

# The ensemble of algorithms for coronary heart disease detection based on electrocardiogram

V N Guryanova<sup>1</sup>

<sup>1</sup>Lomonosov Moscow State University, Leninskie Gory 1, Moscow, Russia, 119991

**Abstract.** Coronary heart disease (CHD) is the leading cause of death in the world. This disease can be asymptomatic for a long time and over time can progress and result in death. Today electrocardiogram (ECG) can be done at home with the help of special equipment from CardioQvark. In this paper the possibility of CHD detection based on such ECGs was explored. Different approaches to the classification of such electrocardiograms were surveyed. New algorithms and modifications to existing algorithms were proposed. A new method – the ensemble of different algorithms – has shown the best performance.

## 1. Introduction

Coronary heart disease (CHD) [1] is a group of diseases, that is defined by lack of oxygen supply to the heart muscle through the coronary arteries. According to World Health Organization, this disease is the leading cause of death in the world.

At the initial stages of the disease, most people do not show any symptoms of this disease. It is very important to identify the CHD in time to slow the course of the disease and prevent the patient's death.

Traditionally, CHD can be detected with the help of specialists and a number of tests. It should be noted that these tests take a significant amount of patient's time, and also require high qualification of the specialist who will conduct them. Since there are very few specialists and the number of potential patients is growing every year, the task of automatically determining CHD is extremely urgent at the present time. Currently, it is extremely relevant to create a device that will help determine the disease or its probability at home. Such devices will allow the person to be sent to a doctor in case of a high probability of having a CHD.

An electrocardiogram (ECG) – is a signal that reflects the electrical activity of the heart. ECG is one of the most affordable ways of diagnosing heart disease now due to its non-invasiveness and low cost. Currently, there are many different studies that show that the ECG can be used to determine CHD [2], [3], [4], [5].

CardioQvark (project site: [www.cardioqvark.ru](http://www.cardioqvark.ru)) has created a device in a form of a smartphone case that allows you to make ECG measurements at home. The CardioQVARK device is a portable electrocardiograph in the form of a smartphone case (iPhone 5 / 5s / SE / 6 / 6s), allowing to register data of bio-electrical activity of the heart from the first ECG lead and next leads: aVR, aVL, aVF, Vi (i = 1 ... 6) using the patient's cable. In this work, the possibility of CHD detection based on such ECGs from the first lead was explored.

There are different approaches to ECG classification problem. Some of them were surveyed in this work. New algorithms and modifications to existing algorithms were proposed. In order to improve the performance of the classification, an ensemble of 5 different methods was built. Each of these methods will be described below.

## 2. Data Description

All research was conducted on the basis of the following medical centers: NGHCI "Semashko Central Clinical Hospital 2", Federal State Scientific Institution "Petrovsky Russian Scientific Center of Surgery", Federal State-Funded Health Care Institution "City Clinical Hospital 4 Health Care Moscow Department", Federal State-Funded Health Care Institution "Moscow Clinical Scientific Center of Moscow Department Health Care", State Autonomous Health Care Institution of the Moscow Region "Clinical Center for Restorative Medicine and Rehabilitation".

A voluntary anonymous study included patients over 18 years of age. Annotated impersonal electrocardiograms (ECG) were recorded from the first ECG-lead using a CardioQVARK cardio monitor. The duration of each recording was 5 minutes. The measurement was taken in the sitting position, with support for the back, hands on the knees or on the table. Data was collected in dynamics of 3-10 observations with an interval of at least 12 hours between measurements.

The sample that was used for this task consists of 1798 cardiograms. It contained 1055 cardiograms of healthy patients and 743 cardiograms of patients with CHD. The sampling frequency was 1000 Hz.

Signals were preprocessed before applying machine learning algorithms. For preprocessing, a low-pass and high-pass Butterworth filters [6] of the second order were used. For low-pass filter cutoff frequency was 0.3 Hz. For high-pass filter cutoff frequency was 15 Hz. The signal trend was extracted using a median filter [6]. Then the trend was subtracted from the preprocessed signal.

## 3. Algorithms Description

Below are descriptions of the algorithms that were used to build the ensemble.

### 3.1. The algorithm based on the HRV signal

The idea which was used for this algorithm was described in the article [2]. In the ECG signal, R-peaks can be distinguished, which correspond to the person's pulse [6]. ECG signal is used to create heart rate variability signal (HRV signal). It is calculated as follows.

- R-peaks are computed.
- The intervals between two R-peaks (RR-intervals) are measured.
- Each value of the RR-interval is converted to  $60/RR$ .

The main idea of this method is a construction of various groups of features from HRV signal.

The first group of features includes various entropic features, which indicate a measure of unpredictability in the signal. The following types of entropies are used: approximate entropy, sample entropy, and Shannon entropy. Each type of entropy is described in detail below.

Approximate entropy is calculated as follows. Here  $x = (x_0, x_1, \dots, x_{N-1})$  is the HRV signal of length  $N$ .

- The integer  $m$  and the real  $r$  are fixed.
- A set of vectors of the form  $\overline{x}_i^m = (x_i, x_{i+1}, \dots, x_{i+m-1})$ , where  $i \in [0, N - m]$  is composed.
- The values  $C_i^m(r)$  are calculated as follows:

$$C_i^m(r) = \frac{|\overline{x}_k^m : d(\overline{x}_i^m, \overline{x}_k^m) \leq r, k \in [0, N - m]|}{N - m + 1},$$

where

$$d(\overline{x}_i^m, \overline{x}_k^m) = \max_{a \in [0, m-1]} |\overline{x}_i^m(a) - \overline{x}_k^m(a)|,$$

where the  $\overline{x}_i^m(a)$  – component  $a$  of the vector  $\overline{x}_i^m$ .

- The values  $\Phi^m(r)$  are defined as:

$$\Phi^m(r) = (N - M + 1)^{-l} \sum_{i=1}^{N-m+1} \log(C_i^m(r)).$$

- Approximate entropy (ApEn) is defined as

$$\text{ApEn}(r) = \Phi^m(r) - \Phi^{m+1}(r).$$

In this paper, the approximate entropy was realized for  $m = 10$  and  $r = 0.2\text{std}(x)$ ,  $\text{std}(x)$  – standard deviation of the signal  $x$ .

The Sample entropy is calculated as follows.

- Vectors  $\overline{x}_i^m$  of length  $m$  and vectors  $\overline{x}_i^{m+1}$  of length  $m + 1$  are formed similar to those that were formed in approximate entropy.
- The values  $A$  and  $B$  are calculated as follows:

$$A(r) = \left| (\overline{x}_k^{m+1}, \overline{x}_l^{m+1}) : d(\overline{x}_k^{m+1}, \overline{x}_l^{m+1}) \leq r, 0 \leq k \leq l \leq N - m - 1 \right|,$$

$$B(r) = \left| (\overline{x}_k^m, \overline{x}_l^m) : d(\overline{x}_k^m, \overline{x}_l^m) \leq r, 0 \leq k \leq l \leq N - m \right|,$$

where  $d$  is determined as in approximate entropy.

- Sample entropy (SampEn) is defined as

$$\text{SampEn}(r) = -\log \frac{A(r)}{B(r)}.$$

In this paper, the sample entropy was determined for  $m = 10$ ,  $r = 0.2\text{std}(x)$ .

The Shannon entropy is calculated as

$$\text{ShanEn} = \sum_{f=1}^k p_f \log p_f,$$

where  $k$  is the number of different elements in the signal  $x$ ,  $p_f$  is the frequency of the element  $f$  in the signal  $x$ .

The following group of features is based on a recurrence plot. This plot shows the frequency and duration of repetitions in the signal. The element  $(i, j)$  of the given plot is defined as 1, if  $\|x_i - x_j\| < \varepsilon$ , or as 0 otherwise, where  $x$  is the HRV signal.

Based on this plot, the following features are calculated,  $N$  – the number of elements in the HRV signal,  $l_{d \min}$  – the minimum length of the plot diagonal,  $l_{v \min}$  – minimal length of the vertical line,  $l_{d \max}$  – maximum length of the diagonal,  $l_{v \max}$  – maximum length of the vertical line:

- Density of points (REC)

$$\text{REC} = \frac{1}{N^2} \sum_{i,j=0}^N R(i, j).$$

- The percentage of points that form the diagonal lines (DET)

$$\text{DET} = \frac{\sum_{l=l_d \min}^{l_d \max} lP(l)}{\sum_{i,j}^N R(i, j)},$$

$P(l)$  – is the number of diagonals of length  $l$ .

- The average length of the diagonals ( $L_{\text{mean}}$ )

$$L_{\text{mean}} = \frac{\sum_{l=l_d \min}^{l_d \max} lP(l)}{\sum_{l=l_d \min}^{l_d \max} P(l)}.$$

- Entropy of diagonal lines ( $\text{EN}_d$ )

$$\text{EN}_d = - \sum_{l=l_d \min}^{l_d \max} p_l \log p_l,$$

where  $p_l$  is the frequency of diagonal lines of length  $l$ .

- Entropy of vertical lines ( $\text{EN}_v$ )

$$\text{EN}_v = - \sum_{l=l_v \min}^{l_v \max} p_{vl} \log p_{vl},$$

where  $p_{vl}$  is the frequency of vertical lines of length  $l$ .

For the calculation of these features, the PyRQA library [7] was used.

Another group of features that is used for this approach is a group of features based on the Poincare plot. The Poincare plot is constructed as follows: for the signal  $x = (x_0, x_1, \dots, x_{N-1})$ , the plot consists of the points

$$(x_0, x_1), (x_1, x_2), \dots, (x_i, x_{i+1})$$

and so on. In this case, RR-intervals were used as a signal  $x$ . The following features are constructed:

- The standard deviation of distances from the points of the plot to the line  $y = x$ . This feature describes the local variability of RR-intervals.
- The standard deviation of distances from the points of the plot to the line  $y = -x - 2RR_{\text{mean}}$ .  $RR_{\text{mean}}$  – is the average value of RR-intervals. This feature describes the long-term variability of RR-intervals.

The feature, which is based on the detrended fluctuation analysis. This method allows determining the self-dependence of the signal. The following cumulative sum is defined as:

$$x_{\text{cumsum}}(t) = \sum_{i=1}^t (x(i) - \mu),$$

$x$  is a signal consisting of RR-intervals,  $\mu$  is the mean of the signal  $x$ .

The data is segmented with a window of size  $\Delta n$ . On each segment polynomial is found for the data, which most accurately represents it (usually linear). The union of all such polynomials forms a function  $x_{\Delta n}(t)$ , which is an approximation of the original function  $x_{\text{cumsum}}(t)$ . Then there is the following function:

$$F(\Delta n) = \sqrt{\frac{1}{N} \sum_{t=1}^N [x_{\text{cumsum}}(t) - x_{\Delta n}(t)]^2},$$

where  $N$  is the length of the signal consisting of RR-intervals.

The feature is the slope of the line  $\log F(\Delta n)$  to  $\log(\Delta n)$ . More information about this approach is written in [8]. In this paper, the implementation of this feature was computed using publicly available software Nonlinear measures for dynamical systems or nolds, version 0.3.2, which can be downloaded from (<https://pypi.python.org/pypi/nolds>).

The next feature that was used in this approach is the correlation of dimensions. This feature is a quantitative characteristic of the signal trajectory and is defined as follows.

- Vectors  $\overline{x}_i^m$  of length  $m$  are similar to those that were formed in approximate entropy are constructed.
- The value  $g$  is defined as:

$$g(r) = |\{(\overline{x}_k^m, \overline{x}_l^m) : d(\overline{x}_k^m, \overline{x}_l^m) \leq r, 0 \leq k \leq l \leq N - m - 1\}|,$$

$g$  – is the number of pairs of vectors whose distance is less than or equal to  $r$ .

- The value  $C(r)$  is defined as

$$C(r) = \frac{g(r)}{N^2},$$

where  $N$  – is the length of the signal  $x$ .

- The correlation of dimensions ( $D2$ ) is defined as

$$D2 = \lim_{r \rightarrow 0} \frac{\log C(r)}{\log(r)}.$$

In this paper, the implementation of this feature was computed using publicly available software Nonlinear measures for dynamical systems or nolds, version 0.3.2, which can be downloaded from (<https://pypi.python.org/pypi/nolds>).

The gradient boosting from the xgboost package [9] was used as the classifier in this approach.

### 3.2. The algorithm based on 3 different feature spaces

This algorithm is a mixture of 3 different feature spaces, which were previously used in the classification of biomedical signals.

The first group of features consists of the parameters of the Hjorth's parameters: activity, mobility, complexity [10]. These parameters were originally used as features for electroencephalograms. Later, these parameters were used in many works, including the classification of ECG signals [11].

The second group of features consists of statistical signal features: mean, standard deviation, signal minimum, signal maximum, skew, kurtosis, selective quantiles of order: 0.1, 0.25, 0.5, 0.75, 0.9, sums and sums of signal values squares that are above / below certain values of quantiles: 0.1, 0.25, 0.5, 0.75, 0.9.

The next group of features was suggested by Uspenskiy for the disease detection by patient's ECG [12]. To calculate these features, it is necessary to calculate the amplitudes of the R-peaks ( $A(n)$ ), the distances between the R-peaks ( $T(n)$ ), and the arctangent of their ratio

$$\alpha(n) = \text{arctg} \frac{A(n)}{T(n)}.$$

**Table 1.** Signal encoding for Uspenskiy features

	A	B	C	D	E	F
$A(n+1) - A(n)$	+	-	+	-	+	-
$T(n+1) - T(n)$	+	-	-	+	+	-
$\alpha(n+1) - \alpha(n)$	+	+	+	-	-	-

It is assumed that the values of  $A(n)$ ,  $T(n)$  are not important, but signs of their increments are. The method of signal encoding based on all possible signs of increments of these quantities is presented in Table 1.

After the code representation of the signal is received, the three-gram selection is performed. The feature space is the number of occurrences of each of the possible three-grams in a given code sequence derived from the signal.

The logistic regression from the scikit-learn package [13] was used as the classifier in this approach.

### 3.3. The algorithm based on R-peak's neighborhood

The idea used in this algorithm was described in the article [14] for determining the state of the patient's heart in which he should be sent to the cardiac service. The feature space for this approach is constructed as follows.

- The neighborhoods of the signal R-peaks are allocated: 200 points before R-peak and 500 after.
- The averaged neighborhood is used as a feature space.

Neural network with the architecture described in Table 2 was used as the classification model for this algorithm.

**Table 2.** Neural Network Structure for the algorithm based on R-peaks neighborhoods

Input Layer	Shape = (700)
Dense Layer	Units Number = 90 Activation Function = sigmoid
Dense Layer	Units Number = 1 Activation Function = sigmoid

In this work, the neural network was implemented using the libraries Theano [15] and Lasagne [16].

### 3.4. The algorithm based on wavelet transformation

The idea for this algorithm was described in paper [17] for epilepsy detection based on ECG-signal and for identification of a person based on ECG-signal.

Wavelet signal transformation is a convolution of the signal with functions  $\Psi(t)$ , called wavelets. Such wavelet functions should possess specific properties:

$$\int_{-\infty}^{+\infty} \Psi(t)dt = 0 \quad \int_{-\infty}^{+\infty} |\Psi(t)|^2 dt < \infty.$$

Wavelet transformation allows achieving signal compression, with good performance of reproduced original signal [6].

All wavelet functions used in a wavelet transformation can be represented with prototype function  $\psi(t)$  using scaling and shift. In case of discrete wavelet transformation all wavelet function can be written as:

$$\psi_{m,n}(t) = \frac{1}{\sqrt{2^m}} (2^{-m}t - n).$$

In discrete wavelet transformation of  $\psi_{m,n}^*$  function can be separated into two parts. The two types of functions correspond to coefficients of approximation and detail coefficients. In the present work approximation coefficients are used for new representation of the signal, and Daubechies wavelets [18] are used as a wavelet function.

After getting the coefficients of wavelet transformations, local segments are extracted from the signal. The segments are extracted by moving a window of certain length  $w$  with step  $s$ , with all elements within one window going into a separate segment. After such a procedure every signal is represented as a set of local segments.

All local segments in the training set are separated into  $k$  clusters using k-means algorithm. After this, every segment is replaced with the number of the cluster it belongs to. This way every signal is represented as a text of codewords, with each word representing the certain cluster.

In the implementation of described algorithm these parameter values were used:  $w = 100$ ,  $s = 30$ ,  $k = 200$ . K-means algorithm implementation from scikit-learn package was used.

In the train sample every local segment is replaced with the cluster it is the closest to. That means that for every local segment  $s_i$  in test sample cluster  $c$  it belongs to is determined using this formula:

$$c = \operatorname{argmin} d(b_j, s_i), \quad d(b_j, s_i) = \sqrt{\sum_{k=1}^w (s_i^k - b_j^k)^2},$$

where  $b_j$  is a cluster center  $j$ ,  $s_i^k$  ( $b_j^k$ ) is a  $k$ th element of local segment  $i$  (cluster  $j$ ).

Using this approach and transforming the input signal into text it is possible to use natural language processing algorithms. The paper's authors that suggested such encoding used the bag of words as a feature space. Feature description is a number of occurrences of each code word in a specific signal.

Features based on word2vec technology were used in order to extract dependencies between local segments in the signal. This approach was suggested by Google and it allows to use context-aware text processing, reducing the dimensions of the data [19].

Word2Vec model was trained using the length of embedding vector equals to 80. A mean of all vectors in the signal was used as a feature. The model was trained using package gensim [20]. The logistic regression from scikit-learn package [13] was used as a classification algorithm.

### 3.5. The algorithm based on bispectrum

Bispectrum is a function of two variables  $f_1$  and  $f_2$  that specify the frequencies, expressed by the following formula [21]:

$$B(f_1, f_2) = X(f_1)X(f_2)X^*(f_1 + f_2),$$

where  $X(f)$  is the Fourier transform of the signal, and  $X^*(f)$  is the complex conjugate of it. The signal bispectrum is usually calculated using a fast Fourier transform. A detailed description of the algorithm for the bispectrum computation of the signal can be found in [22].

During calculation the bispectrum of the signal, we obtain a two-dimensional matrix whose elements are complex numbers.

Based on the resulting matrix, the elements of which can be denoted as  $a(i, j)$ , we can associate each signal with a certain image. This image is calculated as follows. A new matrix  $B = ||b_{i,j}||$  is calculated, the elements of which are equal to:

$$b(i, j) = \sqrt{\text{Re}^2 a(i, j) + \text{Im}^2 a(i, j)},$$

where Re denotes the real part of the complex number, and Im denotes the imaginary part of the complex number. The contour plot of matrix B is used as the image.

The authors of the article [23] have shown that coronary heart disease can be detected by analyzing images obtained from a signal bispectrum. The method described in the above article was to allocate the area of the region within the level lines to conclude that the patient had CHD. The results obtained by the authors allow concluding that bispectrum images can be used to detect CHD.

It was proposed to use neural networks for classification of such images. The architecture of the neural network is described in Table 3.

**Table 3.** Neural Network Structure for the algorithm based on bispectrum

Input Layer	Shape = (3, 80, 80)
Convolution layer	Filter size = (32, 5, 5) Offset = (2,2)
Dense Layer	Units Number = 30 Activation Function = LeakyRelu
Dense Layer	Units Number = 1 Activation Function = Sigmoid

In this work, the neural network was implemented using the libraries Theano [15] and Lasagne [16].

#### 4. Methods of constructing ensembles of algorithms

In order to increase classification performance, it was suggested to use ensembles of the algorithms. Several methods of ensembling are described in this section.

##### 4.1. Majority voting

Given a set of algorithms  $A = (A_1, A_2, \dots, A_n)$  and output a vector of predictions  $a = (a_1, a_2, \dots, a_n)$  then the resulting answer  $a$  of the ensemble is equal to

$$a = \text{mode}(a_1, a_2, \dots, a_n),$$

where mode is a statistic, that is equal to the element which is most often encountered in the predictions. If there are several of them a random one of them is chosen.

##### 4.2. EM-algorithm

The main idea of the EM-algorithm [24] is data aggregation from different people about the same event in order to get a correct evaluation. Since the goal of ensembling is the aggregation of several different algorithms, the EM-algorithm is applicable to ensemble creation.



Algorithm description:

$N$  – size of available data,  $n_{il}^k$  – whether the  $k$ -th algorithm has the answer  $l$  ( $l \in \{1, 2\}$ ) to the data  $i$  ( $i = 1 \dots N$ ),

$\pi_{jl}^k$  ( $j \in \{1, 2\}$ ) – is the probability that the  $k$  th algorithm will return  $j$  when the true answer is  $l$ .

$T_{ij} = 1$ , if the true answer for the data  $i$  is  $j$ , otherwise, it equals 0.

$p_j$  – the probability of class  $j$  in the sample.

- Step 1: Initialize the matrices  $\pi$  to the ideal case. T initialize the voting value for the majority.
- Step 2: Recalculate the values of matrices  $\pi$  and  $p_j$ :

$$\pi_{jl}^k = \frac{\sum_i T_{ij} n_{il}^k}{\sum_l \sum_i T_{ij} n_{il}^k} \quad p_j = \frac{\sum_i T_{ij}}{N}$$

- Step 3. Recalculate  $T_{ij}$ :

$$T_{ij} = \frac{p_j \prod_{k=1}^K \prod_{l=1}^2 (\pi_{jl}^k)^{n_{il}^k}}{\sum_{q=1}^2 p_q \prod_{k=1}^K \prod_{l=1}^2 (\pi_{ql}^k)^{n_{il}^k}}$$

Repeat steps 2 and 3 until the matrices  $\pi$  stop changing.

At the end of this algorithm, we obtain the probability of the data belonging to each class in the matrix  $T$ . As an answer, the class is taken, the probability of belonging to which is the greatest.

## 5. Evaluation of Algorithms

Cross-validation was used to evaluate the performance of algorithm. To avoid overfitting the ECGs of one patient did not fall simultaneously into the training and test set. The following performance criteria was introduced:

$$\frac{1}{N} \sum_{i=1}^N \frac{\sum_{j=1}^{n_i} \mathbf{I}_{t_{ij}=p_{ij}}}{n_i},$$

where  $t_{ij}$  is the true value of the target variable for the cardiogram  $j$  of the patient  $i$ ,  $p_{ij}$  – the predicted value of the target variable for the cardiogram  $j$  of patient  $i$ ,  $n_i$  – the number of cardiograms of the patient  $i$ ,  $N$  – the number of patients,  $\mathbf{I}_{t_{ij}=p_{ij}}$  is the indicator function which equals to 1 if  $t_{ij}$  equals to  $p_{ij}$  and equals to 0 otherwise. This criterion is called patient performance. It allows us to evaluate how well the algorithm determines a person's disease by any of his cardiograms. In addition, this performance criterion does not depend on the number of cardiograms for each patient. ROC-AUC score [25] and F-score [25] were used for models evaluation.

## 6. Results

The results of evaluations are shown in Table 4, where the first column shows algorithms type or ensemble type. The algorithm based on wavelet transformation is included in two variants, with word2vec and without.

## 7. Conclusion

In the course of this paper, the following results were obtained. When CardioQvark equipment is used, it is possible to determine CHD with an accuracy of more than 0.81 for patient performance, with an accuracy greater than 0.77 for F-score and with an accuracy greater than 0.87 for ROC-AUC score. Word2vec can increase the performance of the classification method based on the wavelet transformation. Bispectrum can be used to classify CHD. The EM algorithm is applicable for ensemble and in this case, shows the best performance of classification for all selected performance criteria.

**Table 4.** CHD detection results

An Algorithm	Patient Performance	ROC-AUC	F-score
Bispectrum	0.7207	0.7418	0.7244
Wavelet transformation	0.741	0.8	0.6967
Wavelet transformation +word2vec	0.7501	0.8	0.6990
R-peak's neighborhood	0.7602	0.7988	0.703
The HRV signal	0.763	0.744	0.662
3 different feature spaces	0.7632	0.8042	0.70256
majority	0.806		0.77
EM	0.8108	0.8738	0.7784

## 8. References

- [1] Gorbachev V V 2008 *Icardiac ischemia* (Minsk: Vysh. shk) p 479 (in Russian)
- [2] Dua Sumeet 2012 Novel classification of coronary artery disease using heart rate variability analysis *Journal of Mechanics in Medicine and Biology* **12(4)** 1240017-1240019
- [3] Giri D 2013 Automated diagnosis of coronary artery disease affected patients using LDA, PCA, ICA and discrete wavelet transform *Knowledge-Based Systems* **37** 274-282
- [4] Acharya U 2017 Rajendra et al. Application of higher-order spectra for the characterization of coronary artery disease using electrocardiogram signals *Biomedical Signal Processing and Control* **31** 31-43
- [5] Kumar M R B U and Rajendra A 2017 Characterization of coronary artery disease using flexible analytic wavelet transform applied on ECG signals *Biomedical Signal Processing and Control* **31** 301-308
- [6] Rangayyan R M 2015 *Biomedical Signal Analysis* (John Wiley & Sons)
- [7] Rawald T M and Sips N M 2017 PyRQA – Conducting recurrence quantification analysis on very long time series efficiently *Computers & Geosciences* **104** 101-108
- [8] Kantelhardt J W 2001 Detecting long-range correlations with detrended fluctuation analysis *Physica A: Statistical Mechanics and its Applications* **295(3-4)** 441-454
- [9] Tianqi C and Guestrin C 2016 Xgboost: a scalable tree boosting system *Proc. of the 22nd acm sigkdd international conference on knowledge discovery and data mining ACM*
- [10] Hjorth B 1970 EEG analysis based on time domain properties *Electroencephalography and Clinical Neurophysiology* **29(3)** 306-310
- [11] De Cooman T, Carrette E, Boon P, Meurs A and Van Huffel S 2014 September Online seizure detection in adults with temporal lobe epilepsy using single-lead ECG *Proceedings of the 22nd European Signal Processing Conference* 1532-1536
- [12] Uspensky V 2008 *Theory and practice of diagnosis of diseases of internal organs by the method of information analysis of electrocardio signals* (Moscow: Economics and Informatics) p 116 (in Russian)

- [13] Pedregosa F 2011 Scikit-learn: machine learning in Python *Journal of machine learning research* 2825-2830
- [14] Ripoll V J R 2016 ECG assessment based on neural networks with pretraining *Applied Soft Computing* **49** 399-406
- [15] Al-Rfou R, Alain G, Almahairi A, Angermueller C, Bahdanau D and Ballas N 2016 Theano: a Python framework for fast computation of mathematical expressions *Preprint arXiv:1605.02688*
- [16] Dieleman S 2015 *Lasagne: First release* (Geneva, Switzerland: Zenodo)
- [17] Wang J, Liu P, She M F, Nahavandi S and Kouzani A 2013 Bag-of-words representation for biomedical time series classification *Biomedical Signal Processing and Control* **8(6)** 634-644
- [18] Liu C L 2010 *A tutorial of the wavelet transform* (Taiwan: NTUEE)
- [19] Mikolov T 2013 *Efficient estimation of word representations in vector space* (Scottsdale, Arizona: ICLR Workshop)
- [20] Rehurek R and Sojka P 2010 Software framework for topic modelling with large corpora *Proc. of the LREC Workshop on New Challenges for NLP Frameworks*
- [21] Civera M L Z and Surace S 2016 Using bispectral analysis and neural networks to localise cracks in beam-like structures *Proc. of the 8th European Workshop On Structural Health Monitoring* 1542-1551
- [22] Chrysostomos N, Mysore L and Raghuveer R 1987 Bispectrum estimation: A digital signal processing framework *Proc. of the IEEE* **75(7)** 869-891
- [23] Al-Fahoum A, Al-Fraihat A and Al-Araida A 2014 Detection of cardiac ischaemia using bispectral analysis approach *Journal of medical engineering & technology* **38(6)** 311-316
- [24] Dawid A and Skene A 1979 Maximum likelihood estimation of observer error-rates using the EM algorithm *Applied statistics* 20-28
- [25] Sokolova M, Japkowicz N and Szpakowicz S 2006 Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation *Australian conference on artificial intelligence* **4304** 1015-1021