UDC 616.12-073.7

DOI: 10.20535/2523-4455.mea.281741

The Method of Preprocessing of ECG Signals for Detection of Atrial and Ventricular Late Potentials

A. V. Mnevets[f], 0000-0001-5448-4045
N. H. Ivanushkina[s], PhD Assoc.Prof., 0000-0001-8389-7906
Department of Electronic Engineering
National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute» 00syn5v21
Kyiv, Ukraine

Abstract*—This article is aimed at analyzing and improving the methods of preprocessing ECG signals for the task of detecting low-amplitude regular components. This study analyzed the main advantages and disadvantages of existing ECG signal preprocessing methods for the detection of late ventricular and atrial potentials. Based on this analysis, a cardiac cycle averaging method was proposed in order to increase the accuracy of detection of late potentials by various algorithms and improve the quality of preprocessing of the ECG signal aimed at detection of low-amplitude components. The main feature of the proposed method is the division of a large number of cardiocycles for averaging into smaller aggregates (epochs), and the subsequent application of linear matrix decomposition to suppress irregular inclusions. Also, when dividing into epochs, it can be used overlapping. It can reduce the difference between epochs, and increase the number of cardiocycles for averaging. The use of this approach allows to minimize irregular inclusions in the ECG signal and increase the accuracy of the selection of low-amplitude late potentials. In addition, the division into epochs and overlapping makes possible to avoid blurring of low-amplitude high-frequency components during averaging as a result of heart rate variability, as well as to improve the quality of averaging with a reduced number of cardiocycles. To test the proposed method, various approaches were used to assess the ECG signal preprocessing. Mostly, we compared the cardiac cycles obtained as a result of different averaging algorithms and the proposed method with the template. To test the averaging method, an artificial ECG signal was developed with existing noise, late ventricular and atrial potentials, heart rate variability, and a high-amplitude component that occurs at a random location every two heartbeats. The template cardiac cycle was obtained from the original artificial signal without any distortion or noise. Firstly, we visually compared and evaluated different averaging methods with the template. Secondly, we calculated the similarity metrics of the late potentials on the averaged cardiac cycle with the late potentials on the template signal. Based on these metrics, the curves of dependence of the similarity values on the amplitude of late potentials on the ECG signal were calculated. Thirdly, we evaluated the impact of the proposed averaging method on the classification results of various machine learning algorithms on real ECG signals with available late potentials. The overall testing result showed that the proposed averaging method is able to reproduce the morphology of low-amplitude regular components by 10-30% more accurately and improve the classification accuracy by 5-12%.

Keywords — ventricular late potentials; atrial late potentials; electrocardiography; biomedical signal processing; signal processing algorithms; signal denoising; machine learning.

I. Introduction

Modern ECG screening methods make it possible to diagnose pathologies of the cardiovascular system in the early stages. Indeed, in the early stages of the disease, changes in the ECG are so insignificant that they can be overlapped, for example, by the noise of the electrodes and the quantization error of the initial signal [1]. The presence of late potentials can be a precursor of atrial and ventricular fibrillation and cause life-threatening conditions of the cardiovascular system [2].

Detection of late potentials can be used to diagnose atrial and ventricular fibrillation, as well as other life-threatening arrhythmias in the early stages, so it is very important to have sensitive and reliable methods for their determination.

The most common method of detecting low-amplitude components today is high-resolution ECG, as well as R-wave (SAECG) and P-wave (PSAECG) ECG signal averaging. Late potentials, which are studied by high-resolution electrocardiography and have an amplitude of 1-40 microvolts and lie in the frequency range of 40-250 Hz, are low-amplitude and high-frequency inclusions in the ECG [3].

The useful ECG signal consists of periodic pulses (cardiocycles), while the noise from the registration tract,

DOI: 10.20535/2523-4455.mea.281741

movement artifacts, and other abnormalities are non-periodic and harmful components. By averaging the cardiocycles, the conribution of the irregular components is reduced, and on the other hand, increased the regular ones [4].

When late potentials are determined, 250-400 cardiocycles are usually used [5]. A smaller number of cardiocycles leads to insufficient suppression of unwanted noise and irregular components. A higher number of cardiocycles leads to a smoothing of the useful low-amplitude signal.

Despite the fact that the cardiocycle averaging method suppresses noise and irregular components, the main disadvantage is that the resulting averaged signal is a superposition of the noise components of each individual cardiocycle with the useful signal. High-amplitude irregular components can contribute inclusions to the averaged signal, which in terms of shape and frequency composition will have the characteristics of the desired useful signal, but not be it.

Another approach to detecting low-amplitude components, namely late potentials, is a method based on a vector analysis of the ECG. For this method, orthogonal lead systems are used, for example, Frank's lead system [6]. This method has several improvements due to fewer leads and, as a result, fewer factors for mistakes. Although this method has certain advantages, it still has disadvantages associated with direct signal averaging. In addition, cardiographs with a system of Frank leads are not common. With the presence of other additional pathologies, the accuracy of the prediction of late potentials may decrease according to this method, for example, in the case of bundle branches block [7].

Frequency methods are also used to detect late potentials. Most often, frequency methods analyze the frequency composition of the averaged QRS complexes and allow analyzing the signal in a narrow frequency range, and, as a result, increase the ability to distinguish low-amplitude inclusions in the signal [8].

Despite the advantages of the frequency methods of low-amplitude component analysis, the basis for the analysis is the averaged signal. It can carry in frequency composition irregular disturbances that have penetrated into the averaged signal from noisy cardiocycles. Thus, to take advantage of the above methods, the most important thing is to obtain a standard averaged cardiocycle without noise interference and irregular inclusions that distort the diagnostic information. Such irregular inclusions can be: motion artifacts, poor contact between the skin and the electrode, ECG recording interruption artifacts, muscle tremors, electrical activity of the skin surface, convulsions, cough, breathing artifacts, various electromagnetic disturbances [9].

Methods for extracting late potentials based on the wavelet transform are also widely used. They are used both for denoising [10] and for the efficient detection of late potentials [11]. It has both a number of positive and negative features. Among the positives, wavelet transform has good localization in time and frequency domain, it allows analyzing fast and slow changes in the signal, has a high frequency resolution, can use various types of wavelets, which facilitates the search for the necessary components in the signal.

The negative features of the application of wavelet analysis are the complexity of transitions from scales to frequency values, since, depending on different wavelets, the frequency value on a certain scale can change. Also, choosing the right wavelet, and choosing the range of scales to identify the frequency components is not an easy task. Due to the high variability of the shape and frequency composition of the components that are interesting to us, it may be necessary to analyze signal on different types of wavelets, or use complex mathematical functions, which can complicate the calculation.

Also, methods of linear matrix decomposition are used [12] to reduce the amount of noise and irregular inclusions in the signal, as well as to reduce the attenuation effect of regular low-amplitude components (due to the variability of RR and other ECG intervals). These methods make it possible to use a smaller number of cardiocycles by extracting noise disturbances in vectors with smaller singular values than in the useful signal. Also, it can significantly reduce the level of the noise component without using averaging a high number of cardiocycles, as well as the influence of high-amplitude irregular inclusions and variability of ECG intervals.

Singular value decomposition (SVD) in the task of extracting late potentials is used according to various approaches. They have their own advantages, but the main disadvantage of these methods is that the number, or range of singular values, in which the late potentials will be extracted, varies greatly for real signals. This range may depend on many factors, such as the amplitude of late potentials, the number of decomposing vectors, the amplitudes of the ECG signal components, and other non-periodic components that can be isolated in the range of selected singular vectors and distort the resulting signal [13].

Other methods of linear decomposition of matrices, or methods of matrix factorization, such as principal component analysis (PCA), independent component analysis (ICA), non-negative matrix factorization (NMF), factor analysis (FA) are also used in related directions. For example, for detection of cardiac activity before fibrillation attacks, selection of low-amplitude fetal ECG, analysis of visual evoked potentials (VEP) on EEG, denoising [14]–[17]. Although these methods are promising for the task of detecting late potentials, the problem remains of establishing the exact range of main vectors, components, and factors in which late potentials will be highlighted and irregular disturbances suppressed.

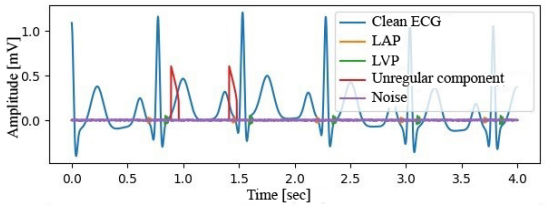
DOI: 10.20535/2523-4455.mea.281741

Thus, for the qualitative selection of late potentials, an preprocessing averaging method is required, which allows obtaininig an average cardiocycle with high accuracy and can minimize the impact of different types of interferences with saving morphology and useful low-amplitude components of the ECG signal during the process of signal averaging.

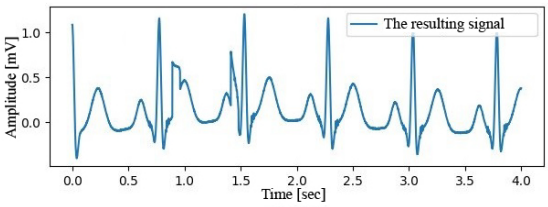
The purpose of the study is to create a method aimed at reducing the impact of noise and high-amplitude irregular disturbances on the useful ECG signal. The method aims to minimize the disadvantages of the above signal processing methods and will be suitable for use with a variety of ECG recordings. In addition, it will be suitable for the further selection of late potentials features (or other low-amplitude components).

II. MATERIALS AND METHODS

To develop the method, an artificial ECG signal was simulated with the presence of late potentials, irregular inclusions, and other features of the signal that can make a negative contribution to the calculation of the averaged cardiocycle. The basis of the artificial signal (Fig. 1) is a modeled cardiocycle based on the "ECGSYN" algorithm [18], which generates a pure ECG signal with duration of 200 seconds, with a heart rate of 80 beats per minute and a sampling frequency of 1000 Hz. The standard deviation of the heart was used with the 2 beats per minute. It was superimposed with a baseline drift caused by breathing. Two regular components were also superimposed on the modeled cardio signal, namely: late atrial potentials (LAP) — signals with an amplitude of 1-20 microvolts, and late ventricular potentials (LVP) — signals with an amplitude of 20-40 microvolts.



a)



b)

Fig. 1 Modeling of the testing ECG signal: a) components of the modeled signal; b) resulting modeled signal

The duration of the signals of the late potentials is 28 milliseconds. They have the following frequency composition:

$$LP(t) = a_1 \times \sin(2\pi f_1)(t) + a_2 \times \sin(2\pi f_2)(t) + \\ + a_3 \times \sin(2\pi f_3)(t),$$

where f_1 – 78 Hz, f_2 – 116 Hz, f_3 – 102 Hz, a_1: a_2: a_3 = = 5: 12: 2.

These particular parameters of the late potentials were chosen taking into account the electrophysiological features of the malfunction of the conduction system of the human heart [19].

Gaussian noise is also superimposed on the modeled signal so that the resulting signal-to-noise ratio is 40 dB. In addition, an irregular high-amplitude component is superimposed on the signal, which can simulate manifestations of movement artifacts, extrasystoles, pacemaker stimulations, or other similar artifacts on the ECG.

These artifacts can cause a negative contribution to the averaged signal. The shape of this component can be described by the following equation:

$$y = x + \left(x^{\frac{1}{10}} \right),$$

where $x \in [0;1]$.

The step of the x-axis and the number of counts for this component depends on its duration and sampling rate. On the modeled signal, the duration of the irregular component is 70 ms. The algorithm for adding an irregular component was configured to randomly generate it every two seconds at a random location in the cardiocycle with given parameters.

Fig. 1 shows how an irregular component is superimposed on the LAP in the second cardiac cycle, which distorts the waveform and can make a negative contribution to the resulting averaged signal.

During the study, preprocessing of the signal was applied to remove interferences, which included the following stages:

1) Removal of baseline wander (0.15-0.3 Hz).
2) Removal of breathing artifacts (0.25-0.5 Hz) [20].
3) Removal of common-mode interference (50, 60 Hz) and second harmonics (100, 120 Hz) [21].
4) Removal of high-frequency components, the frequency composition of which lies outside the frequency range of the useful signal (250 Hz).

Since late potentials lie in the range of 40-250 Hz, we can construct a band-pass filter with cut-off frequencies

DOI: 10.20535/2523-4455.mea.281741

of 1-250 Hz. A zero-phase, 10th-order Butterworth filter with a cutoff frequency of 1-250 Hz was created [22]. Common-mode interference was removed using a second-order digital notch filter with an infinite impulse response.

The quality factor (Q-factor) of the filter is 30, and the center frequencies are 50 and 100 Hz, if there is a 50 Hz common-mode interference, and 60 and 120 Hz, if the common-mode interference is 60 Hz. Since the frequency composition of the modeled signal does not contain common-mode interference, this type of filtering was not applied to the signal.

The division of the useful signal into cardiocycles was carried out in several stages. The first stage is the detection of R peaks. The second stage is the division of the ECG into separate cardiocycles and the formation of sets of cardiocycles (epochs). The "kalidas2017" detector [23] was chosen to detect R peaks. This algorithm was chosen due to its high performance, ability to real-time signal processing, and high accuracy.

The formation of cardiocycle epochs was performed by cutting the cardiocycle from the ECG recording. The beginning of the cardio cycle was set 0.3 seconds before the R peak. The end of the cardio cycle, on the contrary, was set 0.5 seconds after the R peak.

In Fig. 2, a — we can see that during the segmentation of the pure ECG signal, clearly superimposed cardiocycles were obtained. There were no mistakenly selected cardiocycles. There are observed manifestations of heart rate variability on the cardiocycle. Heart rate variability was established in the test signal model.

Due to heart rhythm variability we have a deviation of P, T waves from the central position, and as a result, a divergence in the position of LAP and LVP. Moreover, the divergence of LVP is smaller. It is caused by the fact that the averaging is carried out according to the R peak (SAECG). P peak averaging (PSAECG) was applied to reduce LAP position variance.

In Fig. 2, b, c, d, we can see the cardiocycles after superimposing noise and an irregular high-amplitude component. The quality of the selection of cardiocycles is decreased significantly. Fig. 2, b shows the situation when the R peaks was detected correctly despite nearby high-amplitude interference. But it happens that a high-amplitude component is detected as an R peak, which leads to an erroneous determination of its location (Fig. 2, d), or an erroneous determination as a result of the QRS detector correction algorithm (Fig. 2, c).

TABLE 1 COMPARISON OF INTRA-GROUP VARIANCE INDICATORS FOR A SIGNAL WITH AND WITHOUT AN IRREGULAR HIGH-AMPLITUDE COMPONENT

Signal type	STD	COV	IQR
Unregular component	0.099	615%	0.038
No unregular component	0.035	15%	0.021

To assess the effect of a high-amplitude irregular component on the segmentation of QRS complexes, intra-group variance was estimated in sets of cardiocycles with and without interference.

The standard deviation of (STD), coefficient of covariance (COV), and interquartile distance (IQR) were calculated and values have been entered into a table (Table 1) for counts in each group. We can see how the values of the parameters with the presence of the irregular high amplitude component are higher, which indicates a higher intra-group variance of cardiocycles and, as a result, can lead to a larger error in averaging.

The preprocessing of the signal did not allow significantly reducing the level of Gaussian noise, because the amplitude and frequency composition of such noise can be superimposed on the useful LVP and LAP signal. Moreover, the high amplitude of irregular component that was added to the signal cannot be effectively removed by classical preprocessing methods due to the frequency composition that lies in the frequency range of the investigated useful signal.

The presence of these inaccuracies led to the false detection of R peaks and, as a result, the false detection of cardiocycles. When the detector detects an irregular high-amplitude disturbance as an R peak, the real R peak

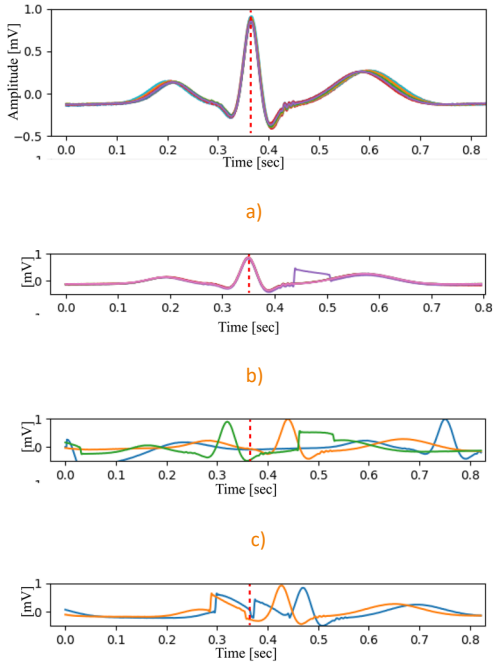


Fig. 2 Accumulation of cardiocycles of the model ECG signal with and without interferences:
a) cardiocycles of the signal without inter-ferences;
b) cardiocycles with existing interferences;
c) incorrect position of the R peak in the interfering signal caused by the correc-tion algorithm of the QRS detector;
d) incorrect position of the R peak in the interfering signal caused by false detection of the inter-fering signal as R peak.

DOI: 10.20535/2523-4455.mea.281741

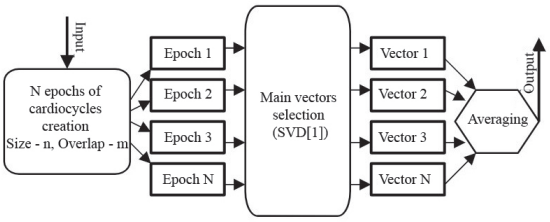


Fig. 3 Block diagram of the algorithm for building an averaged cardiocycle based on the selection of the main vectors as a result of the linear decomposition of the matrices of the epochs of the cardiocycles.

which superimposed on the averaged cardiac cycle in the wrong place becomes a disturbance. It can make a negative contribution to the process of forming the averaged cardiac cycle. This happens due to a similar frequency composition or as a result of missed R peaks detection algorithm, which exists in most ECG detectors [24].

In addition, the irregular component lying near the R peak can overlap with the part of the useful signal in the averaged cardiocycle with subsequent distortion.

To reduce the negative impact of irregular components on the useful signal, and to improve the existing methods of averaging cardiocycles for detecting late potentials, the following algorithm was proposed (Fig. 3).

The principle of the algorithm (Fig. 3) is following: cardiocycles that are selected from the initial signal (usually about 300 segments) are divided into sets (epochs) of *n* cardiocycles in each, and each epoch is selected with an overlap of *m* cardiocycles with the previous and next epoch. Then, a matrix is formed from each epoch, where the columns are formed from counts of a single cardiocycle from the epoch. Then, the singular value decomposition (SVD) method is applied to the obtained matrix.

The SVD method uses an *m*×*n* decomposition of the matrix *A* (where *n* is the number of columns or cardiocycles in the epoch, and *m* is the number of cardiocycle counts in the epoch) in such a way that:

$$A_{[m\times n]} = U_{[m\times r]} W_{[r\times r]} V^T_{[n\times r]},$$

where *A* – the original matrix, *U* – the left singular vectors, *W* – the diagonal matrix of eigenvalues, or singular values, *V* – the right singular vectors.

By zeroing the left and right singular vectors with the smallest singular values, we get a matrix approximation of the original set of cardiocycles, filtering out the least important components (for example, noise). This method is used for the purpose of obtaining a vector that shows the most significant dependence between columns (cardiocycles). This vector is the most informative component of the decomposed cardiocycle epoch matrix. In the last step, the most informative components (vectors) are averaged, forming the initial vector — an averaged cardiocycle with a significant reduction in the influence of negative factors.

An averaged cardiocycle is characterized by: a decrease in high-amplitude inclusions, a decrease in the degree of distortion of the signal shape (due to the inaccuracy of R peak detection), a decrease in the effect of blurring of low-amplitude components after averaging (due to the presence of heart rate variability).

The division into epochs of cardiocycles increases the reliability of the selection of regular low-amplitude components and makes it possible to apply other averaging methods to a reduced number of higher-quality, preprocessed cardiocycles (main vectors). The application of overlap increases the total amount of cardiocycles in processing and minimizes the influence of irregular components. The influence of such components could be manifested in a separate epoch of cardiocycles. In addition, splitting into epochs solves the problem when at a high number of cardiocycles during averaging (due to heart rate variability), the ability to extract late potentials was reduced. Using this method allows us to increase the number of cardiocycles and, as a result, increase the accuracy of low-amplitude components detection. Instead of the SVD algorithm, other matrix factorization algorithms can be used, such as principal component analysis (PCA), non-negative matrix factorization (NMF), factor analysis (FA), etc.

The proposed method aspires to isolate low-amplitude regular components precisely in the first singular vector or principal component (or factor) in order to avoid manifestations of harmful components with lower degree of regularity. Although, there are studies, in which late potentials appear in the second and even the third singular vector [25]. The proposed method provides for a number of measures to maximize the probability of the studied low-amplitude components falling into the first, main vector.

Usually, due to the high amplitude of the peaks of the QRS complex, the approximate form of the QRS complex is distinguished in the first main vector of a matrix decomposition. This occurs when there is a large number of cardiocycles in the matrix, to which the singular decomposition is applied. Using the division into epochs of cardiocycles allows reducing this negative effect. According to the reduced number of cardiocycles in the epoch, compared to the initial number of cardiocycles, a high degree of informativeness is carried not only by the approximate form of the ECG signal but also by less low-amplitude regular components, for example, late potentials.

This kind of sensitivity to low-amplitude regular components can be adjusted by defining the size of the cardiocycle epoch (*n*) and increasing the number of these epochs by overlapping (*m*).

The proposed method also increases the reliability of detecting a useful signal in the main decomposition vector. So, if in one epoch, regular late potentials will not be

DOI: 10.20535/2523-4455.mea.281741

determined in the first principal vector, they will be allocated in the first vector of other epochs. In addition, the reliability of detecting low-amplitude components, such as late potentials in the first, main vector, can be increased by filtering the signal in a narrow frequency range, thereby significantly reducing the influence of high-amplitude low-frequency components, or by cutting out zones of interest from cardiocycles.

This approach allows analyzing a much smaller segment of the signal and, as a result, significantly reduces the impact of unwanted components. For example, to detect LVP, the region of interest can be narrowed down to the ST-T segment.

The allocation of the averaged cardiocycle with late potentials by various methods was also investigated (Fig. 4). The results of averaging according to different methods (Fig. 4, b, c, d) were compared with the reference cardiocycle (Fig. 4, a). The results of averaging cardiocycles for the ECG from the public data bank were also analyzed [24]. Two real signals were assessed with classical averaging (Fig. 4, e, g) and with averaging by the proposed method (Fig. 4, f, h).

To form a reference cardiocycle (template), an artificial cardiocycle without existing disturbances and heart rate variability was formed, then LVP and LAP signals were superimposed on it, according to the same algorithm as for signals with disturbances.

Similarity metrics were also calculated to investigate the quality of detection of low-amplitude regular components (late potentials).

The methods of improving the quality of selection of late potentials were applied to the studied signal, namely, filtering in a narrow frequency range (40-240 Hz), and narrowing the zone of interest to the ST segment for LVP selection (Fig. 5, a), and to the PQ segment to select LAP (Fig. 5, b).

The cosine similarity (1) and the Pearson correlation coefficient (2) were used to evaluate the difference between the template signal and the signal extracted after averaging.

$$CosSimilarity_{A,B} = \frac{A \times B}{\| A \| \times \| B \|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}}, \quad (1)$$

where A, B – n-dimensional vectors, and $||A||$, $||B||$ – Euclidean norms of vectors (A, B) in the space of real numbers.

$$r_{pearson} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \times (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \times \sum_{i=1}^{n} (y_i - \bar{y})^2}}, \quad (2)$$

where x, y – n-dimensional vectors.

On the basis of similarity metrics (1) and (2), the quality of selection of late potentials by different methods was compared depending on their amplitude in the modeled signal.

The classic averaging method of cardiocycle epochs and the proposed averaging method with selection of main vectors based on SVD, PCA, and FA methods were compared (Fig. 6).

An artificial signal with an irregular high-amplitude component, Gaussian noise (with the resulting signal-to-noise ratio of 30 dB), and a heart rate of 80 beats per minute with a standard deviation of 2 beats per minute was applied to the input of various averaging algorithms. The parameters of the proposed averaging method were set to 30 epochs of cardiocycles with an overlap of 15 cardiocycles.

The arithmetic mean was used to average the main vectors obtained from the cardiocycle epochs. Quality index values were calculated for LVP and LAP with amplitudes of 1-40 microvolts with the step of 5 microvolts.

In addition, a study of the influence of the proposed signal preprocessing method on the classification results was conducted.

A test database with available late potentials was developed based on the public database [26] "PTB Diagnostic ECG Database Version: 1.0.0" to test the proposed method. Own development of the database was chosen because the data with available real (not artificial) late potentials are currently not widely distributed and available.

This database includes 549 high-resolution ECGs with a duration of about 115 seconds from 294 subjects with a quantization rate of 16 bits and a sampling frequency of each lead of 1000 Hz.

The database includes 148 patients with myocardial infarction, 18 patients with cardiomyopathy and heart failure, 14 patients with various types of arrhythmias, 7 patients with hypertrophy, as well as 52 healthy volunteers without existing disorders of cardiovascular activity.

These pathologies can be accompanied by the presence of low-amplitude LVP and LAP, on the basis of which a test database with available late potentials can be built.

The main feature of the "PTB Diagnostic ECG Database Version: 1.0.0" database is the inclusion of orthogonal Frank leads (X, Y, Z) in addition to the 12 standard leads.

This allows us to use the method of detecting late potentials based on the method generally accepted by the European Society of Cardiology, the American Heart Association, and the American College of Cardiology [27]. This method is based on the construction of the vector magnitude on the base of three ECG projections of

DOI: 10.20535/2523-4455.mea.281741

the heart from Frank's orthogonal leads. According to this method, using vector magnitude, we can calculate the necessary parameters of the cardiocycle to detect the presence of late potentials [28]. Thus, the following vector magnitude parameters are characteristics of LVP: the duration of the QRS complex > 114 ms, the root-mean-square voltage of the last 40 ms (RMS40) < 20 microvolts, the total duration of the fragment of the QRS complex with voltage < 40 microvolts (LAS) more than 38 milliseconds. In turn, the following parameters of vector magnitude indicate the presence of LAP: P wave duration (PWD) > 115 ms, root-mean-square voltage of the last 20 ms P wave (RMS20) < 2.2 microvolts. The presence of late potentials was declared if at least two of the three conditions for LVP and both conditions for LAP were obtained [11]. Therefore, based on this method, a test database was obtained, in which there are 137 recordings with available LVPs, 47 recordings with available LAPs, 14 recordings with available LVPs and LAPs, and 351 recordings without available late potentials.

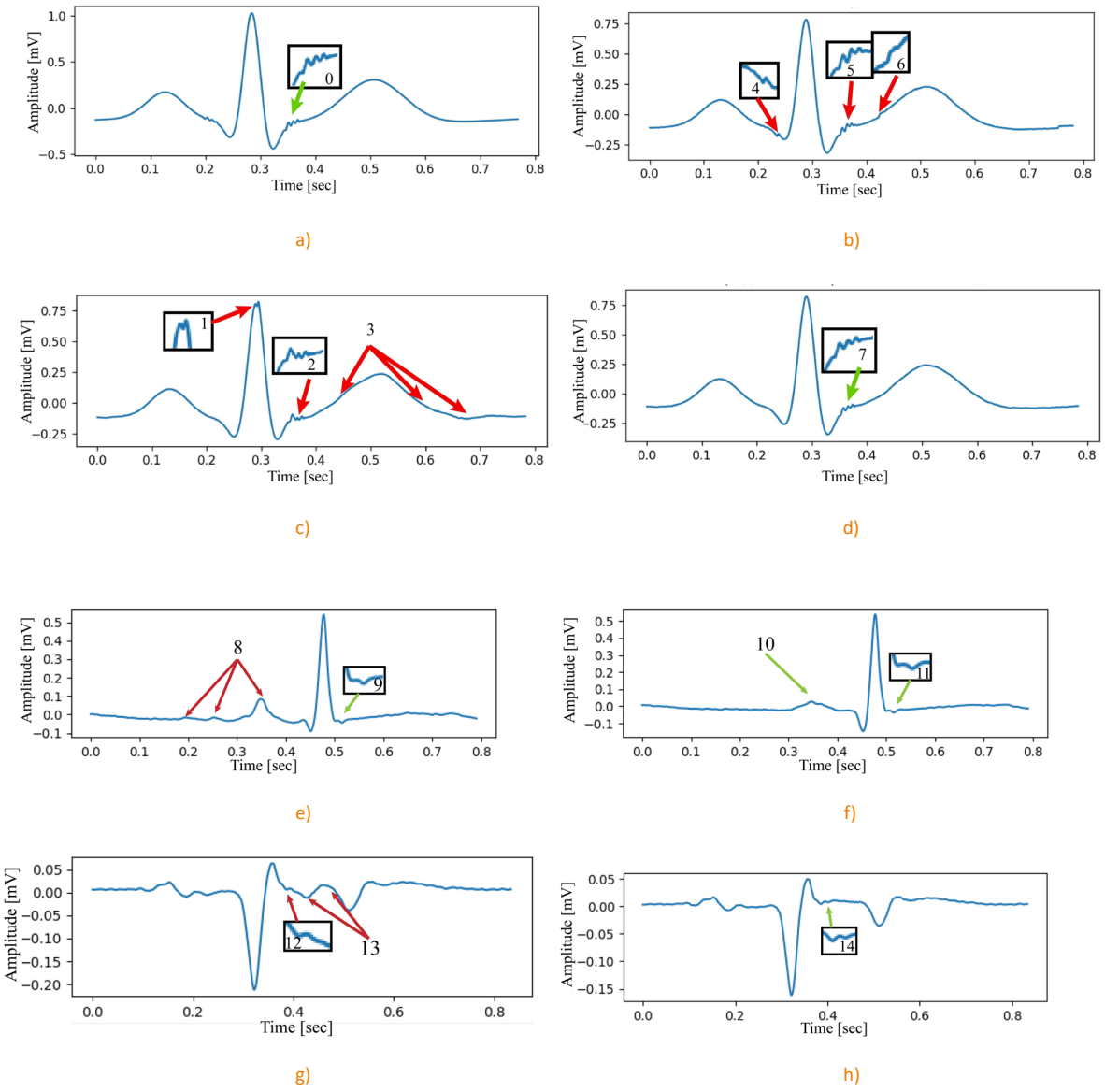
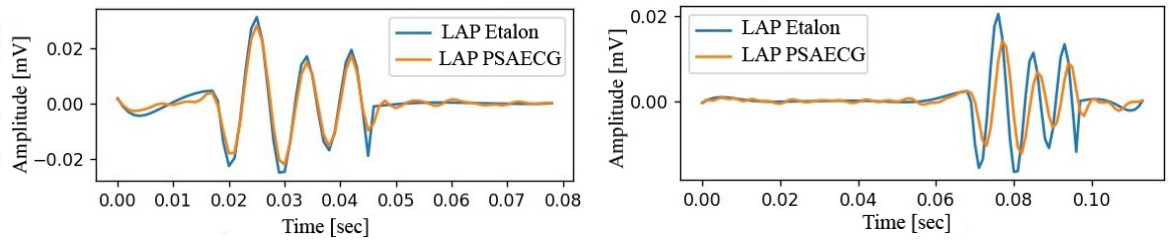


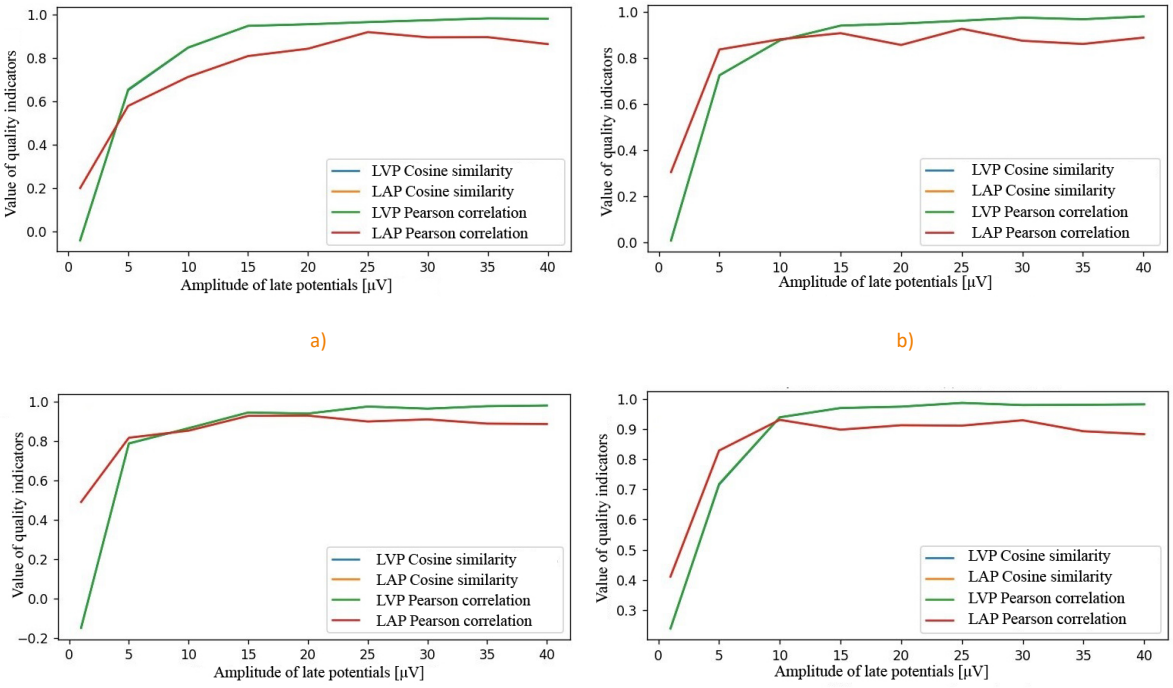
Fig. 4 Cardiocycles averaging by various methods:
a) reference cardiocycle with available late potentials; b) the main vector in the SVD decomposition of one epoch;
c) averaged cardiocycle; d) averaged cardiocycle built on the basis of the proposed method;
e) averaged cardiocycle “a” from the databank obtained by classic averaging; f) averaged cardiocycle “a” from the databank obtained by averaging using proposed method;
g) averaged cardiocycle “b” from the databank obtained by classic averaging; h) averaged cardiocycle “b” from the databank obtained by averaging using proposed method;
0 – LVP on the template cardiocycle; 1, 2, 3 – artifacts on the averaged cardiocycle; 4, 5, 6 – artifacts on the main vector of one epoch; 7 – LVP after averaging using the proposed method; 8, 13 – distortion of the shape of the cardiocycle as a result of signal averaging with interferences; 10 – saving of the shape of the P wave with low-amplitude inclusions when averaging by the proposed method; 12 – distortion of the area with the presence of late potentials as a result of averaging the signal with interferences; 9, 11, 14 – late potentials on SAECG.

DOI: 10.20535/2523-4455.mea.281741



a) b)

Fig. 5 Comparison of late potentials after averaging with template:
a) LVP in template and averaged signal; b) LAP in the template signal, and after applying averaging



a) b)

c) d)

Fig. 6 Quality indicators of late potentials allocation (LAP and LVP) depending on their amplitude and type of averaging method:
a) for classical averaging method; b) for the proposed SVD-based averaging method;
c) for the proposed PCA-based averaging method; d) for the proposed averaging method based on FA.

Classification models based on the support vector machine method (SVM), random forest classifier (RF), logistic regression (LR), and k-nearest neighbors (K-neighbor) were developed for testing the proposed preprocessing method.

For classification, the proposed averaging algorithm (Fig. 3) had the following parameters: epoch size n = = 25 cardiocycles, overlap m = 10 cardiocycles, selection of the main vector (cardiocycle) was based on the SVD method, averaging of the main vectors was carried out by the arithmetic mean.

For the selected classification models, a system of features was calculated, that can characterize the presence of late potentials in time and frequency representation. The set of features was calculated on the basis of one of the existing approaches to the detection of late potentials using machine learning [10].

The basis for calculating the set of features was the vector magnitude (*VM*), which is extended to an arbitrary number of cardiographic leads, and calculated from 12 standard leads (I, II, III, V1, V2, V3, V4, V5, V6) according to the formula [28]:

$$VM = \sqrt{\sum_{k=1}^{N} X_k^2(t)},$$

where X_k – the averaged cardiocycle for each of the selected leads.

Therefore, during the training of the classification model, a set of 5 features used to classify LVPs and a set of 3 features used to classify LAPs were formed for each

DOI: 10.20535/2523-4455.mea.281741

examined ECG signal. For LVP, according to [28], the features based on the vector magnitude graph were calculated:

- the delay between R peak and point j (the point of transition of the S wave into the ST segment), in milliseconds;
- the value of the root-mean-square voltage of the last 40 ms (RMS40) of the QRS complex, in microvolts;
- the value of the total duration of the QRS complex sections with voltage < 40 microvolts, in milliseconds;
- the value of the energy ratio according to the time-frequency representation between the vector magnitude in the time range of LVP and the energy of the QRS complex in the frequency range of 55-300 Hz;
- the value of the energy ratio according to the time-frequency representation between the vector magnitude 80 ms after the end of the QRS complex and the energy of the QRS complex in the frequency range of 55-300 Hz.

Then, by analogy, the following features were calculated for late LAPs:

- the delay between the top of the P peak and the end of the P wave;
- root-mean-square voltage of the last 20 ms P wave (RMS20), in microvolts;
- the value of the energy ratio according to the time-frequency representation between the vector magnitude 80 ms after the end of the P wave and the energy of the P wave in the frequency range of 55-300 Hz.

To study the proposed method, a binary classification was carried out. Classification models were trained to predict normality and pathology signals with and without late potentials using ECG data from databank. On the basis of the developed database with available real ECG signals with LVP, and LAP, as well as control healthy signals, the described features were calculated. Features were calculated using different types of data preprocessing: classical averaging and averaging according to the proposed method. In the obtained datasets of features, normal and pathology classes were coded into the numerical format 0 and 1, respectively. After that, observations with a large number of missing values (2 or more) were removed. The rest of the missing values were replaced by the medians of the corresponding features.

Outliers that lie outside three standard deviations were also removed and replaced by the median of the corresponding sample in the dataset. After that, truncation of the largest class to the size of the smallest was carried out by the method with the random selection of observations to obtain a class sizes ratio of 1:1. The last step of data preparation was data standardization using the "Robust Scaler" algorithm, which is resistant to outliers [29].

As a result, 141 normal observations and 141 pathology observations for LVP signals (processed using classical averaging), 138 normal observations and 138 pathology observations for LVP signals (processed using the proposed averaging method), and 42 and 46 observations, for LAP processed using classical averaging and the proposed averaging method, respectively, were sent to the machine learning algorithms after the feature sets preprocessing.

Based on the obtained results of classification from developed database with real LVP and LAP contained ECG signals, the number of true positive (TP), true negative (TF), false positive (FP), and false negative (FN) results were calculated. These numbers were used to calculate the statistical metrics of the accuracy of the classification models, namely: the probability of making a first-order error (FPR), the probability of making a second-order error (FNR), sensitivity (TPR), specificity (TNR), and overall accuracy (ACC). These metrics were calculated by the following formulas:

$$TPR = \frac{TP}{TP + FN}; TNR = \frac{TN}{TN + FP}$$

$$FPR = \frac{FP}{FP + TN}; FNR = \frac{FN}{FN + TP}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

Also, the average value of each accuracy metric was calculated for further analysis.

Thus, during the study of the influence of the proposed preprocessing method of the ECG signal on the classification results, a summary table (Table 2) was obtained, according to which the comparison was made.

III. Results and discussion

The analysis of the averaged cardiocycle (Fig. 4, c) demonstrates a reduction in the contribution of the manifestations of inaccurate selection of cardiocycles and high-amplitude irregular components to the overall signal shape, but the distortion of the R wave (Fig. 4, 1) due to the presence of a set of irregular components that were detected by the detector as R peak.

DOI: 10.20535/2523-4455.mea.281741

TABLE 2 COMPARISON OF THE RESULTS OF THE CLASSIFICATION OF REAL ECG SIGNALS WITH AVAILABLE ATRIAL AND VENTRICULAR LATE POTENTIALS ACCORDING TO CLASSICAL AVERAGEMENT AND WITH THE PROPOSED AVERAGING METHOD

Classification model	Late ventricular potentials									
	Classical averagement					*Proposed averaging method*				
	FPR	*FNR*	*TPR*	*TNR*	*ACC*	*FPR*	*FNR*	*TPR*	*TNR*	*ACC*
SVM	0,45	0,19	0,81	0,55	0,67	0,35	0,14	0,86	0,65	0,74
RF	0,28	0,35	0,65	0,72	0,69	0,16	0,36	0,64	0,84	0,75
LR	0,45	0,19	0,81	0,55	0,67	0,29	0,23	0,77	0,71	0,74
K-neighbor	0,21	0,35	0,65	0,79	0,73	0,13	0,45	0,55	0,87	0,74
Mean of results	0,35	0,27	0,73	0,65	0,69	0,23	0,3	0,7	0,77	0,74
	Late atrial potentials									
SVM	0,55	0,33	0,67	0,45	0,53	0,17	0,57	0,43	0,83	0,68
RF	0,55	0,00	1,00	0,45	0,65	0,25	0,43	0,57	0,75	0,68
LR	0,55	0,50	0,50	0,45	0,47	0,17	0,57	0,43	0,83	0,68
K-neighbor	0,36	0,50	0,50	0,64	0,59	0,17	0,57	0,43	0,83	0,68
Mean of results	0,50	0,33	0,66	0,50	0,56	0,19	0,54	0,47	0,81	0,68

This phenomenon occurred due to the presence of a set of irregular components, that were detected by the ECG detector as an R peak. There is also a distortion of the shape of the late potentials (Fig. 4, 2). This occurred due to the presence of an irregular high-amplitude component, as well as other noises in this area during the averaging of cardiocycles. In addition, there is a distortion of the T wave shape (Fig. 4, 3). This distortion is the result of the high number of interferences and wrong detection of the R peak position. Because of that, the T waveform on the averaged signal began to differ from the reference one. As a result, bulges and depressions began to appear in places, where they should not be.

The analysis of one epoch of the SVD decomposition (Fig. 4, b) shows, that the main vector has inclusions of high-amplitude irregular components — (Fig. 4, 4, 6), but their amplitude is much lower, than disturbances in the initial signal. The manifestation of these components in one epoch can be explained by the presence of several high-amplitude irregular disturbances in different cardiocycles in this epoch. Also, as a result of extracting the main vector in one epoch, the LVP was slightly distorted (Fig. 4, 5), but these distortions are much smaller than in the case of the averaged cardiocycle. These results (Fig. 4, b) were obtained from an epoch of only 30 cardiocycles, that is a significant improvement compared to the separately averaged signal of 250-300 cardiocycles. The best result was obtained during cardiocycle averaging using the proposed method (Fig. 4, d). Manifestations of the high-amplitude irregular component are almost not observed, and the morphology of the signal is largely similar to the reference cardiocycle. In addition, the section of the cardiocycle with late potentials (Fig. 4, 7) was selected best with the least distortions relative to the reference cardiocycle (Fig. 4, a).

Analyzing signals from the data bank (Fig. 4, e, g, f, h), we found that the classical averaging of cardiocycles can lead to distortion of the morphology of the cardiocycle (Fig. 4, 8, 13), especially in the presence of high-amplitude noise in the real ECG signal. This can significantly distort the results of the classification of various pathologies and late potentials. This effect is especially negative when the distortion of the shape of the signal falls on the area with late potentials (Fig. 4, 12). These deformations can change the form of low-amplitude components and make them indistinguishable for a doctor, as well as artificial algorithms for detecting late potentials.

Averaging of real signals form the databank over the proposed method, in turn, shows a more accurate display of the morphology of the ECG wave (Fig. 4, 10), as well as greater resistance to interference, maintaining low-amplitude inclusions at the same time, in which late atrial potentials can appear. As a result, it makes possible to observe and measure the area with late potentials in more details.

From the graphs of the similarity metrics (Fig. 6) it can be observed, that the proposed method at low amplitudes of late potentials manifests itself much better than the usual averaging of cardiocycles. At LAP with an amplitude of 1 microvolt, the quality index for the proposed SVD-based algorithm is 10% better than for the averaged signal Fig. 6, b), 20% better for the FA-based algorithm (Fig. 6, d), and almost 30% better using PCA-based algorithm (Fig. 6, c). This trend is also observed in LVP quality indicators. In general, at low amplitudes of late potentials, the quality indicators of the proposed method are 10-20% better than the indicators of the classic averaged signal. Even in the case, when LVP has an amplitude of 1 microvolt and the averaged signal shows zero similarity with the template, the proposed method based on factor analysis shows a similarity coefficient of about 0.25. With an increase in the amplitude of late potentials (when exceeding 10-15 microvolts), the sensitivity of the proposed method and the usual averaging of the cardiocycle are equalized and become close to 1. But the quality metrics of the proposed method are 3-20% better in general.

During testing of the proposed method using an input signal containing about 300 cardiocycles, the following parameters showed the best results: epoch size n = 25-30 cardiocycles, overlap size m = 10-15 cardiocycles.

DOI: 10.20535/2523-4455.mea.281741

On the basis of the classification results of the developed database with real LVP and LAP contained ECG signals (Table 2), we can see that the overall classification accuracy for all calculated models varies between 55% and 75%. Such rather low values are the result of the low quality of the original database with real ECG data, which was calculated automatically according to the classic method of LVP and LAP detection. The data has not been manually validated, so the resulting dataset contains false class labels, on which the models are trained. In addition, three to five signs (depending on the type of late potentials) were calculated from the 12-channel ECG according to the above-mentioned method. This number of features could be insufficient for a complete description of the dispersion of signals with available late potentials. But despite this, the quality of the obtained database is enough for testing signal preprocessing methods, comparing and analysis of their impact on the quality of classification.

The analysis of Table 2 shows, that the overall classification accuracy is higher for the proposed averaging method than for the classic method. For LVP, accuracy is 5% higher on average, and for LAP 12% higher on average. In addition, we can see that the probability of making a first-order error according to the FPR metric is also lower with the proposed averaging method, by 12% for LVP and by 31% for LAP. This indicates, that the proposed method is more resistant to disturbances, which could appear on the averaged cardiocycle. This property can also be useful in late potentials determination. And this resistance is better manifested with a decrease in the amplitude of the useful signal.

On the other hand, the probability of making a second-order error by the FNR metric is slightly higher for the proposed averaging method, than for the classical method, on average by 3% for LVP and 21% for LAP. This result is explained by the omission of some late potentials due to the high degree of approximation.

This disadvantage can be corrected by reducing the parameters of epoch size – m and overlap – n, or by changing the method of selection of main vectors to a more sensitive one, for example, factor analysis (FA). We also see a significant increase in the TNR metric, and suppression in the TPR metric. But the suppression of TPR is much smaller than the increase of TNR, which is reflected in the increase in the overall accuracy of the classification by the proposed averaging method.

To improve the sensitivity of the proposed method in order to reduce the probability of second-order error, it is also possible to apply modified approaches to averaging the main vectors (cardiocycles) at the last stage of the proposed averaging method. These approaches could be modified using frequency methods, wavelet analysis, etc.

Conclusions

So, the proposed method carries a number of improvements for the reliable detection of late potentials, even in highly noisy cardiac signals. The analysis of interference resistance on the averaged cardiocycles showed that the proposed method is more resistant to high-amplitude irregular interference, wrong selections of cardiocycles, and distortions of the useful low-amplitude signal (late potentials) than the classical averaging algorithm.

The results of the comparison of the late potentials extracted using the proposed method and the usual averaging based on similarity metrics were obtained. According to these results, the proposed method extracts a signal that is more similar to the template after processing of signal epochs.

This phenomenon is especially evident at small amplitudes of the useful signal.

On the base of the analysis of the impact of the proposed method on the classification results, it was found that its use leads to an improvement in the overall accuracy of the classification, and also to a decrease in the probability of the first kind of error.

This method does not make changes to the morphology of the signal and is suitable for further application of already existing methods for detecting late potentials. In addition, this method can be further developed at the stages of selection of the main vectors or their averaging. It can also improve the reliability and quality of the detection of low-amplitude components.

References

[1] D. C. Bauer, “Association of Major and Minor ECG Abnormalities With Coronary Heart Disease Events”, JAMA, vol. 307, no. 14, p. 1497, Apr. 2012, **DOI:** 10.1001/jama.2012.434.

[2] D. Saito *et al.*, “Atrial late potentials are associated with atrial fibrillation recurrence after catheter ablation,” *J. Arrhythmia*, vol. 38, no. 6, pp. 991–996, Dec. 2022, **DOI:** 10.1002/joa3.12786.

[3] S. Narayanaswamy, “High resolution electrocardiography,” *Indian Pacing Electrophysiol. J.*, vol. 2, no. 2, pp. 50–56, Apr. 2002, **PMID:** 17006557.

[4] P. Gomis and P. Caminal, “Evaluation of very low amplitude intra-QRS potentials during the initial minutes of acute transmural myocardial ischemia,” *J. Electrocardiol.*, vol. 47, no. 4, pp. 512–519, Jul. 2014, **DOI:** 10.1016/j.jelectrocard.2014.04.014.

[5] B. Dinov *et al.*, “Signal-Averaged Electrocardiography as a Noninvasive Tool for Evaluating the Outcomes After Radiofrequency Catheter Ablation of Ventricular Tachycardia in Patients With Ischemic Heart Disease,” *Circ. Arrhythmia Electrophysiol.*, vol. 9, no. 9, Sep. 2016, **DOI:** 10.1161/CIRCEP.115.003673.

[6] R. Jaros, R. Martinek, and L. Danys, “Comparison of Different Electrocardiography with Vectorcardiography Transformations,” *Sensors*, vol. 19, no. 14, p. 3072, Jul. 2019, **DOI:** 10.3390/s19143072.

DOI: 10.20535/2523-4455.mea.281741

[7] G. Speranza, P. Bonato, and R. Antolini, "Analyzing late ventricular potentials," *IEEE Eng. Med. Biol. Mag.*, vol. 15, no. 3, pp. 88–94, 1996, **DOI:** 10.1109/51.499765.

[8] E. Bou Assi, L. Gagliano, S. Rihana, D. K. Nguyen, and M. Sawan, "Bispectrum Features and Multilayer Perceptron Classifier to Enhance Seizure Prediction," *Sci. Rep.*, vol. 8, no. 1, p. 15491, Oct. 2018, **DOI:** 10.1038/s41598-018-33969-9.

[9] A. R. Pérez-Riera, R. Barbosa-Barros, R. Daminello-Raimundo, and L. C. de Abreu, "Main artifacts in electrocardiography," *Ann. Noninvasive Electrocardiol.*, vol. 23, no. 2, p. e12494, Mar. 2018, **DOI:** 10.1111/anec.12494.

[10] Giorgio, Rizzi, and Guaragnella, "Efficient Detection of Ventricular Late Potentials on ECG Signals Based on Wavelet Denoising and SVM Classification," *Information*, vol. 10, no. 11, p. 328, Oct. 2019, **DOI:** 10.3390/info10110328.

[11] X. Fagan, K. Ivanko, and N. Ivanushkina, "Detection of Ventricular Late Potentials in Electrocardiograms Using Machine Learning", in Advances in Intelligent Systems and Computing, vol. 1247, 2021, pp. 487–497, **ISBN:** 9783030555054, **URL:** http://link.springer.com/10.1007/978-3-030-55506-1_44.

[12] J. H. Caltenco, J. L. Lopez-Bonilla, B. E. Carvajal-Gámez, and P. Lam-Estrada, "Singular Value Decomposition," *Bull. Soc. Math. Serv. Stand.*, vol. 11, pp. 13–20, Sep. 2014, **DOI:** 10.18052/www.scipress.com/BSMaSS.11.13.

[13] Y. Biton *et al.*, "Causality analysis of leading singular value decomposition modes identifies rotor as the dominant driving normal mode in fibrillation," *Chaos An Interdiscip. J. Nonlinear Sci.*, vol. 28, no. 1, p. 13128, Jan. 2018, **DOI:** 10.1063/1.5021261.

[14] R. J. Martis, U. R. Acharya, and L. C. Min, "ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform," *Biomed. Signal Process. Control*, vol. 8, no. 5, pp. 437–448, Sep. 2013, **DOI:** 10.1016/j.bspc.2013.01.005.

[15] T. Kurita, "Principal Component Analysis (PCA)," in *Computer Vision*, Boston, MA: Springer US, 2014, pp. 636–639, **DOI:** 10.1007/978-0-387-31439-6_649.

[16] J. Dien, D. J. Beal, and P. Berg, "Optimizing principal components analysis of event-related potentials: Matrix type, factor loading weighting, extraction, and rotations," *Clin. Neurophysiol.*, vol. 116, no. 8, pp. 1808–1825, Aug. 2005, **DOI:** 10.1016/j.clinph.2004.11.025.

[17] S. YU and K. CHOU, "Integration of independent component analysis and neural networks for ECG beat classification," *Expert Syst. Appl.*, vol. 34, no. 4, pp. 2841–2846, May 2008, **DOI:** 10.1016/j.eswa.2007.05.006.

[18] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 3, pp. 289–294, Mar. 2003, **DOI:** 10.1109/TBME.2003.808805.

[19] T. Szél and C. Antzelevitch, "Abnormal Repolarization as the Basis for Late Potentials and Fractionated Electrograms Recorded From Epicardium in Experimental Models of Brugada Syndrome," *J. Am. Coll. Cardiol.*, vol. 63, no. 19, pp. 2037–2045, May 2014, **DOI:** 10.1016/j.jacc.2014.01.067.

[20] S. Kaplan *et al.*, "Filtering respiratory motion artifact from resting state fMRI data in infant and toddler populations," *Neuroimage*, vol. 247, p. 118838, Feb. 2022, **DOI:** 10.1016/j.neuroimage.2021.118838.

[21] M. Butt, N. Razzaq, I. Sadiq, M. Salman, and T. Zaidi, "Power Line Interference removal from ECG signal using SSRLS algorithm," in *2013 IEEE 9th International Colloquium on Signal Processing and its Applications*, 2013, pp. 95–98, **DOI:** 10.1109/CSPA.2013.6530021.

[22] A. E. Cetin, O. N. Gerek, and Y. Yardimci, "Equiripple FIR filter design by the FFT algorithm," *IEEE Signal Process. Mag.*, vol. 14, no. 2, pp. 60–64, Mar. 1997, **DOI:** 10.1109/79.581378.

[23] V. Kalidas and L. Tamil, "Real-time QRS detector using Stationary Wavelet Transform for Automated ECG Analysis," in *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)*, 2017, pp. 457–461, **DOI:** 10.1109/BIBE.2017.00-12.

[24] M. A. Z. Fariha, R. Ikeura, S. Hayakawa, and S. Tsutsumi, "Analysis of Pan-Tompkins Algorithm Performance with Noisy ECG Signals," *J. Phys. Conf. Ser.*, vol. 1532, no. 1, p. 12022, Jun. 2020, **DOI:** 10.1088/1742-6596/1532/1/012022.

[25] C. Guaragnella, M. Rizzi, and A. Giorgio, "Marginal Component Analysis of ECG Signals for Beat-to-Beat Detection of Ventricular Late Potentials," *Electronics*, vol. 8, no. 9, p. 1000, Sep. 2019, **DOI:** 10.3390/electronics8091000.

[26] R. Bousseljot, D. Kreiseler, and A. Schnabel, "Nutzung der EKG-Signaldatenbank CARDIODAT der PTB uber das Internet [Usage of the ECG signal database CARDIODAT of PTB via the internet.]," *Biomed. Tech. Eng.*, pp. 317–318, Jul. 2009, **DOI:** 10.1515/bmte.1995.40.s1.317.

[27] G. Breithardt *et al.*, "Standards for analysis of ventricular late potentials using high-resolution or signal-averaged electrocardiography. A statement by a Task Force Committee of the European Society of Cardiology, the American Heart Association, and the American College of Ca," *Circulation*, vol. 83, no. 4, pp. 1481–1488, Apr. 1991, **DOI:** 10.1161/01.CIR.83.4.1481.

[28] H. Takahashi, M. Takagi, T. Yoshio, M. Yoh, and I. Shiojima, "Analysis of site-specific late potentials using a novel Holter signal-averaged electrocardiography in patients with Brugada syndrome," *Hear. Rhythm*, vol. 19, no. 10, pp. 1650–1658, Oct. 2022, **DOI:** 10.1016/j.hrthm.2022.05.005.

[29] X. H. Cao, I. Stojkovic, and Z. Obradovic, "A robust data scaling algorithm to improve classification accuracies in biomedical data," *BMC Bioinformatics*, vol. 17, no. 1, p. 359, Dec. 2016, **DOI:** 10.1186/s12859-016-1236-x.

Надійшла до редакції 12 червня 2023 року
Прийнята до друку 05 серпня 2023 року

УДК 616.12-073.7

DOI: 10.20535/2523-4455.mea.281741

Метод попередньої обробки ЕКГ сигналів для виявлення пізніх потенціалів передсердь та шлуночків

А. В. Мневець[f], 0000-0001-5448-4045
Н. Г. Іванушкіна[s], к.т.н. доц., 0000-0001-8389-7906
Кафедра електронної інженерії
Національний технічний університет України
"Київський політехнічний інститут імені Ігоря Сікорського" 00syn5v21
Київ, Україна

***Анотація*—Дана стаття націлена на аналіз та вдосконалення методів попередньої обробки електрокардіосигналів для задачі виявлення низькоамплітудних регулярних компонент. В даному дослідженні було проаналізовано основні переваги та недоліки існуючих підходів до попередньої обробки ЕКГ сигналу для задачі виявлення пізніх потенціалів шлуночків та передсердь. На основі проведеного аналізу, з метою підвищення точності виявлення пізніх потенціалів різними алгоритмами, та покращення якості попередньої обробки ЕКГ сигналу для виявлення низькоамплітудних компонент, було запропоновано удосконалений метод усереднення кардіоциклів. Головною особливістю запропонованого методу є розбиття великої кількості кардіоциклів для усереднення на менші сукупності (епохи), та подальше застосування лінійного матричного розкладання для придушення нерегулярних включень. Також за необхідності у разі розбиття на епохи можна використовувати перекриття. Це дозволить зменшити розбіжності між епохами, та збільшити кількість кардіоциклів. Використання даного підходу дає можливість мінімізувати нерегулярні включення в ЕКГ сигнал та підвищити точність виділення низькоамплітудних пізніх потенціалів. Крім того, розбиття на епохи та перекриття дозволяє уникати розмиття низькоамплітудних високочастотних компонент під час усереднення в результаті варіабельності серцевого ритму, а також покращувати якість усередненого сигналу при використанні зменшеної кількості кардіоциклів. Для тестування запропонованого методу використовувались різні підходи для оцінки попередньої обробки ЕКГ сигналу. Переважно було проведено порівняння кардіоциклів, які були отримані в результаті роботи різних алгоритмів усереднення та запропонованого методу, з еталоном. Для тестування методу усереднення, було розроблено штучний ЕКГ сигнал з наявним шумом, наявними пізніми потенціалами шлуночків та передсердь, варіабельністю серцевого ритму, а також з високоамплітудною компонентою, яка виникає у випадковому місці через кожні два серцеві скорочення. Еталонний кардіоцикл було отримано з вихідного штучного сигналу без наявних спотворень та шуму. По перше, було проведено візуальне порівняння та оцінка різних методів усереднення з еталоном. По друге, було розраховано метрики подібності пізніх потенціалів на усередненому кардіоциклі з пізніми потенціалами на еталонному сигналі. На основі даних метрик було побудовано криві залежностей значень метрик подібності від амплітуди пізніх потенціалів на ЕКГ сигналі. По третє, було проведено оцінку впливу запропонованого методу усереднення на результати класифікації реальних ЕКГ сигналів з наявними пізніми потенціалами за допомогою різних алгоритмів машинного навчання. Загальний результат тестування показав, що запропоний метод усереднення за результатами дослідження здатний відтворювати морфологію низькоамплітудних регулярних складових на 10-30% точніше, та покращувати точність класифікації на 5-12%.**

Ключові слова — пізні потенціали шлуночків; пізні потенціали передсердь; електрокардіографія; обробка біомедичних сигналів; алгоритми обробки сигналів; усунення шумів сигналу; машинне навчання.