ECG Data Mining Using SAS

Gulnar Kuatbayeva
LLC Institute of sustainable developmen research
Dr. In economics (RF MSU and RK)
Almaty,Kazakhstan
isdrceo@gmail.com

Ryskhan Satybaldiyeva
Information system
International information technology university
Almaty, Kazakhstan
0000-0002-0678-7583

Zarina Syrymbet
BSc. IS (Information System)
International information technology university
Almaty,Kazakhstan
lady.zazaza00@gmail.com

Akmaral Kuatbayeva
Information System
IITU
Ph.D. in CS assistant-prof. IS department
Almaty,Kazakhstan
ahamala2017@gmail.com

Meruert Mankibayeva
pcg on pvc
City Cardiology Center
laboratory assistant
Almaty,Kazakhstan
erzatovam79@mail.ru

Abstract — ECG analysis is an important part of finding devices, especially in the cardiovascular system. Various parameters obtained with the ECG curve, such as P-R, Q-T, R-R, Q-R-S, and S-T intervals, as well as heart rate variability, serve as safety and efficacy endpoints for many clinical trials. There are different algorithms based on graphs and frequency domains to detect different waveforms and ECG cells [4]. This article discusses the study and determination of ECG using SAS. ECG data, which are used for analysis and articles in clinical, are usually processed. The study of the algorithm of ECG signs using SAS allows you to flexibly and clearly perform a preliminary analysis of the parameters obtained from the ECG. In addition, it allows the programmer or researcher to analyze the same data from a research point of view.

Keywords—ecg, signal, peak, detection, data, analysis heart, rate, variability.

I. INTRODUCTION

An electrocardiogram (ECG) is the most commonly recorded physiological signal. It is an important component of all diagnostic tests related to heart health. The ECG signal contains information about the electrical activity of the heart. A number of ECG parameters are also used as primary and secondary endpoints in drug and drug testing. The ECG signal has a characteristic pattern (as shown in Figure 1) that is repeated with each heartbeat. Of the many morphological markers of the ECG, the most important are the QRS complex and the R peak - and the contribution of the R-peak in the interval R to R is the driving factor [3]. The number of R peaks in a certain time interval is translated into heart rate (beats per minute).

Automated detection of R-peak on the ECG has received widespread attention due to its use in Holter ECG recorders, which are necessary to ensure reliable performance in poor signal conditions [5]. A number of time and frequency domain methods have been proposed to detect the R-peak in the ECG. Although MATLAB, R, C, and other languages are regularly used to perform signal analysis in research environments, SAS is not widely used for signal analysis. In this work, the algorithm developed by Pan & Tompkins is implemented in SAS.

SAS offers an interactive matrix language (IML) procedure that allows users to use matrix-based data elements, operators, and define functions that run on matrices [4]. This ensures the use of any matrix-oriented language as part of the IML procedure.
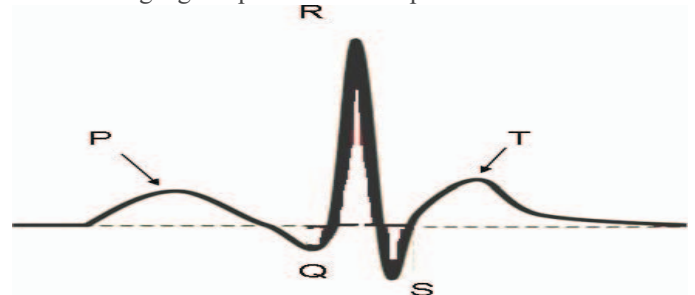

Figure 1: Electrocardiogram (ECG) Waveform for one normal heartbeat showing P, QRS and T waves

II. DATA OVERVIEW

The MIT-BIH arrhythmia database contains ECG signals obtained from lead II and lead V5 from subjects studied by the BIH arrhythmia laboratory. Signals were obtained at a rate of 360 samples per second per channel with a resolution of 11 bits in the range of 10 mV [1]. In addition to ECG signal files, the database also contains signal annotations. Two or more cardiologists independently comment on each signal [2]. Disagreements among cardiologists were resolved to obtain computer-readable reference annotations for each stroke [7]. These annotations are accepted as a reference for evaluating the implementation of the algorithm for detecting peak R. Signals and notes can be downloaded in ASCII format or in binary format. ASCII files are used here for convenience. ECG leads II are used to develop the algorithm. Ten one-minute ECG II clips are used to develop and evaluate the program. Figure 2 below shows an excerpt of the ECG signal from one of the clips.

III. SAS PROGRAMMING OVERVIEW

The difference equation, the integral of the window to be moved, and the threshold value described in the previous section are implemented using the IML procedure. Graphs are formed using the GPLOT procedure.

A. NORMALIZATION

The required data is normalized by subtracting each data point by its average value and dividing by the range.

B. DIFFERENCE EQUATIONS

Normalized ECG data is transmitted through a bandpass filter and then through a derivative filter. The difference equations for these filters are given in equations 1, 2, 3 and 4 in the previous sections. These difference equations are implemented in SAS using the IML procedure.

The difference equations for high-pass and derivative filters are implemented similarly. The output of the bandpass filter is subjected to the squaring function.

Note: Signal values at negative time are taken as zero.

```
PROC IML;

READ all var {ecg_d} into ecg;

filt=j(nrow(ecg),1,0);
filt[1,1]=ecg[1,1];
filt[2,1]=2*filt[1,1]+ecg[2,1];
DO i=3 to nrow(ecg);
IF i<7 then
filt[i,1]=2*filt[i-1,1]-filt[i-2,1]+ecg[i,1];
IF i>=7 & i<13 then
filt[i,1]=2*filt[i-1,1]-filt[i-2,1]+ecg[i,1]-2*ecg[i-6,1];
IF i>=13 then
filt[i,1]=2*filt[i-1,1]-filt[i-2,1]+ecg[i,1]-2*ecg[i-6,1]
+ecg[i-12,1];
END;
varnames = {"lowpass"};
CREATE filtered FROM filt [colname=varnames];
APPEND from filt;
CLOSE filtered;
QUIT;
```

C. MOVING WINDOW INTEGRAL

The integral of the movement window is realized by using nested DO cycles, as shown in Figure 5. Note. Signal values at negative time are considered zero.

```
PROC IML;
USE ecgdata.ecg_squared;

READ all var {squared} into ecg;
filt=j(nrow(ecg),1,0);
DO i=1 to 32;
sum=0;
DO j=i to 1 by -1;
sum=sum+ecg[j];
END;
filt[i]=sum/32;
END;
DO i=33 to nrow(ecg);
sum=0;
DO j=i to i-32 by -1;
sum=sum+ecg[j];
END;
filt[i]=sum/32;
END;
varnames = {"integral"};
CREATE integral FROM filt [colname=varnames];
APPEND from filt;
CLOSE integral;
```

D. PEAK DETECTION AND THRESHOLDING

The peak of the integral waveform of a moving window is determined by determining the change in inclination. The waveform is sought to reduce the slope, and the moment at which the slope becomes negative and remains negative for nine consecutive data points is considered a peak [6]. The slope is calculated by the following formula:

Slope = x (n) - x (n-1)

The time corresponding to half the peak amplitude corresponds to the QRS complex. This moment is detected by finding a window with 40 samples to the peak.

IV.   RESULTS AND DISCUSSION

The results of each block in Figure 2 are presented below.

The ECG signal after normalization is shown in Figure 2. Note that normalization limits the signal range to (1 to +1). This reduces patient variability and helps standardize the thresholds used in the algorithm. Note the presence of high-frequency noise, which is usually caused by interference from power lines and muscle noise.
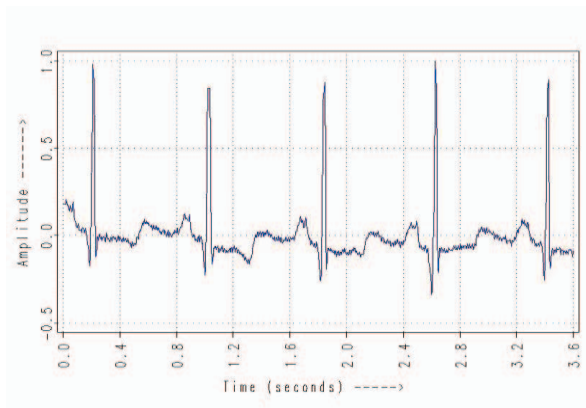

Figure 2: ECG data after normalization

Figure 3 shows the ECG signal after low-frequency and high-frequency filtering. High frequency noise is no longer present in the filtered signal.
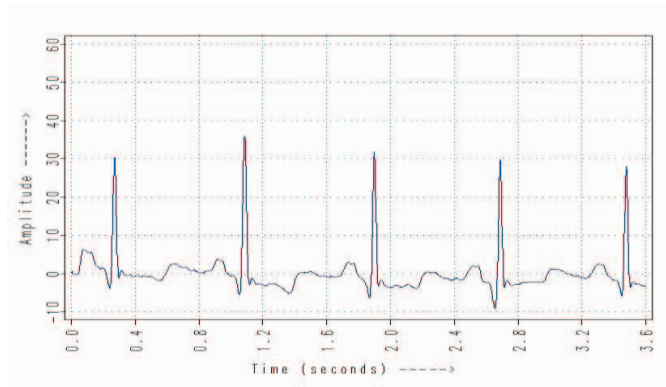

Figure 3: ECG data after High-pass filtering

Figure 4 shows the ECG filtered by the bandpass, after using the function of filtering derivatives and squaring. Note that the P and T waves are attenuated and the peak-to-peak signal corresponding to the QRS complex is amplified.
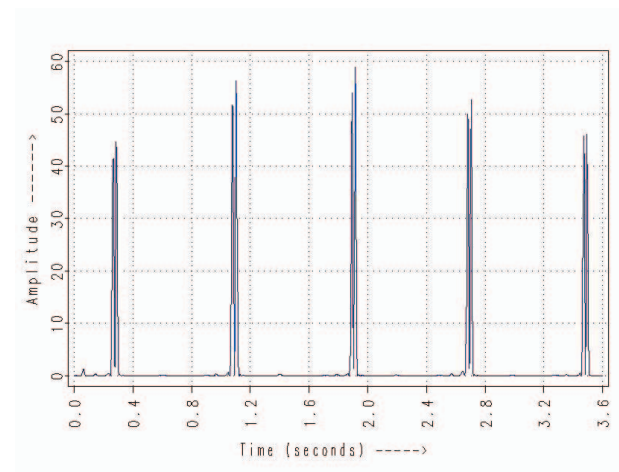

Figure 4: Band-pass filtered ECG after subjecting to derivative filtering and squaring function

Figure 5 shows the output of the integrated stage of the moving window. This signal is noiseless and has a waveform similar to a rectangular wave, which reaches a maximum when the QRS complex occurs.
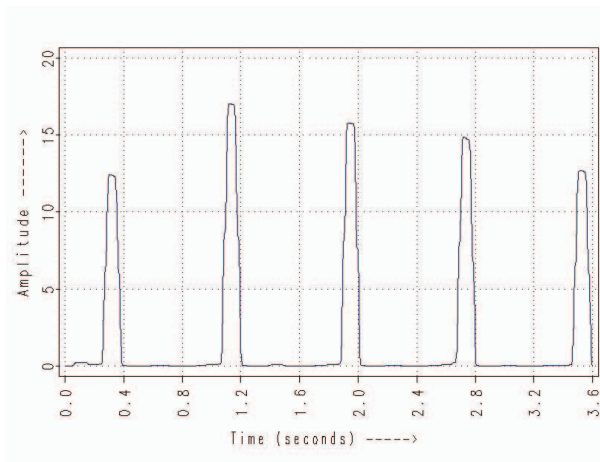
Figure 5: Output from Moving Integral Filter

V. CONCLUSION

The main focus of this work is to demonstrate that the analysis of ECG signals and other physiological signals can be successfully performed using SAS with relative ease. This approach can be extended to develop automated and semi-automated image processing algorithms that are used to quantify tumors and plaques in cancer and cardiovascular studies. Using SAS / IML to analyze and detect signals can be very useful in scenarios where statistical analysis has to be performed on multiple parameters derived from the signal of interest. Using SAS alone can help researchers focus on analysis and avoid data movement between different tools, one for feature detection and the other for statistical analysis (SAS). Using SAS alone also helps reduce costs.

In clinical trials, ECG data have already been pre-processed. Perhaps with the help of SAS and raw data, this allows to expand research and analysis. The macro can calculate the R-R interval, heart rate variability and instantaneous pulse according to ECG data.

REFERENCES

Signals were obtained at a rate of 360 samples per second per channel with a resolution of 11 bits in the range of 10 mV [1]. Two or more cardiologists independently comment on each signal [2]. Of the many morphological markers of the ECG, the most important are the QRS complex and the R peak - and the contribution of the R-peak in the interval R to R is the driving factor [3]. SAS offers an interactive matrix language (IML) procedure that allows users to use matrix-based data elements, operators, and define functions that run on matrices [4]. Automated detection of R-peak on the ECG has received widespread attention due to its use in Holter ECG recorders, which are necessary to ensure reliable performance in poor signal conditions [5]. The waveform is sought to reduce the slope, and the moment at which the slope becomes negative and remains negative for nine consecutive data points is considered a peak [6]. Disagreements among cardiologists were resolved to obtain computer-readable reference annotations for each stroke [7].

[1] *Task force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. "Heart Rate Variability: Standards of measurement, physiological interpretation, and clinical use", European Heart Journal; 17; 1996; p. 354-381*

[2] *De Chazal P et. al, "Automated Processing of Single-Lead Electrocardiogram for the Detection of Obstructive Sleep Apnea", IEEE Transactions on Biomedical Engineering; 50(6); 2003; p. 686-696.*

[3] *Benitez D et. al., "The use of the Hilbert Transform in ECG signal analysis", Computers in Biology and Medicine; 31; 2001; p. 399-406.*

[4] *Friesen GM, Jannett CT, et. al., "A Comparison of the Noise Sensitivity of Nine QRS detection Algorithms"; IEEE Trans. of Biomedical Engg., Vol. 37; No. 1, Jan 1990; p. 85-98.*

[5] *Hamilton PS, Tompkins WJ: Quantitative investigation of QRS detection rules using the MITiBlH arrhythmia database. IEEE Trans Biomed Eng BME- 33: 1157, 1986.*

[6] *Pan J, Tompkins WJ: A real-time QRS detection algorithm. IEEE Trans Biomed Eng BME-32:230.1985.*

[7] *Moody GB, et. al, "PhysioNet: a Web-based resource for the study of physiologic signals", IEEE Engineering in Medicine & Biology Magazine; 20 (3); 2001; p. 70-75.*

[8] *Uskenbayeva, R.K., Kuandykov, A.A., Rakhmetulayeva, S.B., Bolshibayeva, A.K. Properties of platforms for the transformation and automation of business processes, International Conference on Control, Automation and Systems, 2019.*

[9] *Uskenbayeva, R., Kuandykov, A., Kalpeyeva, Z., Kassymova, A. Formalization of Applications for Processing in the e-Commerce System 2019 Proceedings 21st IEEE Conference on Business Informatics, CBI 2019*

[10] *Modelling situational room for healthcare Kuatbayeva, A.A. 2014 World Applied Sciences Journal*

[11] *Fuzzy logic in healthcare situation room modelling Kuatbayeva, A.A. 2014 ACM International Conference Proceeding Series*

[12] *Kuatbayeva G.K. Ecological-economic formation in education potential for sustainable development*