Testing the Quality of the Algorithms for Automated Electrocardiogram Analysis

Piotr Augustyniak[1] , Dariusz Miszczak[2]

[1]Institute of Automatics, University of Mining and Metallurgy,
al. Mickiewicza 30, 30-059 Kraków

[2]Zaklady Elektroniki Medycznej ASPEL - ZABIERZÓW S. C.
os. Sienkiewicza 33, 32-080 Zabierzów

ABSTRACT

The study presents a method for testing the quality of automated electrocardiogram analysis software. The described method uses the cardiological databases which are worldwide standards and makes it possible to both carry on development of the software and compare the quality of cardiological software. This method can be a qualification criterion for approval of analysis software.

1. STANDARD ECG DATABASES

An ECG database is a set of test signals which contains most of the typical heart dysfunctions observed in electrocardiography. Signals are series of samples representing electrocardiographic records performed using different leads and different duration times (10s ... 24 h). As a diagnostic reference, a standard database also contains results of a detailed analysis of the represented signal performed independently by several cardiological centers, sometimes in the light of other analyses results for a particular patient. The database contains signals recorded at different heart rates (typically 20 ... 210/min) also differing in quality, that means the distortion level, isoelectric line variations, etc. This allows to evaluate the operation of the tested software in different recording conditions. The cardiological database is not only an objective representation of the current state of knowledge in electrophysiology of the heart but also a set of signals observed in practical ECG recording.

The MIT-BIH [1], [2] developed in the 70'ies is at present the most popular database. It contains 8 groups of records for basic heart pathologies. This database contains records of a duration time of 15 min to 24 h, using two electrodes. The signal description includes the QRS complex (ventricular complex) occurrence moment (No of the sample), the type of the complex and eventually additional remarks. The aim of this database is to test the so called holter systems for long-term heart function monitoring and to classify the pathologies which are incidentally arising in the record.

In 1990, CSE [3], [4] developed a new ECG database, which is a set of test signals representing records of the most known heart pathologies. The records are performed using a standard set of 12 leads, completed by the pseudoorthogonal Frank leads, which allows testing software for electrocardiography and vectorcardiography, selection of the optimum leads, etc. The database contains 250 real and 310 model signals, arising from duplication of typical heart evolutions. The duration of each record is 10 sec. The description of the represented signal includes fiducial points of the starting and ending points of the heart evolution waves that are typical for a given record, the type of QRS complex, applied drugs, etc. The descriptions of the signals provided in the files are completed by a so called CSE Atlas presenting the graphical shape of all the ECG records. The task of the CSE is to test the supporting software for clinical ECG diagnostics installed commonly in modern electrocardiographs.

2. SUPERVISION OF THE AUTOMATED ECG DIAGNOSTICS

Performing certain diagnostic tasks with the help of adequate software undoubtedly simplifies the doctor's work, though one should be aware of the consequences arising from automated interpretation errors. To protect the patient against the consequences of a false diagnosis and also the physician from the responsibility for a false interpretation of the electrocardiogram, a strict survey of the equipment and also of the software commissioned for applications in health services is a necessity. The role of the qualification criterion for the software can be played by the method of testing the quality of the algorithm for automated electrocardiogram analysis described in the present paper.

3. EVALUATION OF THE QUALITY OF THE ALGORITHM FOR AUTOMATED ELECTROCARDIOGRAM ANALYSIS

Modern software supporting the ECG diagnostics usually perform the following numerical procedures:

- detection of the QRS complexes
- classification of the QRS complexes (sometimes also of whole heart evolutions)
- determination of the fiducial points of the electrocardiogram

Evaluation of the quality of the algorithms for automated electrocardiogram analysis is carried out by applying the tested software for determining the diagnostic parameters of a database test signal and then comparing the obtained results with the real values recorded in the description. This apparently simple procedure requires dealing with a few problems:

- The software should be evaluated by a manufacturer-independent institution, provided the producer's rights for confidentiality. The main issue is to perform an evaluation of the software output, without assessing the source-code. That requires an initial determination of the format and range of the input data, intermediate parameters being evaluated and the diagnostic parameters which are the final output
- The test signals represented in the databases should be adapted to the assumed operating conditions of the evaluated software. Usually it is necessary to rescale the amplitude and resample both the signal and the real values from the database description. It is significant to carry it out without disturbing the baseline nature of the test signals. The best solution, hence the most difficult to implement in practice would be adapting the evaluated software to the test signal (same sampling frequency) by the manufacturer
- Results of the tests should be presented to the manufacturer in a statistical form, which allows further application of the same signal set for objective evaluation of the next version.

4. PRESENTATION OF THE EVALUATION RESULTS

4.1 Evaluation of the QRS complex detector can be carried out based on the MIT-BIH or CSE databases, and the results are usually presented as three percentage values:
- number of correct results
- number of false positive results
- number of false negative results

4.2 Evaluation of the classification quality can be carried out based on the MIT-BIH or CSE databases. The results of this evaluation have been classed as follows:
- number of correct results,
- number of false negative results,
- number of false positive results that in fact belong to other classes, arranged in a decreasing order.

Such a classification allows to evaluate the specificity of a given class along with an eventual optimization of its characteristic features set as well as to determine the probability of a false classification and to indicate the classes are the most commonly mistaken.

4.3 Evaluation of the fiducial points determination quality is by nature a deviation in the value of calculation performed by the evaluated software using a test signal form the given parameter readout and description in the database. Adequate descriptions are only available in the CSE database.

The presentation of the evaluation results can be as follows:
- statistic specification of deviations in obtained results from the real values, given in number of samples for particular test signals (Fig. 1). The specification is presentedas a histogram, each column with the test signal numbers and values of the deviations is also indexed with the total number of signals having a given number of deviations, which is also graphically represented by the column's height. Such a specification allows to observe the distribution of the deviations and has development purposes (introduction of corrections in the following versions of the software).
- specification of the absolute values of the standard deviations obtained from the real values in [ms], in the decreasing order of test signals is presented in Figure 2. The test signals are arranged on the X-axis in the order of decreasing value of the deviation (without respect to the signal number). Each signal is related to the absolute value of the standard deviation calculated based on the set of test signals after elimination of "worse" signals (the ones on the left hand side of the axis). Such a specification allows to observe the sensitivity of the software to extreme situations (high level of interferences or variability in the isoelectric line) and is used for comparing two software products (two applications of two different manufacturers, two versions of the same software or an application with a reference).

5. CONCLUSIONS

The described method for evaluating the quality of automated electrocardiogram analysis algorithms has been applied by the authors to test the system software of the Ascard-3 electrocardiograph. The particular steps of the test are performed according to the procedures described in 3. and aim to systematically improve the available software. It is also planned to obtain a CSE attestation, that will allow homologation of this Polish-made electrocardiograph in many countries of the world.

-9	66
-8	
-7	94
-6	
-5	
-4	15 90
-3	26 42 82
-2	1 12 24 30 35 44 49 62 69 105
-1	4 6 7 8 13 14 22 29 31 34 38 46 47 51 65 68 71 72 74 96 97 99 107 119
0	3 11 17 19 25 36 37 41 53 57 59 61 75 76 78 81 83 84 85 87 91 95 106 116 117
1	16 28 32 33 55 77 80 86 88 102 110 114 122 123 124
2	2 20 27 101 125
3	113 121
4	
5	48 108
6	112
7	103
8	
9	
10	109

Fig. 1. Statistic specification of deviations in obtained results from the real values, given in number of samples for particular test signals. Exemple for P-onset fiducial point. First column: deviation value (in samples), next columns - corresponding signals numbers from CSE data base.

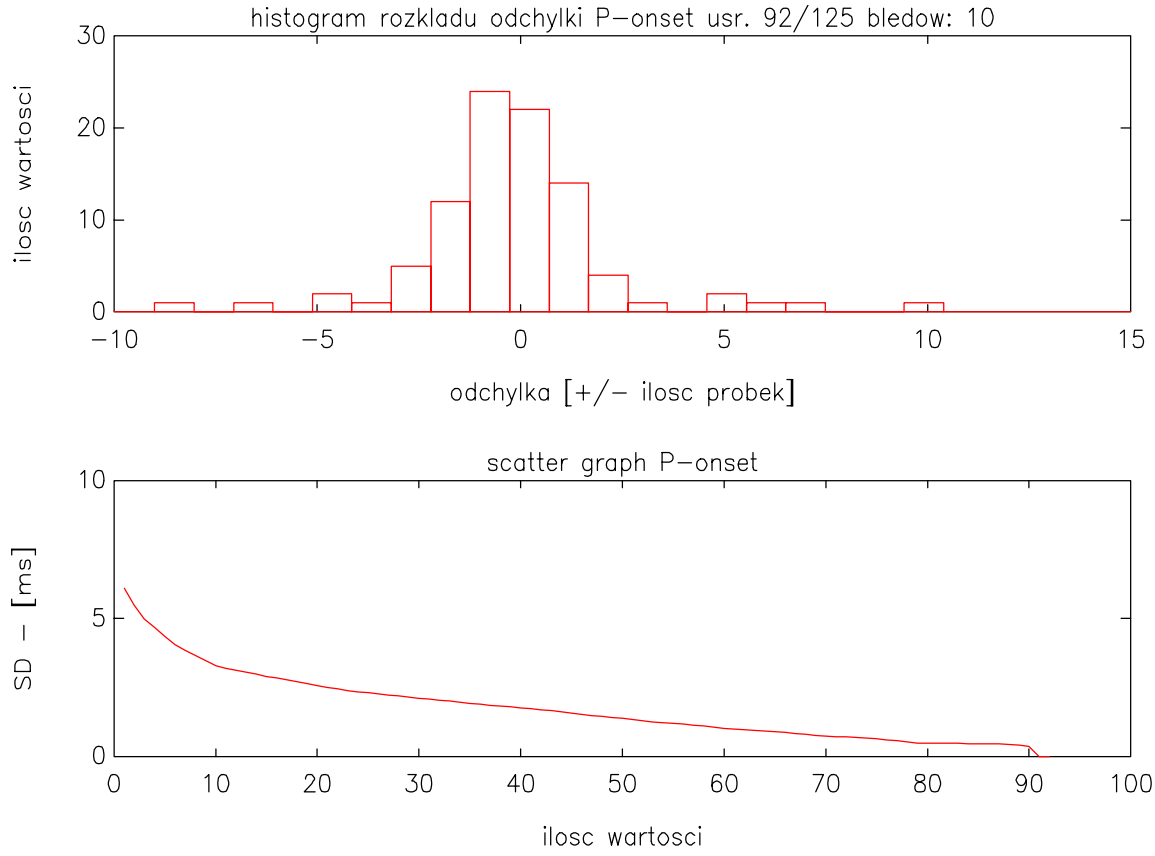

Fig. 2. Specification of the absolute values of the standard deviations SD obtained from the real values in [ms], in the decreasing order of test signals. Exemple for P-onset.

REFERENCES

[1] Comm. Electrocardiogr. Amer. Heart Assoc.: "Recommendation for Standarization of Leads and of Specifications for Instrument in Electrocardiology and Vectocardiology" Circulation, LII, 2 1975, pp. 11-31.

[2] Moody G., MarkR. "MIT-BIH Arrythmia Database Directory" Massachsetts Institute of Technology, Biomedical Engineering Center 1988.

[3] Willems J. L. "Common Standards for Quantitative Electrocardiography" 10th CSE Progress Report, 1990. Leuven: ACCO publ., 1990, 384p.

[4] Willems J. L., Arnaud. P., Van Bemmel J. H. ... "Assessment of the performance of electrocardiographic computer programs with the use of a reference data base" Circulation, 1985, Vol. 71 N° 3, p 523-534.