding the inclusion of any record from the same patient in both sets. We followed the inter-patient paradigm to evaluate our approach.

In the next section, the used database, the selected features and the proposed approach for the ECG classification are detailed. The employed performance measurements, the experiments and the obtained results are explained on Section 3. Finally, Section 4 details the conclusions extracted from this work.

2. Material and methods

The well-known Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) arrhythmia database [8], from Physionet [32], was employed to train and test our classification models, allowing in turn the comparison of our results with those from the state-of-the-art methods.

2.1. MIT-BIH arrhythmia database

This database contains 48 ECG records of about 30 min, sampled at 360 Hz with 11-bit resolution from 47 different patients. Each record comprises two signals, the first one is, for all the records, the modified-lead II (MLII), whereas the second one corresponds

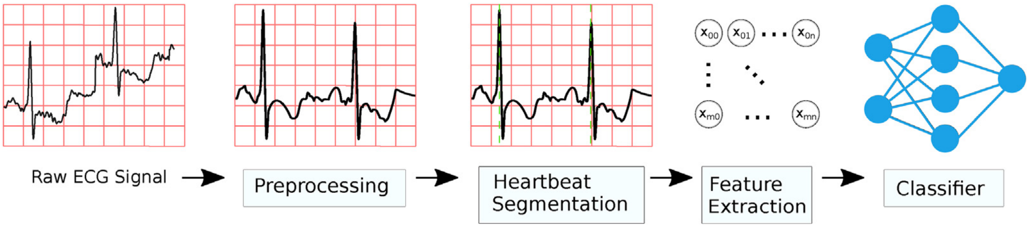

Fig. 2. Diagram of general steps of a full automatic arrhythmia classification system.

Table 1
MIT-BIH labelling and the standard AAMI classes.

AAMI		MIT-BIH
Normal	(N)	N, L, R
Supraventricular ectopic beat	(SVEB)	e, j, A, a, J, S
Ventricular ectopic beat	(VEB)	V, E
Fusion	(F)	F
Unknown beat	(Q)	/, f, Q

to V1, V2, V4, or V5, depending on the record. Therefore, only the MLII is provided by all the records. The database contains approximately 110,000 beats, all of them were independently annotated by two or more expert cardiologists and the disagreements were resolved. Following the Association for the Advancement of Medical Instrumentation AAMI recommended practice [33], the MIT-BIH heartbeat types are grouped into five heartbeat classes as shown in Table 1 . As recommended by the AAMI, the records with paced beats were not considered, namely 102, 104, 107, and 217. The database is highly imbalanced, as near a 90% of the beats belong to the class N whereas the remaining 3%, 6%, and 1% of the beats belong to classes SVEB, VEB, and F. We adopted the decision of ignoring the Q AAMI class like other authors [22,9], since it is practically non-existent. Only 15 samples belong to class F. In order to make a fair comparison between our results and those from other previous works, we used the popular inter-patient scheme division proposed by Chazal et al. [2], which divided the database in two datasets. Each dataset contains data from 22 records with a similar proportion of beat types:

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

The first dataset was employed for training whereas the second one was used to evaluate the performance of the model. None of the patients was repeated in both datasets.

In this work, only the lead II was considered to describe the morphology of the signal. This decision was motivated by the following facts: it is the only lead that is present for all the records from the MIT-BIH arrhythmia database; it is also the most commonly used lead by the experts to analyse the ECG signals; and finally, Chazal proved that using only one lead is sufficient for the arrhythmia classification task [34].

2.2. Signal preprocessing

Before computing the features from the ECG signals, a preprocessing step was applied. Most of the previous works of the literature [2,9,22] usually performed the baseline removal (see Fig. 2) followed by a high-frequency noise filtering at this step. In this case, we have just performed the baseline removal. To compute the baseline of the signal, two consecutive median filters of 200-ms and 600-ms were applied. Then, this baseline was subtracted from the original signal, producing the baseline corrected ECG signal. We made the decision of not performing any high-frequency noise filtering in order to preserve the signal as raw as possible for the feature extraction step.

2.3. Selected features

In practice, a QRS detection algorithm like the proposed by Pan and Tompkins [4] would be required in order to segment the full signal into beats. However, this work is focused on the classification step, therefore the QRS annotations included in MIT-BIH were employed.

For each beat, a window of size 180 was centred around its R-peak and, then, all the features were computed inside that region. Fig. 3(a) shows the mean values of all the beats from the MIT-BIH group by the four AAMI classes, whereas Fig. 3(b–f) show the mean values obtained over each feature descriptor. The following features were employed since they showed a good performance on similar previous works.

2.3.1. Wavelets

The wavelet transforms present the capability of allowing information extraction from both frequency and time domains, which make them suitable for the ECG description. The use of wavelet transforms were successfully proved by different authors on ECG classification [9,10]. Here, we used the Daubechies wavelet function (db1) with 3 levels of decomposition, making a 23-dimensional descriptor.

2.3.2. HOS

The use of higher order statistics (HOS), i.e., cumulants of the second, third, and fourth order were proposed as a better alternative for the morphological ECG description in [11,12]. Here, a 10-dimensional feature was created dividing each beat into 5 intervals, computing the kurtosis and skewness value over each one.

2.3.3. 1D-LBP

A 1D variant of the well-known descriptor, the 2D-Local binary patterns (LBP), was previously proposed for feature extraction of raw Electroencephalogram (EEG) signals in [35]. The 1D version maintains the idea of the original 2D version. For each data point in a beat, a binary code is produced by the comparison of its value with the value of their neighbours. Then, a histogram that contains the frequency of each binary pattern is built. Here, we used the 8 neighbour uniform LBP code, making a 59-dimensional descriptor.

2.3.4. Our morphological descriptor

We proposed a morphological descriptor that relies on several amplitude values from the beats. Instead of directly employing several amplitude values like other previous works [2,22], our descriptor relies on the Euclidean distance (sample, amplitude) between the R-peak and four points of the beat. The selection of this points depends on the amplitude values over the following intervals:

- *max(beat[0, 40]).*
- *min(beat[75, 85]).*
- *min(beat[95, 105]).*
- *max(beat[150, 180]).*

where *beat* is the 180-dimensional array, centered on the R-peak that contains the amplitude values.

2.3.5. R-R Intervals

A descriptor based on these intervals is certainly the most employed feature for the classification of ECGs in the literature [15]. Besides the morphological features, R-R intervals computed from the time between consequent beats were employed. The next intervals were extracted:

- Pre-RR: indicating the distance between the actual heartbeat and the previous one.
- Post-RR: indicating the distance between the actual heartbeat and the next one.
- Local-RR: containing the average of the 10 previous Pre-RR values.

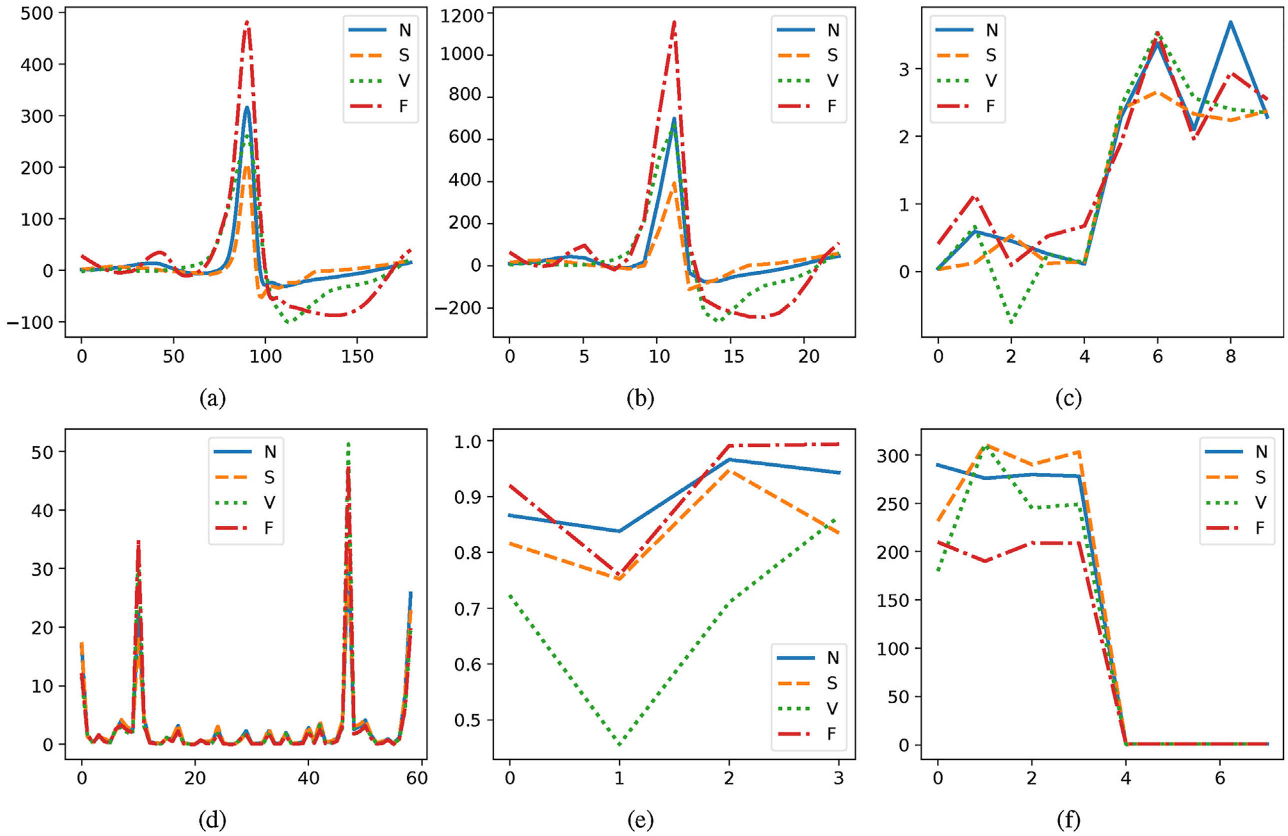

Fig. 3. Average beats from the MIT-BIH database grouped by the four AAMI class (N, SVEB, VEB, and F) (a) 180 window centered on R-peak from the raw lead II signal. (b) Wavelets of family 'db1' level 3 decomposition. (c) 5 HOS intervals: skewness and kurtosis. (d) Histogram U-LBP 1D of 8 bits. (e) Our morphological descriptor. (f) 4 R-R intervals followed by its 4 normalised values. (Best seen in color).

- Global-RR: containing the average of the Pre-RR values produced in the last 20 min.

In addition, the normalised version of these intervals [13], i.e., dividing them by its mean value within the same ECG record, were also employed, making a total of eight R-R features.

2.4. Classifier models

Due to their good performance showed on previous ECG classification works [21,22], we employed SVMs as the classifiers for all our experiments.

2.4.1. Support vector machine

SVM [36] are maximum margin classifiers that map input vectors to a higher dimensional space where a maximal separating hyperplane is constructed to differentiate two different classes. The main drawback of SVMs is their limitation to binary classification problems. In the literature, there are two common alternatives to solve that, namely one-against-all (OAA) and one-against-one (OAO). Being N the number of classes, in the first alternative N SVM models are constructed, i.e., one per class; whereas in the second alternative a SVM is constructed between each pair of classes, resulting in $N(N-1)/2$ models. Finally, a voting system is required for both alternatives in order to get a final decision. In this work, the OAO approach was used, since it is more suitable to work with imbalanced data and requires less time for training than OAA when the number of samples is significantly large [37].

Table 2
Features employed for ECG description.

Feature name	Description	Size
R-R	R-R intervals + normalized R-R intervals	8
Wavelets	Family *db1* with 3 level of decomposition	23
HOS	Skewness and Kurtosis with 5 intervals	10
LBP	Uniform-LBP 1D 8 bits	59
Our Morph.	Our morphological descriptor	4

Given N classes and L models, the final decision of a new observation $\mathbf{x}$ is computed using the pairwise a posteriori probability $P(y_m|f_l(\mathbf{x}))$ that follows a sigmoid function:

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

where $M=2$, since each model of the OAO is in fact a binary classifier, $y_1=+1$ and $y_2=-1$ denotes positive and negative classes, respectively, and $f_l(\mathbf{x})$ measures the decision value given by the l SVM. In our problem $N=4$, and, consequently, $L=6$ due to the OAO strategy. The probabilities of the different models $P(y_m|f_l(\mathbf{x}))$ are accumulated over their associated class n in δ_n and then, the majority voting rule is applied to assign the final decision:

$$\arg\max_n \delta_n. \qquad (2)$$

2.4.2. Combining multiple SVMs

In this work, an independent OAO SVM model is trained for each type of employed feature (see Table 2). Let δ_{tn} be the accumulated probabilities of the t OAO, considering all the features $T=5$. To combine the accumulated probabilities of the different OAO models, some of the most commonly used combination rules were tested:

Table 3
Description of confusion matrix from AAMI classes: N: normal, S: Supraventricular, V: Ventricular, and F: fusion.

	Algorithm				
	n	s	v	f	Sum
Reference					
N	*Nn*	*Ns*	*Nv*	*Nf*	$\sum N$
S	*Sn*	*Ss*	*Sv*	*Sf*	$\sum S$
V	*Vn*	*Vs*	*Vv*	*Vf*	$\sum V$
F	*Fn*	*Fs*	*Fv*	*Ff*	$\sum F$
Sum	$\sum n$	$\sum s$	$\sum v$	$\sum f$	$\sum$

- Product rule: It is a severe rule, since if just one model assigns a close to zero probability for one class, the final output for this class will also be close to zero:

$$\prod_{t=1}^{T} \delta_{tn}. \tag{3}$$

- Sum rule: Opposite to the product rule, the sum rule presents a more relaxed behaviour:

$$\sum_{t=1}^{T} \delta_{tn}. \tag{4}$$

- Majority rule: It adds a vote to each class depending on their rank position. This rule does not consider the differences at probability level between the outputs, it only consider the rank order:

$$\sum_{t=1}^{T} \nu_{tn}, \tag{5}$$

where ν_{tn} contains a vote inversely proportional to the rank position of class n in δ_t.

Finally, once the accumulated probabilities are combined, the final decision is selected using the majority voting rule (Eq. (2)).

3. Experimentation

The data were standardized (z-score), i.e., subtracting the mean and dividing the standard deviation of the training data. Since the MIT-BIH database is highly imbalanced, several weights equal to the ratio between the two classes that compose each *l* model were employed to compensate these differences. The Radial Basis Function (*RBF*) kernel was fixed for all the experimentation process. The same values for *C* and γ were selected for all the *L* models. The *gamma* value was fixed to 1/*size*(*features*). To adjust the penalty parameter *C*, a 10-fold cross-validation strategy was performed over the training dataset (DS1), varying *C* over the grid {0.001, 0.01, 0.1, 1, 10, 100}. Once the best parameters were selected, the models were trained again over the full training set (DS1) and tested over the evaluation set (DS2) following the inter-patient division [2].

3.1. Performance measurements

Following the AAMI specifications, the performance measurements were computed from the confusion matrix (Table 3). They include some particularities in the measurements computation [2], e.g., they do not reward or penalize a classifier for the classification of ventricular fusion (F) as (VEBs), *Fv*. The confusion matrix provides a complete description of any classification results. However, due to imbalance of the MIT-BIH database, the overall accuracy or the mean accuracy do not represent well how *good* a classifier is. Suppose we have a classifier that only assigns the output of normal class (N) for all the new incoming data. This classifier would achieve a value of the overall accuracy higher than 89%. In the other hand, the mean accuracy would give the same importance to the majority and minority classes. Therefore, these performance measurement does not seem appropriate to represent the quality of the classifiers on this database. To overcome this problem, Mar et al. proposed a new index, which they named *j*κ index [9], as a combination of two values: the *j* index [38] and the Cohen's Kappa (κ) index [39]:

$$j\kappa\ index = w_1 \kappa + w_2\, j\ index, \tag{6}$$

where $w_1 = 1/2$, and $w_2 = 1/8$ since κ takes values in the [0,1] range and *j index* in the [0,4] range. The *j* index evaluates the discrimination of the most important arrhythmias (SVEB, VEB, according to the AAMI standard [9]):

$$j\,index = Se_S + Se_V + P^+_S + P^+_V, \tag{7}$$

being *Se* and P^+ the sensitivity and positive predictive value of each class. Finally, the Cohen's Kappa (κ) is a measure of agreement that globally evaluates the confusion matrix. It was reported as a performance measurement more robust than the overall accuracy or the mean accuracy on imbalance datasets [40]:

$$\begin{aligned}
\kappa &= \frac{P_o - P_e}{1 - P_e},\\
P_o &= \frac{Nn + Ss + Vv + Ff}{\sum},\\
P_e &= \frac{\sum N \sum n + \sum S \sum s + \sum V \sum v + \sum F \sum f}{\sum^2},
\end{aligned} \tag{8}$$

where P_o is the observed probability, being equal to the overall accuracy, and P_e corresponds with the chance agreement. Note that the term P_e takes into consideration the number of samples of each class. Assuming equally distributed data over the four classes, P_e will be a constant, and hence κ and the overall accuracy will be linearly dependent.

We used the *j*κ *index* for the evaluation, since this index takes into account, in a single score, the misclassification and the imbalance that is present between all the considered classes, thanks to the included κ index, and at the same time emphasises the discrimination of the most important arrhythmias (SVEB and VEB), thanks to the *j* index.

3.2. Experiment 1: Features evaluation

An OAO SVM model was independently trained for each feature in order to compare their single performance. Table 4 shows the results that were obtained for the different models over the evaluation set (DS2) from the MIT-BIH database. The included performance measurements are: the sensitivity (*Se*) and the positive predictive value (P^+) for each class, the overall accuracy (*Acc*), the mean *Se* and P^+ of the four classes, the *j index*, the Cohen's kappa (κ *index*) and the *j*κ *index*. The best results regarding the most important arrhythmias (SVEB and VEB) are obtained by the *HOS* feature, which achieved the best *j index*, followed very closely by our morphological descriptor. Conversely, the *wavelet* presents a low score for *j index*. The model of this feature classifies most of the beats as class *N* or *VEB*, achieving the highest overall accuracy (*Acc*) due to the imbalance, but at the same time it obtain a low score for mean *Se* due to the minority classes (*SVEB* and *F*). In regards to the *j*κ *index*, i.e., considering the four classes with an emphasis on the discrimination of the most important arrhythmias, R-R is the best descriptor. Finally, the *LBP* obtain the worst *j*κ *index* score by far.

Table 4
Results of the OAO SVM classifiers trained with the different features over MIT-BIH (DS2). Best feature per measurement in bold.

	N		SVEB		VEB		F		Average					
Features	*Se*	P^+	*Se*	P^+	*Se*	P^+	*Se*	P^+	*Se*	P^+	*Acc*	*j index*	*κ index*	*jκ index*
R-R	0.769	**0.989**	0.505	**0.264**	0.802	0.472	**0.874**	**0.058**	**0.738**	0.446	0.762	2.044	0.368	**0.439**
HOS	0.572	0.977	**0.719**	0.106	0.736	**0.690**	0.765	0.045	0.698	**0.455**	0.589	**2.251**	0.216	0.389
Wavelet	**0.857**	0.953	0.106	0.079	**0.959**	0.426	0.013	0.056	0.484	0.378	**0.826**	1.570	**0.384**	0.388
Our Morph.	0.468	0.958	0.707	0.112	0.771	0.610	0.030	0.001	0.494	0.420	0.494	2.201	0.154	0.352
LBP	0.744	0.921	0.005	0.008	0.524	0.293	0.003	0.000	0.319	0.307	0.686	0.846	0.132	0.172

Table 5
All the possible combinations of the five features tested with the single SVM model and the ensembles of SVMs over MIT-BIH (DS2). Ensembles of SVMs are combined with product, sum, and majority rules. Best results per configuration and measurement in bold.

Features					Single-SVM				Product rule				Sum rule				Majority rule			
R-R	Wavelets	HOS	LBP	Our Morph.	Acc	*Se*	P^+	$j\kappa$ index	Acc	*Se*	P^+	$j\kappa$ index	Acc	*Se*	P^+	$j\kappa$ index	Acc	*Se*	P^+	$j\kappa$ index
•	•				0.848	0.518	0.465	0.486	0.866	**0.568**	0.459	0.525	0.857	0.560	0.450	0.501	**0.893**	0.531	**0.472**	**0.531**
•		•			**0.874**	0.552	0.531	0.555	0.837	**0.865**	**0.581**	**0.637**	0.836	0.863	0.579	0.635	0.853	0.781	0.577	0.588
•			•		**0.902**	0.484	**0.470**	**0.499**	0.836	**0.516**	0.438	0.450	0.832	0.497	0.427	0.425	0.834	0.474	0.432	0.389
•				•	0.792	0.617	0.470	0.526	0.879	**0.673**	**0.579**	**0.669**	0.873	0.667	0.573	0.656	**0.899**	0.616	0.578	0.650
	•	•			0.841	0.440	0.473	0.400	0.835	0.665	0.519	0.519	0.821	**0.672**	0.504	0.500	**0.923**	0.502	**0.524**	**0.571**
	•		•		**0.909**	**0.456**	**0.438**	**0.458**	0.820	0.438	0.345	0.333	0.821	0.442	0.349	0.340	0.823	0.395	0.324	0.278
	•			•	0.844	0.428	**0.466**	0.385	0.779	0.500	0.458	0.422	0.765	**0.520**	0.462	**0.434**	**0.873**	0.447	0.461	0.429
		•	•		**0.909**	0.448	0.442	**0.457**	0.768	**0.542**	**0.472**	0.385	0.761	0.536	0.459	0.370	0.833	0.443	0.463	0.360
		•		•	0.779	0.647	0.505	0.511	0.765	0.646	0.510	0.512	0.745	**0.661**	0.510	0.504	**0.845**	0.594	**0.532**	**0.549**
			•	•	**0.912**	**0.431**	0.411	**0.429**	0.746	0.422	**0.424**	0.328	0.728	0.429	0.423	0.330	0.820	0.378	0.401	0.282
•	•	•			0.901	0.522	0.530	0.565	0.901	0.815	**0.606**	0.690	0.896	**0.818**	0.603	0.679	**0.916**	0.687	0.595	**0.706**
•	•		•		**0.926**	0.491	**0.503**	**0.558**	0.869	**0.503**	0.411	0.462	0.869	0.500	0.408	0.457	0.867	0.471	0.399	0.425
•	•			•	0.883	0.507	0.500	0.518	**0.926**	0.650	**0.572**	**0.711**	0.921	**0.651**	0.566	0.703	0.917	0.640	0.562	0.688
•		•	•		**0.930**	0.525	0.565	0.592	0.874	**0.759**	**0.581**	**0.630**	0.870	0.759	0.578	0.616	0.876	0.706	0.576	0.617
•		•		•	0.884	0.696	0.558	0.640	**0.921**	0.782	**0.635**	**0.742**	0.907	**0.801**	0.622	0.719	0.890	0.706	0.586	0.668
•			•	•	**0.930**	0.491	0.509	0.569	0.911	**0.626**	**0.562**	**0.670**	0.903	0.618	0.560	0.656	0.892	0.608	0.550	0.611
	•	•	•		**0.923**	**0.470**	**0.476**	**0.514**	0.861	0.469	0.403	0.417	0.853	0.470	0.394	0.408	0.864	0.449	0.400	0.398
	•	•		•	**0.876**	0.442	0.491	0.425	0.845	0.607	0.517	0.563	0.831	**0.650**	0.516	0.555	0.845	0.601	**0.518**	**0.564**
	•		•	•	**0.912**	**0.458**	**0.443**	**0.469**	0.837	0.445	0.374	0.366	0.825	0.444	0.371	0.358	0.811	0.407	0.344	0.294
		•	•	•	**0.917**	0.454	0.434	0.476	0.830	0.592	0.507	**0.524**	0.823	**0.600**	**0.508**	0.521	0.824	0.568	0.500	0.501
•	•	•	•		**0.933**	0.509	**0.547**	**0.604**	0.902	**0.602**	0.528	0.581	0.893	0.590	0.506	0.548	0.913	0.553	0.534	0.587
•	•	•		•	0.900	0.523	0.532	0.567	**0.945**	0.703	0.664	**0.773**	0.943	**0.736**	**0.674**	0.771	0.943	0.640	0.620	0.745
•	•		•	•	**0.926**	0.494	**0.510**	**0.562**	0.908	**0.525**	0.480	0.552	0.902	0.515	0.465	0.530	0.897	0.509	0.468	0.517
•		•	•	•	**0.940**	0.505	0.584	0.627	0.930	0.707	**0.625**	**0.732**	0.920	**0.727**	0.614	0.712	0.906	0.655	0.589	0.653
	•	•	•	•	**0.922**	0.470	**0.471**	**0.508**	0.866	0.502	0.469	0.480	0.856	**0.506**	0.458	0.469	0.890	0.468	0.458	0.467
•	•	•	•	•	0.933	0.509	0.551	0.606	**0.938**	**0.625**	**0.617**	**0.707**	0.933	0.621	0.596	0.692	0.934	0.621	0.587	0.704

3.3. *Experiment 2: Comparison of single SVM vs. combination of multiple SVMs*

The goal of this experiment is: evaluate if an ensemble of OAO SVMs, combining the decision of the previous models, improves the results over a single OAO SVM model trained with all the features together. Table 5 contains the results obtained for all the possible configurations of the employed features for the two alternatives. For the ensemble case, three combination rules were employed: the product, the sum and the majority rule. As we previously said, we compare the results of the methods with the *jκ index* measurement. The values of the overall accuracy (*Acc*), the mean *Se* and P^+ are also displayed. Results in Table 5 show that, in general, ensembles of SVMs produce superior scores than a single SVM, especially when the product rule is employed. However, there is an exception when the LBP descriptor is present. Note that single SVM models have only superior *jκ index* values than their ensemble approaches when the LBP descriptor is present. This is due to the fact that when the ensemble of SVMs is used, all the features add the same amount of confidence to the final decision. This causes a deterioration of the performance if a feature is significantly worst than the rest. On the other hand, when a single SVM is employed, the training process itself may discard the feature dimension that behave worst. In general, the more features are added the better *jκ index* is achieved. Not surprisingly, the best configurations are those that include the R-R interval, which was the best single feature. The higher *jκ index* = 0.773 score was achieved by the configuration $R-R$, *Wavelets*, *HOS*, and *Our Morph*. with an ensemble of SVMs using the product rule. On the other hand, the best single SVM configuration of *jκ index* = 0.640 was achieved by $R-R$, *HOS*, and *Our Morph*.

3.4. *Experiment 3: Comparison with the state-of-the-art*

The goal of this experiment is to compare the result of our best configurations against other classification approaches, which also employed the MIT-BIH public database with the same inter-patient division. As indicated, this is a well-known public database used as reference for validation of computational proposals of the issue. Table 6 includes the comparison of our best configuration of ensemble of SVMs and single SVM, next to some of the best state-of-the-art methods. Results on Table 6 show that our ensemble achieves more than a 10% of improvement regarding to the *jκ index* in comparison with the Zhang et al. method [22], which is the second highest one. On the other hand, our approach with a single SVM behaves similar to the state-of-the-art methods. As we can see, looking at class level, the highest positive predictive value is obtained by our ensemble method for all classes, except for the normal class (*N*). This means that our method tends to be more conservative at assigning abnormal classes (*SVEB*, *VEB*, *F*) as the

Table 6
Results on MIT-BIH (DS2) comparing our best configurations against state-of-the-art methods.

	N		SVEB		VEB		F		Average					
Classifier	*Se*	P^+	*Se*	P^+	*Se*	P^+	*Se*	P^+	*Se*	P^+	*Acc*	*j index*	*κ index*	*jκ index*
Our Ensemble SVM	**0.959**	0.982	0.781	**0.497**	**0.947**	**0.939**	0.124	**0.236**	0.703	**0.664**	**0.945**	**3.165**	**0.755**	**0.773**
Zhang et al. [22]	0.889	0.990	0.791	0.359	0.855	0.927	**0.938**	0.137	**0.868**	0.604	0.883	2.934	0.592	0.663
Our Single SVM	0.895	0.982	0.670	0.349	0.933	0.849	0.286	0.055	0.696	0.559	0.884	2.800	0.579	0.640
Mar et al. [9]	0.896	0.991	**0.832**	0.335	0.868	0.759	0.611	0.166	0.802	0.564	0.899	2.798	0.599	0.649
Chazal et al. [2]	0.871	**0.992**	0.760	0.385	0.803	0.866	0.894	0.086	0.832	0.570	0.862	2.767	0.532	0.612

Table 7
Confusion matrix over (DS2) MIT-BIH of our best configuration: ensemble of SVM (*R* – *R*, *W*, *HOS*, *Our. Morph*.) using the product rule.

	Algorithm				
	n	s	v	f	Total
Reference					
N	42,244	1540	99	150	44,033
S	427	1601	21	1	2050
V	90	75	3051	4	3220
F	256	2	82	48	388
Total	43,017	3218	3253	203	49,691

normal class (*N*) than the other methods. Regarding the sensitivity, our methods achieved higher values comparing to the state-of-the-art methods for the majority classes, *N* and *VEB*. But at the same time, the lowest sensitivity for class *F* was achieved by our methods, causing their low value of mean sensitivity. This must be due to the inclusion of the *Wavelets* and *Our Morphology* features, which obtained considerably lower sensitivity than the *RR* or *HOS* features for this class (see Table 4). Looking at the confusion matrix (Table 7), it is noticeable that most of the *F* beats were misclassified as class *N* and *VEB*. However, note also that due to the highly imbalanced data samples from *F* class correspond only with a 1% of the total samples from MIT-BIH. Considering more appropriate measurements for this database that take into account the imbalance of the data, like *κ index* and *jκ index*, our methods present good results, especially our ensemble of SVMs approach.

4. Conclusions

A new approach for ECG classification based on an ensemble of SVMs was proposed. All the experiments were performed on the MIT-BIH public database, following an inter-patient scheme division. In order to evaluate the results the *jκ index*, which were proposed as an adequate performance measurement for this database, has been employed. We tested several feature descriptors, including: R-R intervals, wavelets, HOS, LBP, and our own morphological descriptor. In the first experiment, a SVM model was trained for each descriptor, being R-R intervals the one that obtained the highest *jκ index*. For the second experiment, we evaluated the improvement of the ensemble of SVMs against a single SVM. All the possible combinations of the five feature descriptors and the three combination rules, the product, the sum, and the majority were tested. The obtained results show that, in the majority of the cases, our approach combining multiple SVM models is superior than concatenating all the features and training a single SVM model. Only when the LBP descriptor is employed, which was the worst single descriptor, our ensemble approach does not improve. For the best configuration, employing an ensemble of SVMs using R-R interval, wavelets, HOS and our morphological descriptor, combined with the product rule, the score obtained for *jκ index* is over a 10% better than the previous machine learning approaches of the state-of-the-art. Additionally, it must be emphasised that our method only requires the QRS detection for the segmentation step and one single lead (lead II) for the feature extraction. Instead of that, other state-of-the-art methods may require many leads [9,22] and a more complex segmentation step [2,9,22] that includes the computation of the position and duration of P, QRS, and T waves. The highest complexity in the segmentation implies a higher error probability during this step.

Possible future works, include the use of multiple leads, and also the addition of more sophisticated data fusion methodologies, employing techniques such as Dempster–Shafer theory of the evidence [41]. Ideally, each classifier model from a certain feature descriptor behaves better than the others at specific cases, hence, a system that assigns more confidence to the right model at those cases, would increase the performance of the system.

All the code developed in this work is publicly available on the repository.[1]

Acknowledgements

This work was partially supported by the Research Project RTC-2016-5143-1, financed by the Spanish Ministry of Economy, Industry and Competitiveness and the European Regional Development Fund (ERDF). Also, this work has received financial support from the ERDF and the Xunta de Galicia, Centro singular de investigación de Galicia accreditation 2016–2019, Ref. ED431G/01; and Grupos de Referencia Competitiva, Ref. ED431C 2016-047.

References

[1] E.J. da S. Luz, T.M. Nunes, V.H.C. de Albuquerque, J.P. Papa, D. Menotti, ECG arrhythmia classification based on optimum-path forest, Expert Syst. Appl. 40 (9) (2013) 3561–3573, http://dx.doi.org/10.1016/j.eswa.2012.12.063.

[2] 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, http://dx.doi.org/10.1109/TBME.2004.827359.

[3] S. Meek, F. Morris, Introduction. I – Leads, rate, rhythm, and cardiac axis, BMJ 324 (7334) (2002) 415–418, http://dx.doi.org/10.1136/bmj.324.7334.415.

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

[5] Y.C. Yeh, W.J. Wang, QRS complexes detection for ECG signal: the difference operation method, Comput. Methods Prog. Biomed. 91 (3) (2008) 245–254, http://dx.doi.org/10.1016/j.cmpb.2008.04.006.

[6] H. Li, X. Wang, L. Chen, E. Li, Denoising and R-Peak detection of electrocardiogram signal based on EMD and improved approximate envelope, Circ. Syst. Signal Process. 33 (4) (2014) 1261–1276, http://dx.doi.org/10.1007/s00034-013-9691-3.

[7] H. Li, X. Wang, Detection of electrocardiogram characteristic points using lifting wavelet transform and Hilbert transform, Trans. Inst. Meas. Control 35 (5) (2013) 574–582, http://dx.doi.org/10.1177/0142331212460720.

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

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

[10] A.S. Al-Fahoum, I. Howitt, Combined wavelet transformation and radial basis neural networks for classifying life-threatening cardiac arrhythmias, Med. Biol. Eng. Comput. 37 (5) (1999) 566–573, http://dx.doi.org/10.1007/BF02513350.

[1] https://github.com/mondejar/ecg-classification.

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

[12] G. de Lannoy, D. François, J. Delbeke, M. Verleysen, Weighted SVMs and Feature Relevance Assessment in Supervised Heart Beat Classification, Springer, Berlin, Heidelberg, Berlin, Heidelberg, 2011, pp. 212–223, http://dx.doi.org/10.1007/978-3-642-18472-7_17, ISBN:978-3-642-18472-7.

[13] C.C. Lin, C.M. Yang, Heartbeat classification using normalized RR intervals and morphological features, Math. Probl. Eng. (2014) 10, http://dx.doi.org/10.1155/2014/712474.

[14] S. Chen, W. Hua, Z. Li, J. Li, X. Gao, Heartbeat classification using projected and dynamic features of ECG signal, Biomed. Signal Process. Control 31 (2017) 165–173, http://dx.doi.org/10.1016/j.bspc.2016.07.010.

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

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

[17] H. Khorrami, M. Moavenian, A comparative study of DWT, CWT and DCT transformations in ecg arrhythmias classification, Expert Syst. Appl. 37 (8) (2010) 5751–5757, http://dx.doi.org/10.1016/j.eswa.2010.02.033.

[18] R.J. Martis, U.R. Acharya, K. Mandana, A. 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, http://dx.doi.org/10.1016/j.eswa.2012.04.072.

[19] H. Li, D. Yuan, X. Ma, D. Cui, L. Cao, Genetic algorithm for the optimization of features and neural networks in ECG signals classification, Scientific Reports (2017), http://dx.doi.org/10.1038/srep41011.

[20] W. Lu, H. Hou, J. Chu, Feature fusion for imbalanced ECG data analysis, Biomed. Signal Process. Control 41 (2018) 152–160, http://dx.doi.org/10.1016/j.bspc.2017.11.010.

[21] F. Melgani, Y. Bazi, Classification of electrocardiogram signals with Support Vector Machines and Particle Swarm Optimization, IEEE Trans. Inf. Technol. Biomed. 12 (5) (2008) 667–677, http://dx.doi.org/10.1109/TITB.2008.923147.

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

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

[24] H. Li, H. Liang, C. Miao, L. Cao, X. Feng, C. Tang, et al., Novel ECG signal classification based on KICA nonlinear feature extraction, Circ. Syst. Signal Process. 35 (4) (2016) 1187–1197, http://dx.doi.org/10.1007/s00034-015-0108-3.

[25] T.G. Dietterich, Ensemble Methods in Machine Learning, Springer, Berlin, Heidelberg, Berlin, Heidelberg, 2000, pp. 1–15, http://dx.doi.org/10.1007/3-540-45014-9_1, ISBN:978-3-540-45014-6.

[26] L. Breiman, Bagging predictors, Mach. Learn. 24 (2) (1996) 123–140, http://dx.doi.org/10.1023/A:1018054314350.

[27] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. Syst. Sci. 55 (1) (1997) 119–139, http://dx.doi.org/10.1006/jcss.1997.1504.

[28] T.G. Dietterich, G. Bakiri, Solving multiclass learning problems via error-correcting output codes, J. Artif. Int. Res. 2 (1) (1995) 263–286.

[29] R.P.W. Duin, D.M.J. Tax, Experiments with classifier combining rules, in: Multiple Classifier Systems, Springer, Berlin, Heidelberg, Berlin, Heidelberg, 2000, pp. 16–29, ISBN:978-3-540-45014-6.

[30] B. Waske, J.A. Benediktsson, Fusion of Support Vector Machines for classification of multisensor data, IEEE Trans. Geosci. Remote Sens. 45 (12) (2007) 3858–3866, http://dx.doi.org/10.1109/TGRS.2007.898446.

[31] J. Kittler, M. Hatef, R.P.W. Duin, J. Matas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 20 (3) (1998) 226–239, http://dx.doi.org/10.1109/34.667881.

[32] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, et al., PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) e215–e220.

[33] Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms, Association for the Advancement of Medical Instrumentation, 1998.

[34] P. de Chazal, Detection of supraventricular and ventricular ectopic beats using a single lead ECG, 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (2013) 45–48, http://dx.doi.org/10.1109/EMBC.2013.6609433.

[35] Y. Kaya, M. Uyar, R. Tekin, S. Yldrm, 1d-local binary pattern based feature extraction for classification of epileptic EEG signals, Appl. Math. Comput. 243 (2014) 209–219, http://dx.doi.org/10.1016/j.amc.2014.05.128.

[36] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297, http://dx.doi.org/10.1023/A:1022627411411.

[37] J. Milgram, M. Cheriet, R. Sabourin, "One against one" or "one against all": which one is better for handwriting recognition with SVMs? in: G. Lorette (Ed.), Tenth International Workshop on Frontiers in Handwriting Recognition. Université de Rennes 1, La Baule (France): Suvisoft, 2006 https://hal.inria.fr/inria-00103955.

[38] M.L. Soria, J.P. Martinez, An ECG classification model based on multilead wavelet transform features 2007 Computers in Cardiology (2007) 105–108, http://dx.doi.org/10.1109/CIC.2007.4745432.

[39] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Meas. 20 (1) (1960) 37–46, http://dx.doi.org/10.1177/001316446002000104.

[40] M. Fatourechi, R.K. Ward, S.G. Mason, J. Huggins, A. Schlgl, G.E. Birch, Comparison of evaluation metrics in classification applications with imbalanced datasets, 2008 Seventh International Conference on Machine Learning and Applications (2008) 777–782, http://dx.doi.org/10.1109/ICMLA.2008.34.

[41] T. Denoeux, A k-nearest neighbor classification rule based on Dempster–Shafer theory, IEEE Trans. Syst. Man Cybern. 25 (5) (1995) 804–813, http://dx.doi.org/10.1109/21.376493.