Algorithm for ECG forms clustering based on artificial neural networks

Van G.V.*, Nekrasov I.A.
Saint Petersburg Electrotechnical University, «LETI»
*gennady.van@mail.ru

Abstract. The possibilities of machine learning methods and artificial neural networks use to solve problems of blind ECG cardiocycles clustering were explored. An algorithm for ECG cardiocycles clustering is based on the Kohonen neural network and k-means method is developed. The proposed algorithm is evaluated on annotated standard database MGH/MF waveform database. The accuracy indicators of this algorithm is Acc – 96,92%. This algorithm can be used as the basis for ECG diagnostic software.

1 Introduction
An electrocardiogram (ECG) is a source of receiving diagnostic information used to monitor cardiac activity. One of the urgent problems of modern cardiology is to obtain the most complete information about the electrical potential of the heart, on the basis of which it would be possible to expand the diagnosis of the pathological conditions of the myocardium and its electrophysiological properties. The analysis of cardiocycles extracted from ECG signals may reflect symptoms indicating that the heart needs immediate attention. The approach for cardiocycles clustering using methods of artificial neural networks (ANN) was investigated. In the clustering of cardiac cycles of electrocardiograms and the establishment of cardiodiagnosis, sets of informative features, based on the characteristic properties of QRS complexes, were used.

2 Methodology
In this section, we describe the proposed algorithm for the clustering of ECG cardiocycles in MGH/MF waveform database using Kohonen’s self-organized map. The algorithm’s steps are considered in Figure 1. The implementation of the algorithm and the results of the research were received from Matlab.
2.1 ECG database
The signals required for the analysis were acquired from free available MGH/MF waveform database [6]. MGH/MF waveform database is the collection of 250 patients recording. Signals in the database are sampled at 360 samples/second. Each record consists a .dat file, .hea file and .ari file. Header file consists of information regarding the type signals and type of leads in case of ECG signal, sampling interval, sampling frequency, duration of the signal and units. It comprises of the information to convert the recorded signals from raw units to physical units. The .ari file consists of ECG annotations whereas .dat file comprise of 8 signals.

For the study, fragments from the MGH/MF Waveform Database were used, which contains 48 annotated signals for three standard leads with duration of 60 seconds at a sampling frequency of 250 Hz. Since the problem of blind clustering was solved using the proposed clustering algorithm, the training sample was not used (the neural network training occurred directly on the test data) [6].

2.2 ECG pre-processing
The received raw digital ECG signal of a record is shown in Figure 2a. To increase of algorithm’s efficiency the procedure of preliminary signal processing was used cause the signal contains different noise and artifacts. For the conversion a bandpass filter with a passband frequencies from 5 Hz to 20 Hz. The filtered ECG signal is shown in Figure 2b.
2.3 Informative features
In this stage, the adaptive threshold values of the informative features were created. The choice of informative features was based on the QRS complex specific possesses, it’s own characteristics, frequency components, etc. and also that cardiocycles relating to different clusters possess various morphology [5]. On the basis above the described informative feature’s signs have been chosen:
- QRS complex duration;
- Rise speed of a signal;
- QRS complex power in the range of frequencies corresponding to filter bandwidth;
- Global nonzero maximum in a frequency range;
- Area of the QRS complex.

2.4 Clustering phase
At this stage ECG cardiocycles clustering is made. The phase of a clustering is shown in Figure 3.

![Figure 2 – ECG signal before and after preprocessing: (a) Raw ECG signal, (b) filtered ECG signal.](image)

![Figure 3 – Schematic representation of intermediate steps for clustering phase.](image)
Firstly, the informative features were clustered with the help of Kohonen's self-organizing map [3]. At this stage, a Kohonen hexagonal network with a dimension of $3 \times 4$ neurons was initially formed. The number of learning epochs was calculated by the formula (1).

$$N_{\text{epoch}} = N_{\text{neuron}} \times 1000,$$

where $N_{\text{epoch}}$ – is the number of neural network learning epochs, $N_{\text{neuron}}$ – is the number of neurons in the Kohonen self-organizing map.

Secondly, the calculation of the optimal number of clusters of QRS complexes in the ECG signal was performed. To solve this problem, the Elbow method [4] was used;

Thirdly, using the k-means method, the clustering function (the result of this method) was received, on the basis of which clustering of QRS-complexes was carried out [5]. At this stage of clustering neuron weights of the Kohonen self-organizing map, previously trained, is held.

For the study, fragments from the MGH/MF Waveform Database were used, containing 48 annotated signals for three standard leads with a duration of 60 seconds at a sampling frequency of 250 Hz. Since the problem of blind clustering was solved using the proposed clustering algorithm, the training sample was not used (the neural network training occurred directly on the test data) [6].

### 3 Testing and results

For the calculating of algorithm’s accuracy indicators were used:

Algorithm accuracy indicators can be estimated by calculating $\text{Acc}$ – a probable assessment of correctness cluster determination of cardiocycles of ECG (2):

$$\text{Acc} = \frac{P}{N} \times 100\%;$$

where $P$ is the correct detection of class QRS-complex; $N$ – number of QRS-complexes [7].

### Conclusion

The accuracy indicators of this algorithm is $\text{Acc} = 96.92\%$. Further increase of efficiency of this algorithm is possible due to the use of bigger quantity of informative features; changes of method which have been the basis for the algorithm; changes in preliminary signal processing procedure; changes in output data interpretation procedure.

### References

[1] Hampton J.R. The ECG made easy. London: Churchill Livingstone, 2003.

[2] Van G.V., Podmasteryev K.V. Algorithm for detection the QRS complexes based on support vector machine //Journal of Physics: Conference Series. – IOP Publishing, 2017. – Т. 929. – №. 1. – С. 012041.

[3] T. Kohonen, Self-Organizing Maps (Third Extended Edition), New York, 2001.

[4] Robert L. Thorndike. Who Belong in the family? // Psychometrika, 1996, Vol.18, P. 267-276.

[5] MacQueen, J. Some methods for classification and analysis of multivariate observations.// Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967, Vol. 1, P. 281-297.

[6] Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., and Stanley, H. E. 2000. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 101, 23, e215--e220.

[7] GOST R IEC 60601-2-51-2011. Medical Electrical Equipment. Part 2–51. Particular Requirements for Safety, Including Essential Performance, of Recording and Analysing Single Channel and Multichannel Electrocardiographs.