Multi-type Arrhythmia Classification: Assessment of the Potential of Time and Frequency Domain Features and Different Classifiers

Irena Jekova[1*], Giovanni Bortolan[2], Todor Stoyanov[1], Ivan Dotsinsky[1]

[1]*Institute of Biophysics and Biomedical Engineering*
Bulgarian Academy of Sciences
Acad G. Bonchev Str., Bl. 105, 1113 Sofia, Bulgaria
E-mails: irena@biomed.bas.bg, tstoyanov72@gmail.com, iadoc34@gmail.com

[2]*Institute of Neuroscience*
IN-CNR, Padova, Italy
E-mail: giovanni.bortolan@cnr.it

[*]*Corresponding author*

Received: December 11, 2019 **Accepted: February 28, 2020**

Published: June 30, 2020

Abstract: *Atrial fibrillation (AF) is associated with significant risk of heart failure and consequent death. Its episodic appearance, the wide variety of arrhythmias exhibiting irregular AF-like RR intervals and noises accompanying the ECG acquisition, impede the reliable AF detection. Therefore, the Computing in Cardiology Challenge 2017 organizers encourage the development of methods for classification of short, single-lead ECG as AF, normal sinus rhythm (NSR), other rhythm (OR), or noisy signal (NOISE). This study presents a set of 118 time and frequency domain feature including descriptors of the RR and PP intervals; QRS and P-wave amplitudes; ECG behavior within the TQ intervals, deviation of the TQ and PQRST segments from their first principle component analysis vector; dominant frequency; regularity index, width and area of the power spectrum estimated for the ECG signal with eliminated QRS complexes. Three classification techniques have been applied over the 118 ECG features – linear discriminant analysis (LDA), classification tree (CT) and neural network (NN) approach. The scores over a test subset are: (i) F_{NSR} = 0.81; F_{AF} = 0.61; F_{OR} = 0.53, F1 = 0.65 for CT, which is the most simple model; (ii) F_{NSR} = 0.82; F_{AF} = 0.62; F_{OR} = 0.53, F1 = 0.66 for LDA, which is the model with the most reproducible accuracy results; (iii) F_{NSR} = 0.86; F_{AF} = 0.74; F_{OR} = 0.57, F1 = 0.72 for NN, which is the most accurate model.*

Keywords: *Normal sinus rhythm, Atrial fibrillation, Other rhythm, Noise, Neural network, Classification tree, Linear discriminant analysis.*

Introduction

Atrial fibrillation (AF) is the most common sustained cardiac tachyarrhythmia, with incidence increasing from 0.5% at age of 40-50 years up to 5-15% for 80 years old people [20]. It is characterized by uncoordinated atrial activation and deterioration of atrial mechanical function, associated with significant risk of heart failure and consequent death [11, 21, 27].

There are three general approaches for AF detection:

- Atrial activity analysis associated with investigation of the TQ interval for presence of multiple P-waves [8, 10] or absence of P-waves [4, 15];

- Ventricular response analysis associated with RR intervals investigation via assessment of their median absolute deviation [18], irregularity [23], sample entropy [16], etc.;
- Combination of independent data from the atrial and ventricular contractions analyses [2, 24].

A comparative study of AF detection methods [17] highlights the techniques based on analysis of RR interval irregularity as the most robust against noise, providing the highest sensitivity and specificity, while the combination of RR and atrial activity analysis assures the highest positive predictive value.

The ECG analyses for AF detection are performed either in the time [2, 5, 6, 8, 15, 16, 18, 23, 24] or in the frequency domain, where the dominant AF frequency is usually assessed over a signal with extracted QRS-T complexes [26, 29].

Considering the episodic appearance of AF, the wide variety of arrhythmias exhibiting irregular AF-like RR intervals and the diverse noises accompanying the ECG acquisition, the organizers of the Computing in Cardiology Challenge 2017 encouraged the promotion of methods for classification of short single lead ECG as AF, normal sinus rhythm (NSR), other rhythm (OR), or noisy signal (NOISE). A total of 75 independent teams entered the Challenge by applying a variety of traditional and novel methods for the solution of this multi-type arrhythmia classification problem [7]. Fifteen of these studies (i.e., 20%) were included in the dedicated to the Challenge special issue of Physiological Measurements. The presented decisions were based on assessment of morphological features [1, 3, 6, 12, 25, 31-33]; rhythm descriptors [3, 6, 19, 31-33]; frequency domain parameters [12, 19, 28] and statistical characteristics [1, 19, 30]. The applied rhythm classifiers were discriminant analysis [6, 28]; decision trees [3, 13, 30, 31, 33]; support vector machines [1, 12, 19, 31] and neural networks [22, 25, 28, 32-35].

The aim of this study is to assess and present the potential of a significant number of time and frequency domain features for discrimination between AF, NSR, OR and NOISE and to compare the performance of different decision making approaches.

ECG databases

The ECG signals used for training and validation of the designed methods are from the PhysioNet/CinC Challenge 2017 database. It contains single-lead ECGs recorded via AliveCor device at 300 Hz sampling rate. The recordings are separated to two independent datasets:

- Training dataset, which contains 8528 ECGs annotated in four classes according to the rhythm type and signal quality:
    - 5076 NSR signals, which are 59.5% of the training dataset;
    - 758 AFs, which are 8.9% of the training dataset;
    - 2415 ORs, which are 28.3% of the training dataset;
    - 279 NOISE signals, which are 3.3% of the training dataset.

The duration of the ECG recordings is in the range (9-61 s), approximately the same for all class annotation.

- Test dataset, containing 3658 ECGs unavailable to the public, including:
    - 2437 NSR signals, which are 66.6 % of the training dataset;
    - 286 AFs, which are 7.8% of the training dataset;
    - 683 ORs, which are 18.7% of the training dataset;
    - 252 NOISE signals, which are 6.9% of the training dataset.

The test dataset has been used by the Challenge organizers for scoring purposes.

There are three versions of the annotations of the PhysioNet/CinC Challenge 2017 database and the listed above numbers correspond to the third version [7].

Method

The module for discrimination between NSR, AF, OR and NOISE in a single-lead ECG is developed in Matlab (MathWorks Inc.) and is presented in Fig. 1. It implements a preprocessing stage, feature extraction procedures and rhythm classification block embedding classification tree (CT), linear discriminant analysis (LDA) or neural network classifier (NN).

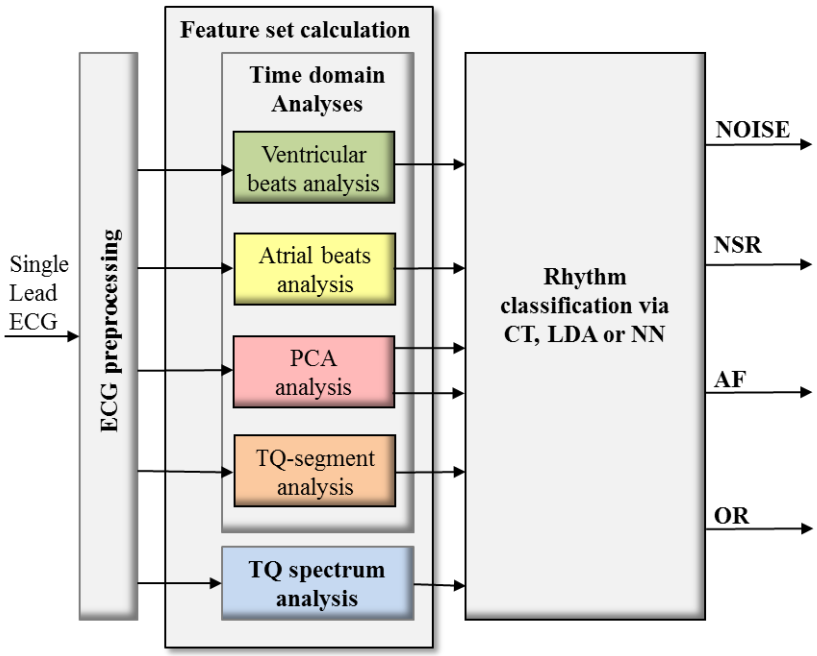


Fig. 1 Flow chart of the module for discrimination between AF, NSR, OR and NOISE

ECG preprocessing

The ECG preprocessing includes high-pass filter with cut-off frequency 1 Hz for baseline wander elimination and a comb filter with first zeroes at 50 Hz for power-line interference suppression.

Feature extraction procedures

Aiming to estimate the differences between the four analyzed classes we elaborated feature extraction procedures for calculation of a redundant feature set based on QRS and P-wave detection and evaluation; Principal Component Analysis (PCA) over the PQRST and TQ segments, and TQ-segment processing in the time and frequency domain.

Ventricular beats analysis

The time domain analysis starts with ventricular beats (VB) detection and classification based on heuristic algorithm including two main criteria:

1. Real-time evaluation of steep edges and sharp peaks for detection of normal QRS complexes;
2. Identification of ectopic beats via RR interval and waveform analysis.

Detailed description of the procedure for ventricular beats detection and classification could be found in [9]. The stability of the detected ventricular contractions is assessed via calculation of:

- 16 amplitude and temporal features describing the heartbeats and the underlying ECG rhythm, including:
  - Statistical descriptors of the amplitudes of all VBs (4 *features*) – mean (MeanAmpVB), minimal (MinAmpVB), maximal (MaxAmpVB) values and standard deviation (StdAmpVB). Higher StdAmpVB is expected for OR and NOISE groups. Higher value of MaxAmpVB is presumed for some NOISE signals.
  - Statistical descriptors of the amplitudes of the normal beats (4 *features*) – mean (MeanAmpN), minimal (MinAmpN), maximal (MaxAmpN) values and standard deviation (StdAmpN). Higher StdAmpN would give a hint for erroneous heartbeat detection due to existence of noise.
  - Statistical descriptors of the RR-intervals of all VBs (4 *features*) – mean (MeanRRVB), minimal (MinRRVB), maximal (MaxRRVB) values and standard deviation (StdRRVB). Lower value of MeanRRVB would suggest for possible tachycardia (part of the OR class), while higher StdRRVB is a sign for either irregular ECG rhythm (OR) or existence of noise that disturbs the correct heartbeat detection.
  - Statistical descriptors of the RR-intervals of all N-beats (4 *features*) – mean (MeanRRN), minimal (MinRRN), maximal (MaxRRN) values and standard deviation (StdRRN). Again lower value of MeanRRN is expected for OR signals and higher StdRRN for ECGs annotated as NOISE.
- N beats ratio (NBeats) (1 *feature*), expected to present higher values for NSR compared to OR:

$$\text{NBeats} = 100\,\text{Number}_{\text{N}}/\text{Number}_{\text{VB}}, (\%)\,.$$

- Probability the rhythm to be AF based on assessment of the RR irregularity (1 *feature*). Based on our observations, the following values are assigned to AF:

$$\text{AF, (\%)} = \begin{cases} 80 & \text{if (MaxRR\_VB-MeanRR\_VB)} > 200 \text{ ms}, \\ 60 & \text{if } 200 \text{ ms} \geq \text{(MaxRR\_VB-MeanRR\_VB)} > 160 \text{ ms}, \\ 0 & \text{if (MaxRR\_VB-MeanRR\_VB)} < 80 \text{ ms}, \\ 40 & \text{otherwise.} \end{cases}$$

Atrial beats analysis

The P-waves are identified via the P-wave detector described in [6] which uses a synthesized P-wave template and applies modified convolution between its samples and moving signal intervals, thus searching for P-wave patterns.

The stability of the detected atrial contractions is assessed via calculation of 11 *features*:

- Statistical descriptors of the P-waves amplitudes (4 *features*), including mean (MeanAmpP), minimal (MinAmpP), maximal (MaxAmpP) values and standard deviation (StdAmpP) – higher value of StdAmpP is presumed for AF compared to NSR.
- Statistical descriptors of the intervals between consecutive P-waves (4 *features*), including mean (MeanPPint), minimal (MinPPint), maximal (MaxPPint) values and standard deviation (StdPPint) – lower MeanPPint and higher StdPPint are supposed for AF compared to NSR and OR.
- Mean value and standard deviation of the P-waves number in each RR interval (2 *features*): MeanPcountRRint, StdPcountRRint – higher values of both parameters are expected for AF compared to NSR.
- Percentage of RR intervals with two or more detected P-waves (1 *feature*) – DoubleP, (%).

PCA analysis

PCA is used for assessment of the ECG beat-to-beat irregularity. The PCA is applied over the PQRST and the TQ segments wrapped respectively between:

- PQRST: (QRSindex – 0.25 SR) and $(\text{QRSindex} + 0.5\text{SR}\sqrt{\text{MeanRR\_VB/SR}})$;
- TQ: (QRSindex – 0.8 MeanRR_VB) and (QRSindex – 100 ms), enclosing the T-wave, P-wave and atrial fibrillation waves (if present).

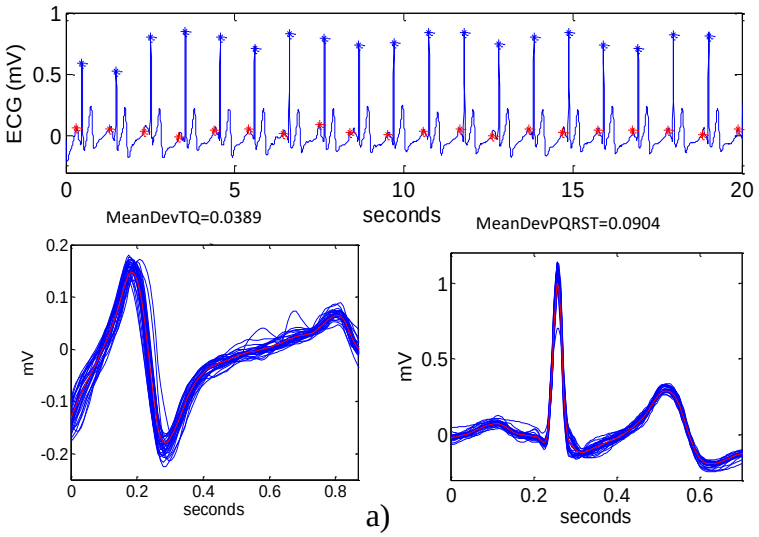
Two features representing the mean value of the deviation between the samples of the respective segments and the corresponding samples of its first PCA vector (Fig. 2) are calculated using the formula:

$$MeanDevECG\,int = \frac{\sum_{i=1}^{LenECG\,int}\sqrt{\sum_{j=1}^{NbVB}\frac{\left(ECG\,int(i,j)-PCA_1(i)\right)^2}{NbVB}}}{LenECG\,int},$$

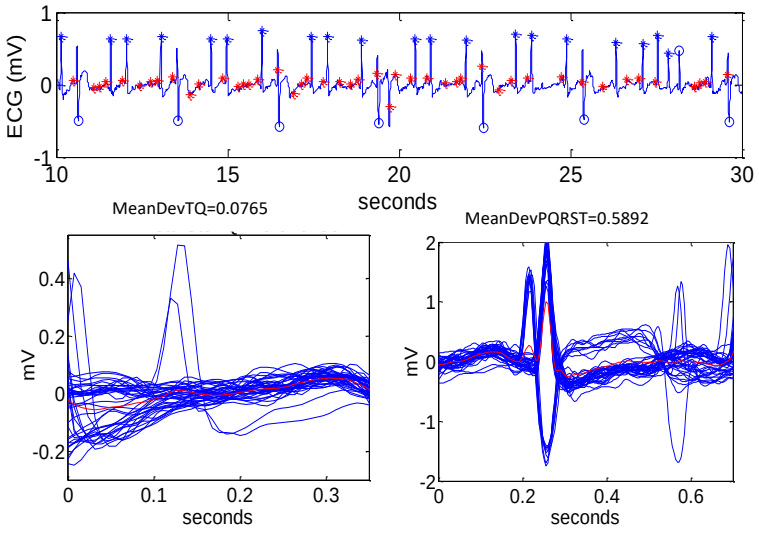
where *LenECGint* is the length of the analyzed ECG interval (PQRST or TQ), *NbVB* is the number of analyzed VBs, $ECG(i, j)$ is the i^{th} ECG sample of the j^{th} VB, $PCA_1(i)$ is the i^{th} sample of the first PCA vector.

The features are:

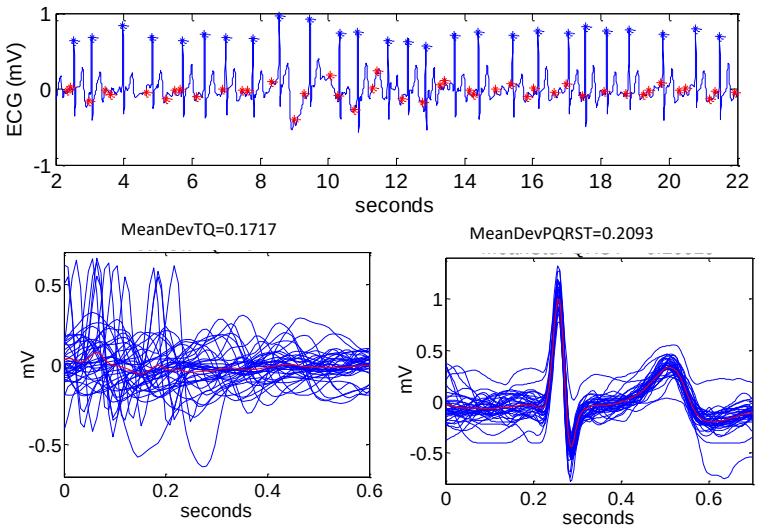
- *MeanDevPQRST* = *MeanDevECGint*, when *ECGint* is the PQRST segment. Before *MeanDevPQRST* calculation all PQRST segments are normalized towards the maximal absolute value of PQRST first PCA vector. This feature is supposed to be appropriate for NSR vs. OR discrimination, as well as for noise detection considering the expected PQRST waveform variability in OR and NOISE signals.
- *MeanDevTQ* = *MeanDevECGint*, when *ECGint* is the TQ segment. This feature is supposed to be representative for the NOISE signals.


ECG (mV)
seconds
MeanDevTQ=0.0389
MeanDevPQRST=0.0904
mV
seconds

a)


ECG (mV)
seconds
MeanDevTQ=0.0765
MeanDevPQRST=0.5892
mV
seconds

b)


ECG (mV)
seconds
MeanDevTQ=0.1717
MeanDevPQRST=0.2093
mV
seconds

c)

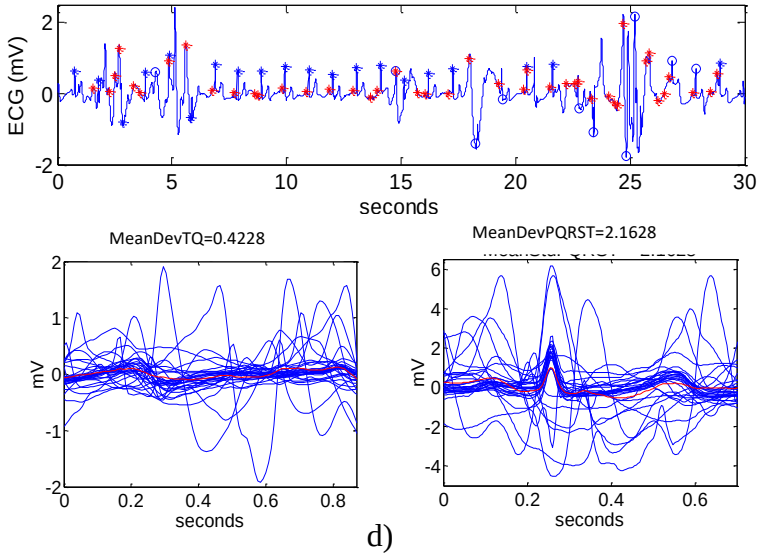


Fig. 2 Examples of: (a) NSR; (b) AF; (c) OR; (d) NOISE;
the respective TQ, PQRST segments (in blue) and TQ, PQRST first PCA vectors (in red).
The detected normal beats are marked with blue asterisks (*),
ectopic beats with blue circles (o), and P-waves with red asterisks (*).

TQ-segment analysis

Considering that the AF influences the waveforms appearing between the T-wave end and the Q-wave by mimicking ventricular fibrillation (VF) like patterns, two analysis procedures typical for VF detection are applied over the ECG signals after elimination of the QRS-T segments (i.e., over the TQ intervals) and the following features are calculated in the time domain:

- TQ-signal complexity (C) estimated via the following procedure:
  - Conversion of the ECG segment to a binary string 0/1 by comparison with a suitably selected threshold. The mean value of the ECG data points in the analyzed window is calculated and is subtracted from each ECG sample, thus generating a new signal – *ECGnew*. V_p and V_n are the positive and negative peak values in *ECGnew*, respectively. P_c and N_c are the number of *ECGnew* values in the interval $[0 < x_i < 0.1V_p]$ and $[0.1V_n < x_i < 0]$, respectively. If $(P_c + N_c) < 0.4n$ (n is the number of samples in the analyzed ECG window) the threshold is selected to be $T_d = 0$; else if $P_c < N_c$, $T_d = 0.2V_p$, otherwise $Td = 0.2V_n$. The ECG data is converted in a string by the following logical rule: if $ECGnew(i) < T_d$, $S_i = 0$, else $S_i = 1$.
  - Computation of the complexity measure by scanning the binary string S from left to right and increasing a counter $c(n) = c(n) + 1$ each time when a new subsequence of 0 and 1 is encountered.
  - Computation of the normalized $C = \frac{c(n)}{b(n)}$, where $b(n) = \frac{n}{\log_2^n}$ is the asymptotic behavior of $c(n)$ for a random string.
- Detailed description of the method for calculation of C could be found in [36]. C is calculated for the first, the middle and last 6 s segment of the TQ signal and the median value is further considered (1 *feature*). This feature is expected to present higher values for AF signals.

- TQ-signal leakage after application of a narrow band-stop filter over 3 s segments of the TQ-signal adjusted at its main frequency [14]. The TQ-signal is considered to be of quazi-sinusoidal in case of AF and its mean period is obtained from the equation:

$$T = 2\pi \frac{\sum_{i=1}^{len} |TQ_i|}{\sum_{i=1}^{len} |TQ_i - TQ_{i-1}|},$$

where $TQ(i)$ are the TQ-signal samples and *len* is the number of data points in the analyzed interval, i.e., 3 s.

The narrow band-stop filter is simulated by combining the TQ-signal data with a copy of the data shifted by half a period. The leakage is computed as:

$$Leak = \frac{\sum_{i=1}^{m} \left| TQ_i + TQ_{i-\frac{T}{2}} \right|}{\sum_{i=1}^{m} \left(|TQ_i| + \left| TQ_{i-\frac{T}{2}} \right| \right)}.$$

For assessment of the entire TQ-signal we consider the statistical descriptors of the *Leak* values (4 *features*) – mean (MeanLeakTQ), minimal (MinLeakTQ), maximal (MaxLeakTQ) values and standard deviation (StdLeakTQ), which are supposed to be low for AF signals.

The frequency domain analysis of the TQ-signal included calculation of its power spectrum (PS) over 4 s non-overlapping intervals and estimation of the following spectral features:

- Dominant frequency (DF), which is the frequency corresponding to the maximum PS power in the range (3-15) Hz [26]. The mean (MeanDF), minimal (MinDF), maximal (MaxDF) DF values and the DF standard deviation (StdDF) calculated over the entire TQ-signal are further considered (4 *features*).
- Regularity index (*RI*), calculated according to [26], which quantifies the sharpness of the dominant peak in PS:

$$RI = \frac{\sum_{i=DF-0.75Hz}^{DF+0.75Hz} PS(i)}{\sum_{j=3Hz}^{15Hz} PS(j)}.$$

The mean (MeanRI), minimal (MinRI), maximal (MaxRI) RI values and RI standard deviation over the entire TQ-signal are involved into analysis (4 *features*).

- Spectral width at nine different levels (0.1 to 0.9) of the maximum power in the range (3-15) Hz – mean (MeanSpecWidth_level), minimal (MinSpecWidth_level), maximal (MaxSpecWidth_level), and standard deviation (StdSpecWidth_level) values are measured, i.e., 9×4 *features*.
- Spectral area between the first and the last cross of nine different levels (0.1 to 0.9) of the maximum power in the range (3-15) Hz – again the mean (MeanAreaWidth_level),

minimal (MinAreaWidth_level), maximal (MaxAreaWidth_level), and standard deviation (StdAreaWidth_level) values are applied, i.e., 9×4 *features*.

Decision making

The decision making in this study is performed by applying three independent classification approaches over the feature set – LDA and CT during the official Challenge phase, and NN in the follow-up phase.

Linear discriminant analysis

The detection of NOISE is performed via a single threshold rule MeanDevTQ ≥ 0.5, which minimally influences the correct classification of the NSR, AF, OR rhythms (see the distribution in Fig. 3, section Results). The remaining three rhythm types are discriminated via stepwise analysis of LDA models.

Three linear discriminant functions are calculated:

$$F_{NSR} = \sum_{i=1}^{n} w_{NSR}(i) feature(i) + a_{NSR},$$

$$F_{AF} = \sum_{i=1}^{n} w_{AF}(i)\, feature(i) + a_{AF},$$

$$F_{OR} = \sum_{i=1}^{n} w_{OR}(i)\, feature(i) + a_{OR}.$$

Here F_{NSR}, F_{AF}, F_{OR} are related to the possibility the rhythm described by the feature vector to be NSR, AF, OR, respectively, while w_{NSR}, w_{AF}, w_{OR}, and a_{NSR}, a_{AF}, a_{OR} are the corresponding discriminant coefficients and constants. Stepwise linear discriminant analysis is applied for iterative selection of non-redundant feature set and the rhythm is classified according to the maximal discriminant function.

Classification tree

The noise detection is performed via a single threshold rule MeanDevTQ ≥ 0.5, which minimally influences the correct detection of the NSR, AF, OR rhythms (see the distribution in Fig. 3, section Results). The three rhythm types are discriminated via a CT model, generated and pruned by means of the statistical package Statistica (v. 12.3, Dell Inc.), using the following settings:

- Three categories of the classification variable according to the rhythm type: NSR, AF, OR;
- Splitting based on Gini index;
- Optimal prior probabilities for NSR vs. AF vs. OR – 0.4 vs. 0.2 vs. 0.4;
- Pruning criterion based on misclassification rate;
- Apriori probabilities – 0.4 for NSR and OR, 0.2 for AF.

Neural networks

Multilayer Neural Network architecture with Levenberg-Marquardt back propagation algorithm is applied. The NN's performance function was computed by mean squared normalized error.

In order to maintain the generalization property, the training procedure stops if the validation performance degrades for a number of 6 consecutive iterations.

Two architectures with two hidden layers were tested:

- A1 – 118 input nodes, first hidden layer with 8 nodes, second hidden layer with 4 nodes, and 4 output nodes (NSR, AF, OR, NOISE);
- A2 – 118 input nodes, first hidden layer with 16 nodes, second hidden layer with 4 nodes, and 4 output nodes (NSR, AF, OR, NOISE).

Two strategies have been used for improving the generalization property and the achieved accuracy:

- Introduction of duplicated patients in the learning procedure, in order to have homogeneous composition, and to assure similar consistency/weight of all the classes.
- Application of multiple NNs and classification based on a majority rule as the decision-making system (the most voted classification).

We designed four rhythm classifiers based on different number of NNs:

- 3 NN with A1 architecture;
- 6 NN with A1 architecture;
- 7 NN with A1 architecture;
- 6 ANN with A1 (3) and A2 (3) architecture.

The various NNs are chosen from a forest of NN with the choice of the best 3, 6 or 7 NNs based on their performance on the training database. An optimal version of majority rule in a plurality system was obtained by a simple algorithm – the sum of the 4 outputs of the considered group of NN's, which has the particular property of the presence of an implicit order. The rhythm with highest value corresponds to the output with highest confidence. For the majority of the ECGs, the different NN in the considered group show a high agreement in the classification. The 6 ANN model provided the best results over the training dataset and it is further considered. For 63% of records, the output of the 6 NNs produce the same classification, while in 92.3% of cases at least 4 of 6 NN perform the same classification.

Accuracy assessment

The comparison of feature distributions for each pair of ECG classes is done in Statistica (Dell Inc.) via Student's t-test. A value of $p \leq 0.05$ is considered statistically significant.

Evaluation of the Challenge entries is done by rhythm specific (F_{RHYTHM}) and common ($F1$) score, proposed by the Challenge organizers and calculated as:

$$F_{RHYTHM} = \frac{2TP_{RHYTHM}}{2TP_{RHYTHM} + FN_{RHYTHM} + FP_{RHYTHM}},$$

$$F1 = \frac{F_{NSR} + F_{AF} + F_{OR}}{3},$$

where *TP*, *FN*, *FP* stand for true positive, false negative, false positive detections.

A standard accuracy estimator is also calculated for the training dataset as follows:

$$A_{RHYTHM} = \frac{TP_{RHYTHM}}{TP_{RHYTHM} + FN_{RHYTHM}}.$$

Results

The examples in Fig. 2 illustrate the operation of the heartbeat detector and the P-wave detector together with the calculation of the PCA analysis features.

The applied t-test showed statistically significant difference between the distribution of each feature for at least two of the ECG classes.

The application of LDA, CT analysis and NN over the 118 features revealed their potential for multi-type arrhythmia classification. Table 1 represents the 15 features included in the CT, the first 15 features that were used by the stepwise LDA and 12 features indicated by the applied NNs. We tested the weights (absolute values) in the input layers of the 6 ANN model, and we considered the sum of absolute values of the 8 weights for every feature in A1 architecture, and the sum of absolute values of the 16 weights for every feature in A2 architecture. After sorting the weights' sums for the 3 NN with A1 architecture and 3 NN with A2 architecture, the common features among the first 20 sorted for each NN are included in the table. Statistical distribution of the features is presented in Figs. 3-7.

Table 1. Features highlighted by the LDA, CT and NN classifiers

	Time-domain features				Frequency-domain features
Model	**Ventricular beats**	**Atrial beats**	**PCA analysis**	**TQ analysis**	
LDA	AF, (%) NBeats, (%) MinRRN MeanRRVB MaxAmpVB	MeanAmpP MeanPPint StdPPint DoubleP, (%) StdPcount RRint	MeanDevTQ MeanDevPQRST	C	StdRI MinSpecArea_04
CT	AF, (%) NBeats, (%) MeanRRN MinRRN StdRRVB StdRRN	MeanPPint	MeanDevPQRST MeanDevTQ	MeanLeakTQ MinLeakTQ MeanPeriodTQ C	MeanDF StdRI
6 ANN	MeanRRVB MinRRVB MaxRRVB MeanRRN MinRRN	MeanPPint StdPPint MeanPcount RRint MeanAmpP	MeanDevTQ	C	StdSpecArea_09

Three multi-type arrhythmia classification models for discrimination between NSR, AF, OR and NOISE were trained and submitted for independent evaluation on the hidden test dataset:

- CT model based on the 15 features listed in Table 1;
- LDA model based on 57 features as follows:

- 15 time-domain features describing the ventricular beats behavior, including all statistical descriptors of the RR-intervals of VBs and N-beats (8 *features*); all statistical descriptors of the amplitudes of VBs (*4 features*); MinAmpN; Nbeats, (%); AF, (%);
- 10 time-domain features describing the atrial beats behavior, including all statistical descriptors of the intervals between consecutive P-waves (*4 features*); MeanAmpP; MinAmpP; StdAmpP; MeanPcountRRint; StdPcountRRint; DoubleP, (%);
- 5 TQ-signal features, including C; MeanLeakTQ; MinLeakTQ; MeanPeriodTQ;
- 2 PCA features – MeanDevTQ; MeanDevPQRST;
- 25 frequency-domain features – 10 representing the TQ-signal spectral width at different levels; 13 characterizing the TQ-signal spectral area between the first and the last cross of different levels; MeanDF, StdRI.

- 6 ANN model based on all 118 features.

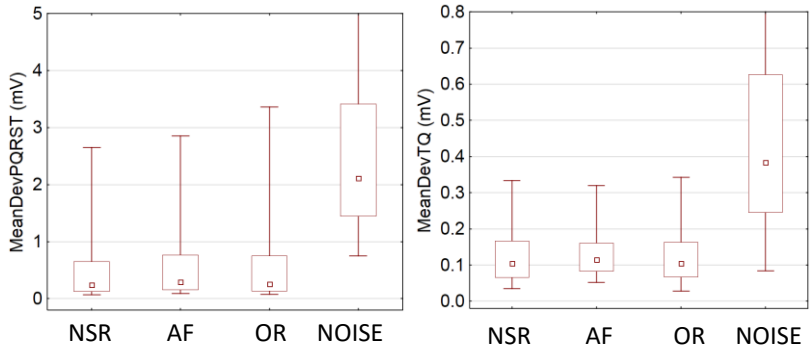


Fig. 3 Quartile range (box) and (5-95)% range (whiskers) of the PCA analysis features listed in Table 1

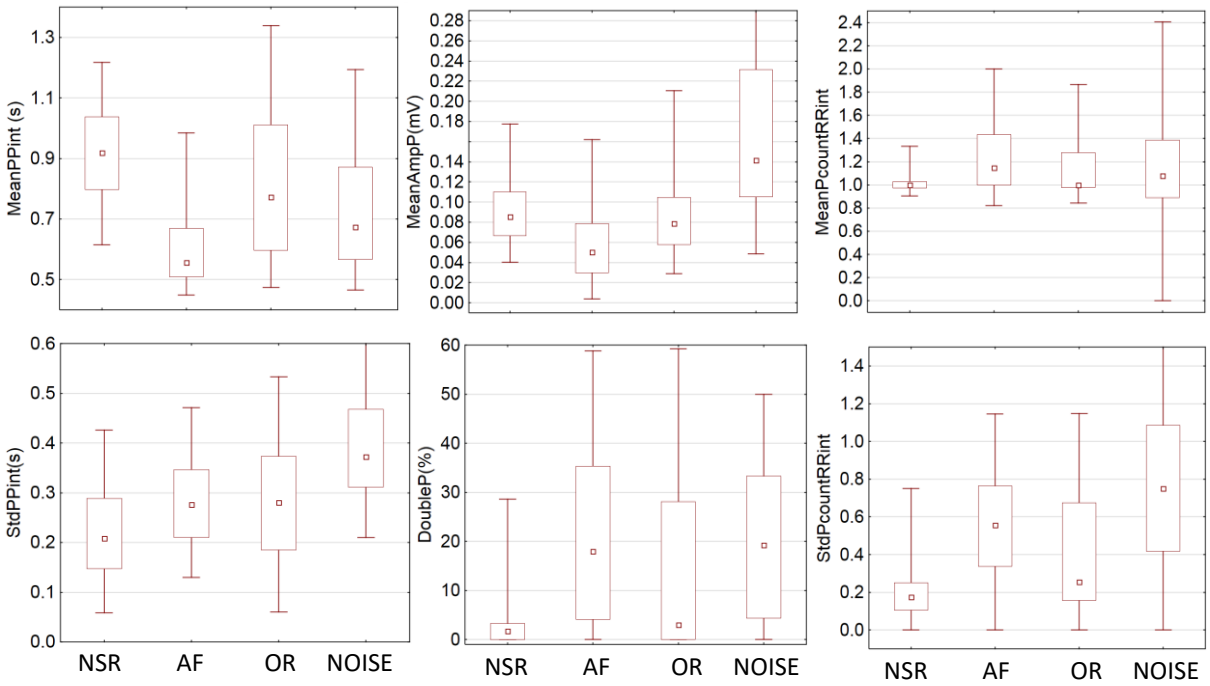


Fig. 4 Quartile range (box) and (5-95)% range (whiskers) of the time-domain atrial beat features listed in Table 1

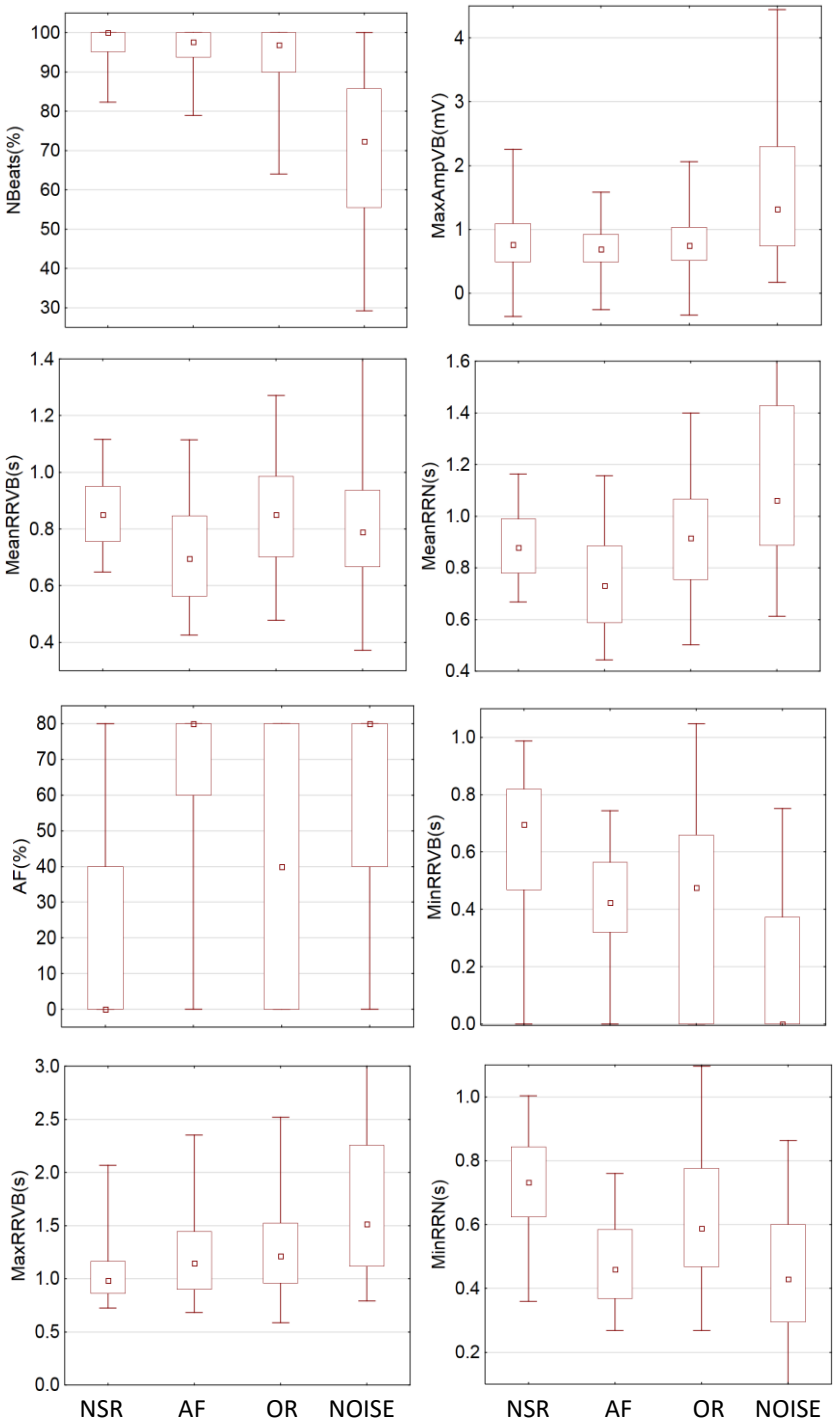


Fig. 5 Quartile range (box) and (5-95)% range (whiskers) of the time-domain ventricular beat features listed in Table 1

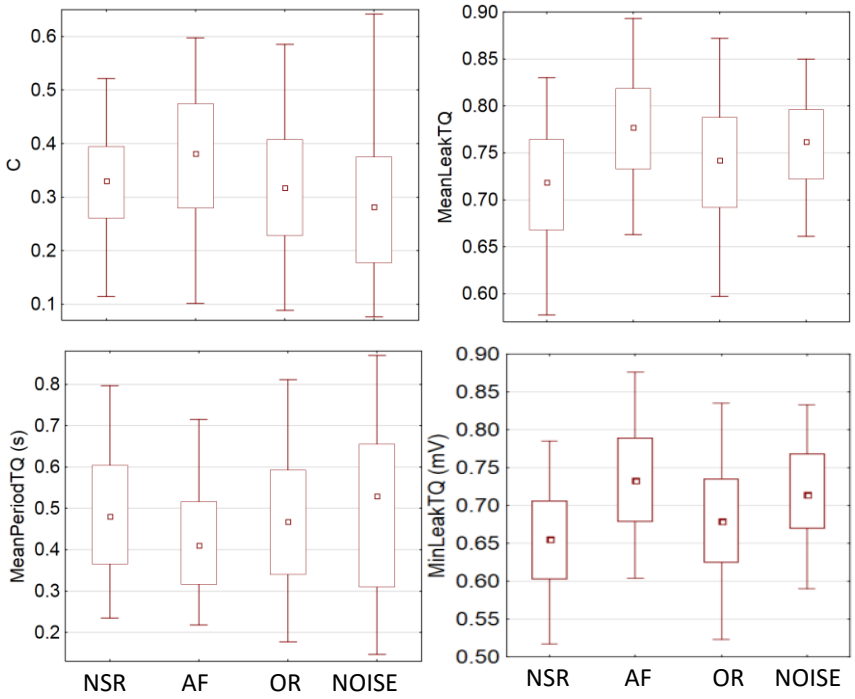


Fig. 6 Quartile range (box) and (5-95)% range (whiskers) of the time-domain TQ features listed in Table 1

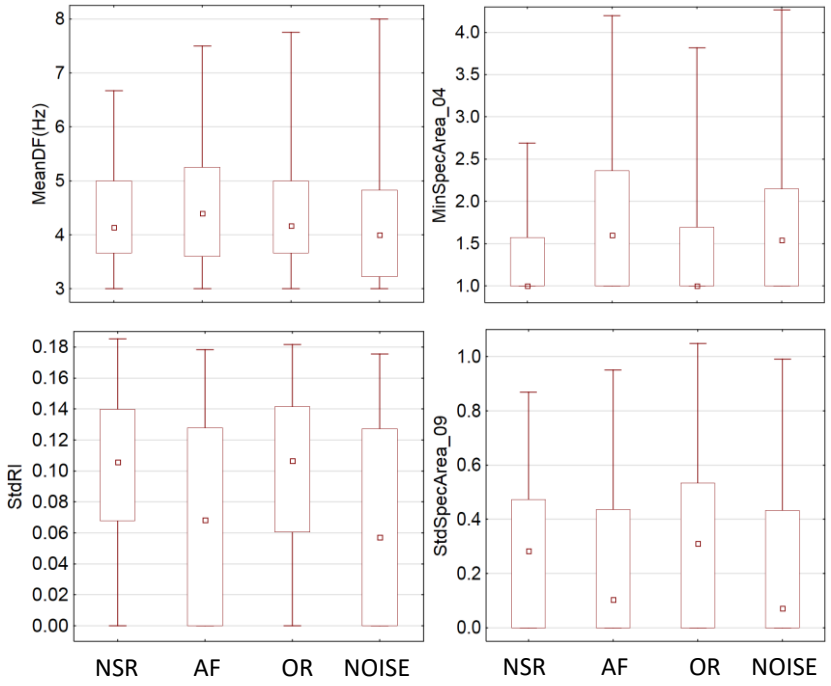


Fig. 7 Quartile range (box) and (5-95)% range (whiskers) of the frequency-domain features listed in Table 1

Table 2 presents the training accuracy for each of the 3 arrhythmia classification models. The Challenge scores achieved on the training and the hidden test dataset are presented in Table 3. The test results are shown for the total hidden test set (where available) and for a subset of the test set (27%). The last is done for two reasons: (i) only two of our entries are scored over the total test set by the Challenge organizers; (ii) for comparative purposes, since the competitors' scores in [7] are reported only for the 27% subset of the test set.

Table 2. Accuracy for classification of NSR, AF, OR, NOISE by LDA, CT and 6 ANN achieved during the training process

	LDA	CT	6 ANN
A_{NSR}	82.2%	78.0%	85.1%
A_{AF}	59.8%	60.3%	75.1%
A_{OR}	52.5%	66.1%	69.7%
A_{NOISE}	20.0%	35.8%	87.3%

Table 3. Scores for classification of NSR, AF, OR, NOISE and common score achieved by LDA, CT and 6ANN on the training dataset and the hidden test dataset

	LDA		CT			6 ANN		
	Train	Test (27%)	Train	Test (27%)	Total test	Train	Test (27%)	Total test
F_{NSR}	0.82	0.82	0.82	0.81	0.83	0.86	0.86	0.84
F_{AF}	0.60	0.62	0.62	0.61	0.60	0.72	0.74	0.66
F_{OR}	0.56	0.53	0.61	0.53	0.51	0.70	0.57	0.54
F_{NOISE}	0.30	NA	0.38	NA	0.36	0.73	NA	NA
$F1$	0.66	0.66	0.68	0.65	0.64	0.76	0.72	0.68

NA – not available

The Challenge organizers assessed average running time 10.3%, 10.8%, 12.9% of the quota for CT, LDA, 6 ANN, respectively.

Discussion and conclusions

This study is dedicated to the classification of ECG signals in four classes – NSR, AF, OR and NOISE. It investigates the potential application of time-domain and frequency-domain features as well as the reliability of three different classifiers. We tried to answer two general questions:

1) Which features are suitable for solving this multi-type arrthythmia classification problem?

   Although the applied t-test showed statistically significant difference between the distributions of each feature for at least two of the ECG classes, for the four ECG types subjected to classification all features present considerably overlapping distributions (see the examples in Figs. 3-7). The time-domain features, especially those describing the ventricular and atrial beats (Figs. 4 and 5), show distinguishable behavior for NSR vs. AF – longer RR and PP intervals, higher P-wave amplitudes, smaller deviation of the PP intervals and the number of P-waves within one RR interval, lower DoubleP, (%), AF, (%) and MeanPcountRRint for NSR. The general complication comes from the fact that the values of these features calculated for the OR class, which contains wide variety of arrhythmia types, are similar to those typical either for NSR or for AF. The distributions of the PCA analysis features (Fig. 3) suggest their potential for detection of noisy ECG signals. Despite the great overlapping of the TQ-signal features and the frequency-domain features for all

classes, MeanLeakTQ, MinLeakTQ, MeanPeriodTQ, C, MeanDF, StdRI are included in the CT model; C, StdRI, MinSpecArea_04 are among the first involved by the stepwise LDA; and C, StdSpecArea_09 have one of the highest weights in the input layers of the 6ANN model.

2) Which is the most appropriate classifier for such multi-type arrthythmia classification problem?

   The answer of this question is not unique, since it depends on several criteria, such as accuracy, complexity and reproducibility of the results. The best accuracy results (Tables 2 and 3) are achieved for the 6 ANN model. The complexity criterion indisputably highlights the CT model as the simplest one based on only 15 *features*. However, these two models show statistically significant score drop of 8% points (6 ANN) and 4% points (CT) between training and test performance. A possible reason for this drop could be the considerable changes in the test dataset annotations during the Challenge, which was accompanied with negligible changes in the training set annotations (v2 vs. v3, described in [7]). This inevitably leads to training on data which is not annotated by applying the criteria used for the test data labeling. LDA is the model with the best reproducibility showing equal scores for the training and the subset (27%) of the test dataset.

The performance of all algorithms submitted to the Challenge is summarized in [7]. They provide scores between 0.86 and 0.20 on the 27% subset of the hidden test dataset for more than 120 Challenge entries. The scores show gradual decrease from 0.86 down to 0.5 for the first half of the ranked entries. Considering these officially announced Challenge algorithm performances our 6 ANN model, presenting score of 0.72, falls within the first 25 entries.

Despite some limitations of the Challenge competition, such as selection of not the most appropriate screening metric $F1$ and the disadvantages in the datasets annotations [7], the elaborated algorithms are a step forward in solving the non-trivial problem for multi-type arrhythmia classification. In this respect, this study demonstrates the potential of time and frequency domain features for discrimination between NSR, AF, OR and noisy ECGs as well as provides useful information about the applicability of CT, LDA and NN classification approaches in terms of simplicity, reproducibility and accuracy.

References

1. Athif M., P. Yasawardene, C. Daluwatte (2018). Detecting Atrial Fibrillation from Short Single Lead ECGs Using Statistical and Morphological Features, Physiological Measurement, 39(6), 064002.
2. Babaeizadeh S., R. Gregg, E. Helfenbein, J. Lindauer, S. Zhou (2009). Improvements in Atrial Fibrillation Detection for Real-time Monitoring, Journal of Electrocardiology, 42(6), 522-526.
3. Chen Y., X. Wang, Y. Jung, V. Abedi, R. Zand, M. Bikak, M. Adibuzzaman (2018). Classification of Short Single-lead Electrocardiograms (ECGs) for Atrial Fibrillation Detection Using Piecewise Linear Spline and XGBoost, Physiological Measurement, 39(6), 104006.
4. Christov I., G. Bortolan, I. Daskalov (2001). Sequential Analysis for Automatic Detection of Atrial Fibrillation and Flutter, IEEE Computers in Cardiology, 28, 293-296.

5. Christov I., V. Krasteva, I. Simova, T. Neycheva, R. Schmid (2017). Multi-parametric Analyses for Atrial Fibrillation Classification in ECG, Computing in Cardiology, 44, doi: 10.22489/CinC.2017.175-021.
6. Christov I., V. Krasteva, I. Simova, T. Neycheva, R. Schmid (2018). Ranking of the Most Reliable Beat Morphology and Heart Rate Variability Features for the Detection of Atrial Fibrillation in Short Single-lead ECG, Physiological Measurement, 39(6), 094005.
7. Clifford G., C. Liu, B. Moody, L. W. Lehman, I. Silva, Q. Li, A. E. Johnson, R. G. Mark (2017). AF Classification from a Short Single Lead ECG Recording: the Physionet Computing in Cardiology Challenge 2017, Computing in Cardiology, 44, doi:10.22489/CinC.2017.065-469.
8. Dotsinsky I. (2007). Atrial Wave Detection Algorithm for Discovery of Some Rhythm Abnormalities, Physiological Measurement, 28, 595-610.
9. Dotsinsky I., T. Stoyanov (2004). Ventricular Beat Detection in Single Channel Electrocardiograms, BioMedical Engineering Online, 3:3, doi: 10.1186/1475-925X-3-3.
10. Du X., N. Rao, M. Qian, D. Liu, J. Li, W. Feng, L. Yin, X. Chen (2014). A Novel Method for Real-time Atrial Fibrillation Detection in Electrocardiograms Using Multiple Parameters, Annals of Noninvasive Electrocardiology, 19(3), 217-225.
11. Fuster V., L. Ryden, R. Asinger, D. Cannom, H. Crijns, R. Frye, et al. (2001). ACC/AHA/ESC Guidelines for the Management of Patients with Atrial Fibrillation: Executive Summary, Journal of the American College of Cardiology, 38(4), 1231-1265.
12. Gliner V, Y. Yaniv (2018). An SVM Approach for Identifying Atrial Fibrillation, Physiological Measurement, 39(6), 094007.
13. Kropf M., D. Hayn, D. Morris, A. K. Radhakrishnan, E. Belyavskiy, A. Frydas, E. Pieske-Kraigher, B. Pieske, G. Schreier (2018). Cardiac Anomaly Detection Based on Time and Frequency Domain Features Using Tree-based Classifiers, Physiological Measurement, 39(6), 114001.
14. Kuo S., R. Dillman (1978). Computer Detection of Ventricular Fibrillation, Computing in Cardiology, 347-349.
15. Ladavich S., B. Ghoraani (2015). Rate-independent Detection of Atrial Fibrillation by Statistical Modeling of Atrial Activity, Biomed Signal Process Control, 18(4), 274-281.
16. Lake D. E., J. R. Moorman (2011). Accurate Estimation of Entropy in Very Short Physiological Time Series: The Problem of Atrial Fibrillation Detection in Implanted Ventricular Devices, American Journal of Physiology-Heart and Circulatory Physiology, 300(1), H319-H325.
17. Larburu N., T. Lopetegi, I. Romero (2011). Comparative Study of Algorithms for Atrial Fibrillation Detection, Computing in Cardiology, 38, 265-268.
18. Linker D. (2016). Accurate, Automated Detection of Atrial Fibrillation in Ambulatory Recordings, Cardiovascular Engineering and Technology, 7(2), 182-189.
19. Liu N., M. Sun, L. Wang, W. Zhou, H. Dang, X. Zhou (2018). A Support Vector Machine Approach for AF Classification from a Short Single-lead ECG Recording, Physiological Measurement, 39(6), 064004.
20. Naccarelli G., H. Varker, J. Lin, K. Schulman (2009). Increasing Prevalence of Atrial Fibrillation and Flutter in the United States, American Journal of Cardiology, 104(11), 1534-1539.
21. Naydenov S., N. Runev, E. Manov, D. Vasileva, Y. Rangelov, N. Naydenova (2018). Risk Factors, Co-morbidities and Treatment of In-hospital Patients with Atrial Fibrillation in Bulgaria, Medicina (Kaunas), 54(3), 34, doi: 10.3390/medicina 54030034.
22. Parvaneh S., J. Rubin, A. Rahman, B. Conroy, S. Babaeizadeh (2018). Analyzing Single-lead Short ECG Recordings Using Dense Convolutional Neural Networks and

Feature-based Post-processing to Detect Atrial Fibrillation, Physiological Measurement, 39(6), 084003.
23. Petrėnas A., V. Marozas, L. Sörnmo (2015). Low-complexity Detection of Atrial Fibrillation in Continuous Long-term Monitoring, Computers in Biology and Medicine, 65(10), 184-191.
24. Petrėnas A., L. Sörnmo, A. Lukoševicius, V. Marozas (2015). Detection of Occult Paroxysmal Atrial Fibrillation, Medical & Biological Engineering & Computing, 53(4), 287-297.
25. Plesinger F., P. Nejedly, I. Viscor, J. Halamek, P. Jurak (2018). Parallel Use of a Convolutional Neural Network and Bagged Tree Ensemble for the Classification of Holter ECG, Physiological Measurement, 39(6), 094002.
26. Roonizi E., R. Sassi (2016). Dominant Atrial Fibrillatory Frequency Estimation Using an Extended Kalman Smoother, Computing in Cardiology, 43, doi:10.22489/CinC.2016.286-359.
27. Runev N., S. Dimitrov (2015). Management of Stroke Prevention in Bulgarian Patients with Non-valvular Atrial Fibrillation (BUL-AF Survey), International Journal of Cardiology, 187, 683-685.
28. Sadr N., M. Jayawardhana, T. Pham, R. Tang, A. Balaei, P. de Chazal (2018). A Low-complexity Algorithm for Detection of Atrial Fibrillation Using an ECG, Physiological Measurement, 39(6), 064003.
29. Sasaki N., Y. Okumura, I. Watanabe, A. Madry, Y. Hamano, M. Nikaido, et al. (2015). Frequency Analysis of Atrial Fibrillation from the Specific ECG Leads V7-V9: A Lower DF in Lead V9 is a Marker of Potential Atrial Remodeling, Journal of Cardiology, 66, 388-394.
30. Shao M., G. Bin, S. Wu, G. Bin, J. Huang, Z. Zhou (2018). Detection of Atrial Fibrillation from ECG Recordings Using Decision Tree Ensemble with Multi-level Features, Physiological Measurement, 39(6), 094008.
31. Smisek R., J. Hejc, M. Ronzhina, A. Nemcova, L. Marsanova, J. Kolarova, L. Smital, M. Vitek (2018). Multi-stage SVM Approach for Cardiac Arrhythmias Detection in Short Single-lead ECG Recorded by a Wearable Device, Physiological Measurement, 39(6), 094003.
32. Sodmann P., M. Vollmer, N. Nath, L. Kaderali (2018). A Convolutional Neural Network for ECG Annotation as the Basis for Classification of Cardiac Rhythms, Physiological Measurement, 39(6), 104005.
33. Teijeiro T., C. García, D. Castro, P. Félix (2018). Abductive Reasoning as a Basis to Reproduce Expert Criteria in ECG Atrial Fibrillation Identification, Physiological Measurement, 39(6), 084006.
34. Warrick P., M. Homsi (2018). Ensembling Convolutional and Long Short-term Memory Networks for Electrocardiogram Arrhythmia Detection, Physiological Measurement, 39(6), 114002.
35. Xiong Z., M. Nash, E. Cheng, V. Fedorov, M. Stiles, J. Zhao (2018). ECG Signal Classification for the Detection of Cardiac Arrhythmias Using a Convolutional Recurrent Neural Network, Physiological Measurement, 39(6), 094006.
36. Zhang X., Y. Zhu, N. Thakor, Z. Wang (1999). Detecting Ventricular Tachycardia and Fibrillation by Complexity Measure, IEEE Transactions on Biomedical Engineering, 46, 548-555.

Prof. Irena Jekova, Ph.D.
E-mail: irena@biomed.bas.bg

Irena Jekova graduated Technical University – Sofia, Faculty of Electronic Engineering and Technology, specialization Electronic Medical Equipment in 1998. In the period 1999-2010 she worked in the Centre of Biomedical Engineering, Bulgarian Academy of Sciences, defended her Ph.D. thesis in the field of ventricular fibrillation detection (2001) and became Associate Professor (2007). From 2010 she is with the Institute of Biophysics and Biomedical Engineering, where she became Professor in 2017. Her scientific interests are in the field of biomedical data and signals processing.

Prof. Giovanni Bortolan, Ph.D.
E-mail: bortolan.cnr@gmail.com

Giovanni Bortolan received the laurea degree from the University of Padova, Padova, Italy in 1978. From 1981 he was a Research Fellow and presently he is a Senior Researcher at the Institute of Neuroscience IN-CNR (Institute of Biomedical Engineering ISIB-CNR until 2014) of the Italian National Research Council. He published numerous papers in the area of biomedical signal processing, neural networks, and applied fuzzy sets.

Sen. Assist. Prof. Todor Stoyanov, Ph.D.
E-mail: tstoyanov72@gmail.com

Todor Stoyanov graduated as M.Sc. from the Faculty of Electronics, Technical University of Sofia, in 1999. Since 1999 he is with the Centre of Biomedical Engineering and the Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences. He is Research Associate since 2002. He obtained Ph.D. degree in 2005 on computer aided processing and analysis of electrocardiograms. His interests are in developing embedded systems for biomedical signal analysis.

Prof. Ivan Dotsinsky, Ph.D., D.Sci.
E-mail: iadoc34@gmail.com

Ivan Dotsinsky obtained his M.Sc. degree from the Faculty of Electrical Engineering, Technical University of Sofia. His Ph.D. thesis was on the statistical assessment of the reliability of electrical and electronic circuitry. In 1987 he obtained the Dr.Eng.Sci. on Instrumentation of Electrocardiology. Since 1989, he has been Professor in Biomedical Engineering. Since 1994, he is a Professor with the Centre of Biomedical Engineering and the Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences. His interests are mainly in the field of acquisition, preprocessing, analysis and recording of biomedical signals.