Models for Risk Stratification of Sudden Cardiac Death Based on Logical Conjunction and Decision Tree

Marek Kamiński, Rafał Kotas, Piotr Mazur, Bartosz Sakowicz, and Andrzej Napieralski

Abstract—**This article presents methodology of simple model building for risk stratification of sudden cardiac death based on an original ECG signal analysis platform. The method includes the analysis of the ECG signal, extraction of selected parameters and model optimization for predicting sudden cardiac death.**

Index Terms—**sudden cardiac death (SCD), ECG Holter monitoring, risk stratification model**

I. Introduction

CARDIOVASCULAR disease is the most common cause of death in Poland for over 50 years [1, 2]. Research from 2009 shows that this applies to about 45% of deaths (Fig. 1). Studies conducted in Europe also indicate cardiovascular disease as the main factor for mortality. Cardiovascular disease mortality in Eastern Europe is much higher than in the west. Poland is in the middle of this ranking, but it should be noted that compared to Western Europe there is much to be done in this direction. World Health Organization predicts that cardiovascular disease will be no. 1 on the list of causes of mortality until at least 2030. The average age of society has extended with the reduction in mortality due to cardiovascular diseases in Poland.

The main focus of cardiology research is now set to increasing the efficiency and availability of diagnostic methods. For example, sudden cardiac death is relatively easy to avoid through the implantation of a cardiac defibrillator. The problem remains, however, limiting patients to those which are actually threatened with sudden cardiac death. Effective diagnostic methods (adopted by the gold standard) are generally difficult and expensive and may not cover a broad spectrum of patients. Obviously there are performed much cheaper, standard tests but the results can hardly be regarded as conclusive in many cases. One of the possible directions of development of these techniques is a statistical analysis of the data received on the possible broad population of patients in order to develop the model and the observations appearing in the rules which could be used to develop new diagnostic techniques more effective.

M. Kamiński, R. Kotas, P. Mazur, B. Sakowicz and A. Napieralski are with Lodz University of Technology, Department of Microelectronics and Computer Science, Lodz, Poland (e-mails: kaminski@dmcs.pl, rkotas@dmcs.pl, pmaz@dmcs.pl, sakowicz@dmcs.pl, napier@dmcs.pl).

This project is financially supported by the National Science Centre under grant No. DEC-2011/03/B/ST6/03454, for which the authors express their gratitude.

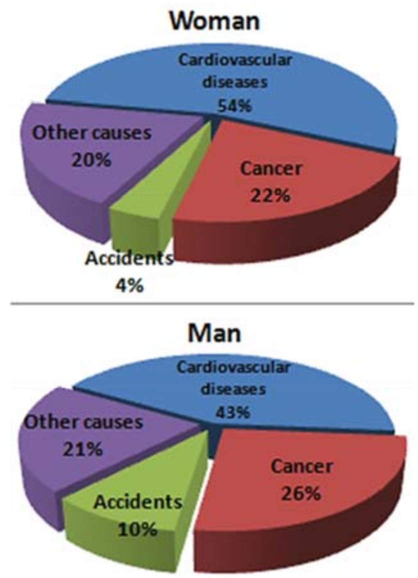

Figure 1. Share (%) of major diseases in the general mortality in Poland.

This paper describes the *Kardio* application developed under grant "Sudden Cardiac Death risk stratification based on functional assessment of autonomic nervous system with the use of Holter methods". The project is conducted at the Lodz University of Technology, Faculty of Electrical, Electronic, Computer and Control Engineering (WEEIA) at the Department of Microelectronics and Computer Science (DMCS). The main goal of *Kardio* application is to allow the basic statistical analysis of a large data in order to develop a model of sudden cardiac death (SCD) risk stratification. The project is scheduled for several years. Currently a simple version of the model similar to decision trees is implemented. This paper presents the concept of application architecture, the choice of model parameters, their generation, as well as example results of sample group of patients. Summary contains critical assessment of obtained results as well as future plans and directions of application development.

II. The Operation of *Kardio* Application

Kardio is a standalone application running on any popular operating system. It is developed in Java and all of the features are implemented as a separate plug-in. This choice was based on several advantages of plug-in architecture:

- Users can choose between the features that they want to have in the program, and which do not. (The program can be offered in the future to other centers involved in the processing of electrocardiography (ECG) signals);
- Each plug-in works independently;
- Easy to add new functionality (installation of a new plug-in);
- The tasks in the project could be easily distributed among many developers (with different competencies) by arranging for them to write plug-ins at different levels of complexity.

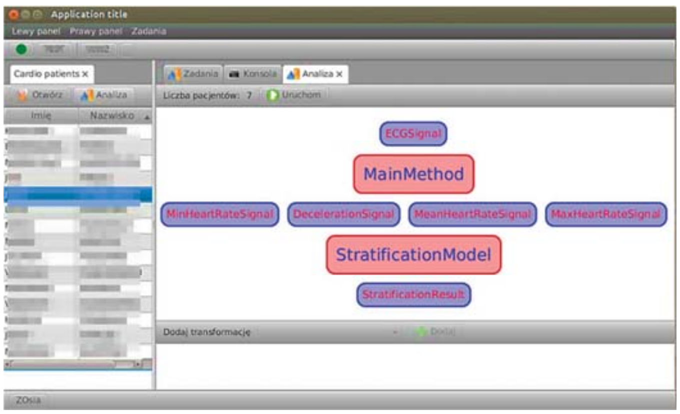

Figure 2. *Kardio* application interface.

Each plug-in is responsible for different function in the program. Plug-ins are divided into several modules, which contain similar functionalities. Fig. 3 shows a block diagram of the application. Number of included plug-ins can obviously increase significantly.

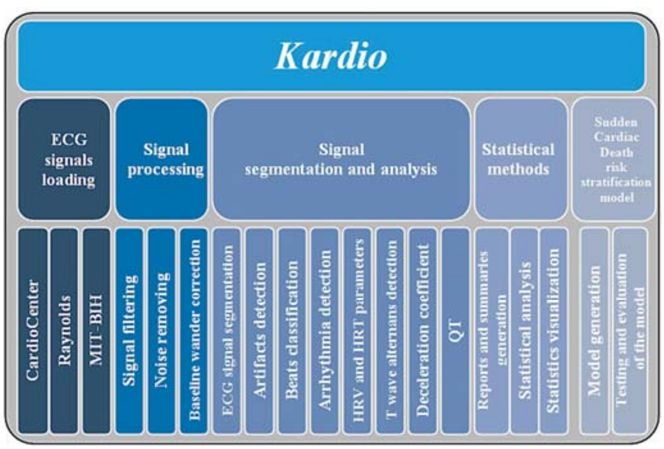

Figure 3. A block diagram of *Kardio* application.

The main groups of plug-in's functionalities:

- loading ECG signals from the available files (different formats and to read a variety of information), the authors previously developed plug-ins to analyze the 8-bit Holter ECG signal obtained from the CardioCenter and Reynolds;
- ECG signal filtering;
- ECG segmentation for each heartbeat;
- ECG waveform parameter calculations (parameter is understood in a broad sense);
- statistical analysis;
- generation of risk stratification model.

Cooperation between the plug-ins is provided by the adopted scheme of plug-in development. Application engine is responsible for linking plug-ins in accordance with the declared rules (Fig. 5). Each plug-in declares the source and destination signals (Fig. 4).

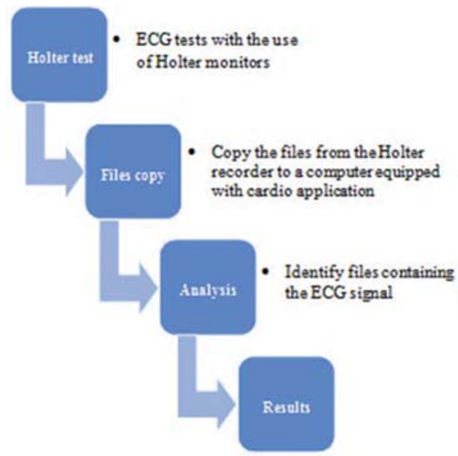

Figure 4. *Kardio* –operating diagram.

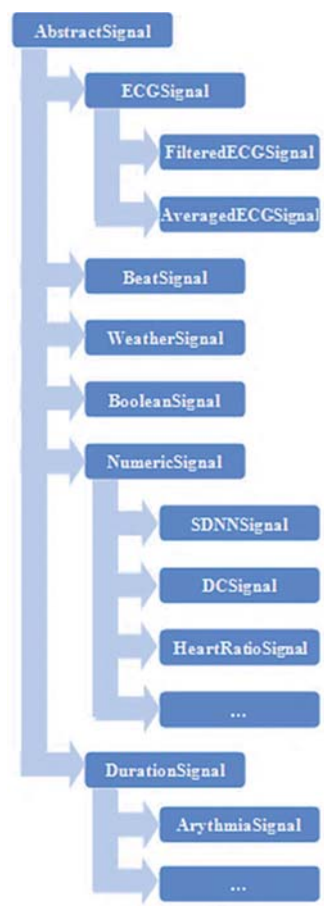

Figure 5. Signal hierarchy in application.

III. Analyzed Parameters

At the beginning of the *Kardio* platform development, the authors have chosen a set of parameters calculated from the ECG signal for the stratification model generation. This choice was based on the opinion of cooperating doctors supported by literature studies.

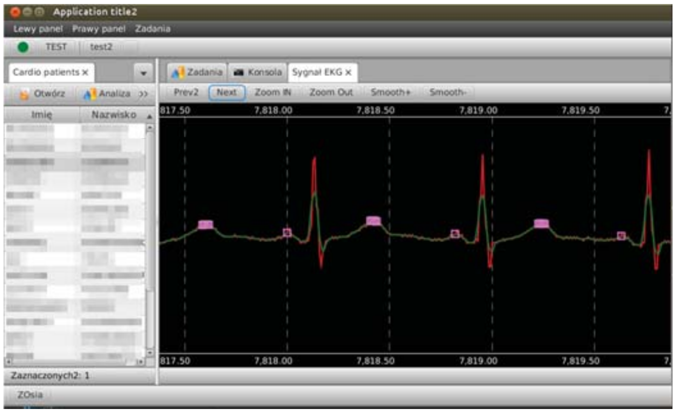

Figure 6. ECG signal visualization.

For SCD risk stratification authors use a set of three markers based on ECG signal: HRT, HRV and DC/AC.

Heart rate turbulence (HRT) is a physiological, two-phase response of sinoatrial node to premature ventricular contraction (PVC). This response consists of short heart rate acceleration phase, after which the heart rate slows to a normal pace. To assess HRT two separate parameters are used: turbulence onset (TO), defined as percentage ratio of heart rate before PVC to the heart rate right after PVC and turbulence slope (TS) which is defined as the most steep regression line trough each 5 consecutive distances in normal sinusoidal rhythm in ECG.

Heart rate variability (HRV) defines short-term time differences in duration of consecutive RR distances in sinusoidal rhythm registered in ECG signal [3, 4]. The most acknowledged parameter is standard deviation of all normal-to-normal RR intervals (SDNN), which value below 50 ms is considered abnormal. It was verified in numerous publications that lowered HRV is an independent risk factor in patients with myocardial infarction.

Deceleration capacity (DC) and corresponding acceleration capacity (AC) provide a quantitative assessment of heart rhythm capacity to decelerate and accelerate respectively [7]. Parameter values are calculated using phase-rectified signal averaging (PRSA) method, after synchronization of all periodic oscillations of RR intervals excluding all non-periodic distortions (artifacts, noise, etc.).

In addition to the set of parameters authors included a number of other parameters which are not found to be predictive. However the platform calculates them (for other purposes) and allows to test them as parameters of the model: T wave alternans (TWA), mean heart rate (HR), mean QRS duration (length of three waves of ECG), and mean QTc (a corrected QT interval).

IV. Risk Stratification Model

An essential element of the application is the possibility of risk stratification model generation. The model is also implemented as plug-in so the author has many possibilities of the development. The authors plan to implement several different versions of the model. The paper presents the first and the simplest concept based on the optimization of the traditional model.

A. Idea

Described markers are used to assess the risk of occurrence of life-threatening cardiac events. Based on their values a risk stratification models are designed, usually consisting of threshold values for parameters above (or below) which the risk is greater. For example turbulence slope (TS) below 2.5 ms or deceleration capacity (DC) below 3.5 ms indicate a high risk. Model efficacy is assessed basing mostly on: positive predictive accuracy (PPA), negative predictive accuracy (NPA), sensitivity, specificity; calculated for control group of patients.

It should be noted that currently used models are relatively simple (based on one or at most two indices), and their parameters are chosen arbitrarily by the medical staff. Predictive model of opportunities depend mainly on the size of the control group. [5] describes a study based on a group of 2343 patients and [6] on a group of 1200 patients.

Comparison of several models shows that the selection of their parameters can have a significant impact on their ability and predictive value. According to the literature it should not be considered final and correct for each group of patients. For analysis of deaths traditional model has proved to be far better for each term studies. In the case of prediction of other models of cardiac events with increased cut-off values were better (Fig. 7).

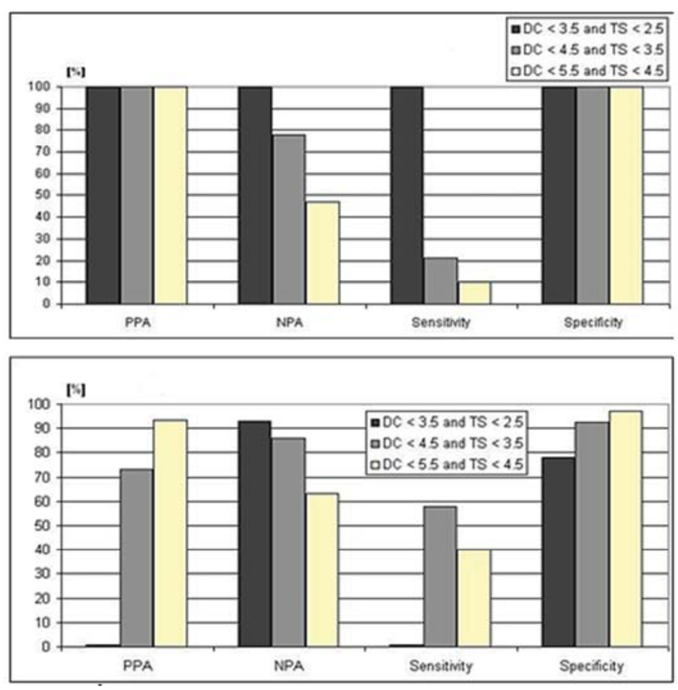

Figure 7. Prediction of mortality (above) and prediction of cardiovascular hospitalization (below).

Similarly, a model based on a combination of the two indicators almost always exceeded the model based only on a single indicator. Dependencies are shown, of course, very modest. The reason is of course a small group of patients in the study, for which the number of cardiac events was very limited. However, you can venture to say that the construction of platforms, opportunities to broaden the analysis and patient groups will help to develop a comprehensive, reliable and useful in clinical risk stratification models.

Observed differences in parameters: positive predictive value (PPV), negative predictive value (NPV), sensitivity and specificity depending on the assumed model parameters vary quite significantly. The natural solution is to use optimization methods to find the model with the best predictive properties. Number of parameters used in the model can be arbitrarily changed.

B. Objective Function

The simplest version of the model resembles the structure of the decision tree. Each parameter divides the group into two subgroups of patients (considered as a group of normal and high risk). Parameter that changes in this case, is the cut-off value. In general the direction of inequality (<; >) can also be variable, but in order to accelerate the calculations it was decided to pre-select the type of inequality (based on the one-parameter model) and in the case of multiparameter model take the fixed type. Conditions can be joined by different logic functions, and in the case of the presented example AND function is selected.

Knowing the fate of patients after the conducted Holter test they can be divided according to assumed criteria into four groups:

- TP - true positive results;
- FP - false positive results;
- FN - false negative results;
- TN - true negative results.

Predictive ability of the model is characterized by four standard parameters:

1) Sensitivity – describes the ability to detect people actually sick (with a particular trait). So, if we examine a group of patients, the sensitivity gives us information about what percentage of them have tested positive (1).

$$sensitivity = \frac{TP}{TP + FN} \tag{1}$$

2) Specificity - describes the ability to detect people actually healthy (without the particular trait). So, if we examine a group of healthy subjects, the specificity give us information about what percentage of them have a negative test result (2).

$$specificity = \frac{TN}{FP + TN} \tag{2}$$

3) Positive predictive value (PPV) - the probability that an individual has the disease given a positive test result. So if the test subject has received a positive test result, the PPV gives him information on how many can be sure that he is suffering from a disease (3).

$$PPV = \frac{TP}{TP + FP} \tag{3}$$

4) Negative predictive value (NPV) - the probability that the individual did not have the disease having a negative test result. So if the test subject received a negative test result of the test, the *NPV* gives him information on how many can be sure that he do not suffer from the disease (4).

$$NPV = \frac{TN}{TN + FN} \tag{4}$$

Optimizing the model only for one of these parameters, is obviously meaningless. The parameter calculated on the basis of two opposing parameters from the list should be assumed as an objective function. Two examples of the most commonly used objective functions:

5) Likelihood ratio of a positive result, (LR+) - a measure that allows to compare the results of several tests fit to the gold-standard and is not dependent on the prevalence of the disease. It is *the* quotient of two opportunities: the chance that a positive test result will get a person with a group of sick patients to the chance that the same effect would be observed among healthy individuals. (5)

$$LR+ = \frac{sensitivity}{1 - specificity} \tag{5}$$

6) Likelihood ratio of a negative result, (LR-) - it is the quotient of two opportunities: the chance that a negative test result *will* get a person with a group of sick patients to the chance that the same effect would be observed among healthy individuals (6).

$$LR- = \frac{1 - sensitivity}{specificity} \tag{6}$$

If in the studied group, the group of actually sick patients is not too large it can often lead to the situation that the number of false-positive cases is zero. This leads to the maximum specificity (specificity = 1) and LR + (infinity). Dealing with such a case authors decided to change the objective function. It is now a combination of both type of reliability ratios (7). Zero in the denominator of the function indicates the ideal model (sensitivity = 1, specificity = 1). The main goal of this optimization is maximizing the value of the objective function.

$$objectiveFunction = \frac{sensitivity \cdot specificity}{2 - sensitivity - specificity} \tag{7}$$

C. Optimization

Presented optimization was based on the brute force method. Although this method has the worst time parameters, however several arguments explain its use:

- the number of parameters is limited to a few. More complex models would rather not find acceptance by doctors;
- measurement of ECG parameters is burdened with quite a significant error, and low accuracy. For example, DC parameter measured between the various channels of signal differs by the average of 0.3 ms. Assuming that DC is in the range from 2 to 13 ms, only about 40 possible division values of the group;
- from the user point of view model generation time is insignificant. Gathering a large group of patients may take a few years (therefore the authors decided to use historical data). Before the model would serve as a diagnostic tool it has to be precisely verified. For this reason, the more profitable is to check all the possibilities for detecting minimum (possibly considering several alternatives) rather than taking the risk of finding only a local minimum;
- brute force method is ideal for parallelization of calculations.

It is obvious that the optimization time dependence of the number of parameters is exponential. In the interests of authors (up to 3 parameters in the model) brute force optimization time is negligible.

D. Problems and development directions of stratification model

This paper presents the preliminary concept of Kardio application development The results obtained so far have shown that the proposed solution brings a number of problems which should be solved in the future:

1) The form of the model

Classification is a method of data analysis for predicting the value of a specific attribute based on a set of test data. New model is based on the contents of the database. It is used to classify newly created objects in the database, or for a deeper understanding of the existing division of elements on a pre-defined class [8]. With many of the available concepts of classification models authors chose and implemented two solutions. Authors also plan the implementation of the third.

The first concept is a currently used approach based on the use of one or more parameters. Conditions for the individual parameters are combined using the logical conjunction. Geometrically, this corresponds to a cut in a single area of the search space which is characterized by an increased risk of disease (Fig. 8). Some novelty is the use of optimization to select the best method of division.

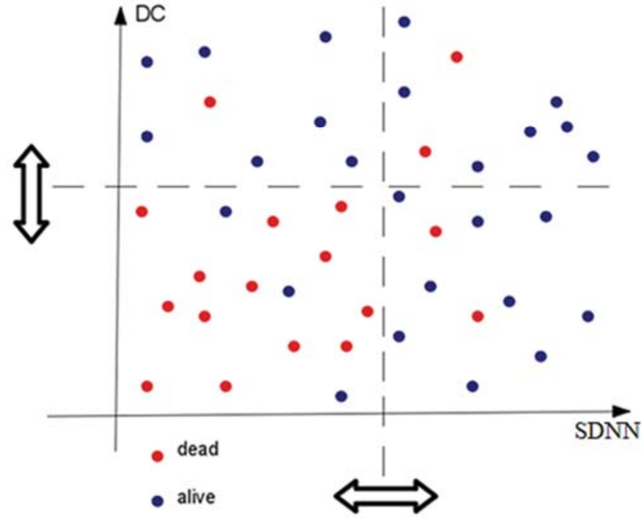

Figure 8. Risk stratification model based on logical conjunction.

The second concept is based on decision trees. Decision tree is a graph in which the vertices correspond to the tests – a comparison of the values of attributes, arcs to the test scores, and leaves to the classes. The apex of the tree is called the root of the tree. Each internal vertex of a decision tree consists of the division, which is responsible for the division of the data set to the appropriate partition (Fig. 9). As a partition can be understood the set of data belonging to one class, resulting from division of the training set.

Bayesian network was selected as the third type of classification model and it will be implemented in the future work [9, 10, 11].

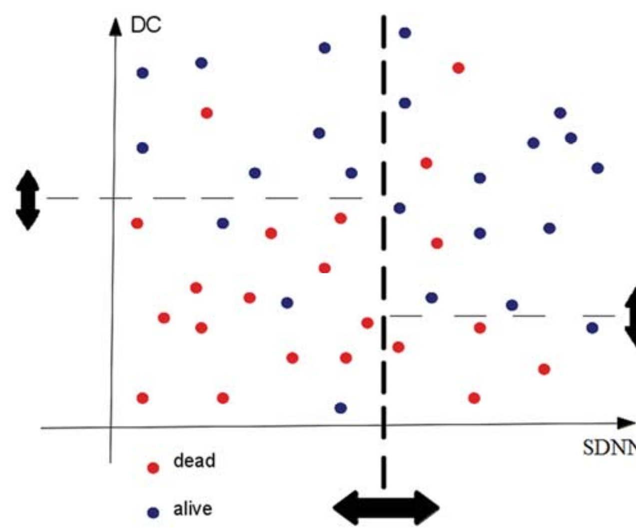

Figure 9. Risk stratification model based on decision tree.

2) The process of database creation

Collecting relevant database is an extremely difficult process. The authors have several thousands of ECGs collected over several years in the Wł. Biegański's hospital in Lodz. Documentation of patients, however, is incomplete (in the context of the fate of patients after completion of the study) and in paper version which is difficult to transfer to the database. In addition, most patients are treated under constant medical supervision and the nature of the risk is thereby disturbed. The ideal set of research would be a group of people subjected to ECG Holter observation which would have made no other medical activities. Creation of such a group of patients is completely unrealistic. For this reason, authors decided to find other than sudden cardiac death cut-off points (cardiac events). It would be preferred if such a phenomenon was possible to detect in the ECG signal. At the moment, ventricular tachycardia appears to be suitable cardiac event. The longer the duration of a threat to life can be considered as an interesting marker for the risk of death. The presence of ventricular tachycardia is ascertained exclusively on the basis of the ECG:

- increased RR frequency >100/min;
- the number of evolutions >3;
- QRS width >120ms;
- the first evolution of tachycardia occurs earlier than the expected evolution of the patient's own rhythm;
- conditionally - the opposite vector of maximum deflection of the QRS complex, especially the final section of the ST in relation to the evolution of sinus rhythm.

The change in the cut-off point for arrhythmia and broadening the base of patients is currently performed.

V. Example Results

A. Study Group

In order to verify the correctness of implemented algorithm a series of tests were performed on ECG recordings from a selected group of patients. The study enrolled a total of 73 patients surviving an acute myocardial infarction (MI) in sinus rhythm. Each patient had three 24-hour sessions of ambulatory ECG monitoring in the first week, 6th week and then 6th month after the MI respectively.

Twenty-four hour ECG recording and analysis were performed using the CardioScan equipment. The analysis included arrhythmia, heart rate variability (HRV) assessed by means of standard deviation of NN intervals (SDNN), heart rate turbulence (HRT) and deceleration capacity (DC) assessment. Heart rate turbulence analysis was performed using the CardioScan software. Only one parameter was assessed, namely the turbulence slope (TS).

Mean follow-up period was 24 months. Two endpoints were defined, the first comprised of mortality and the second comprised of cardiovascular hospitalization during 2-year follow-up. Patients were divided into three groups in accordance to defined endpoints (Table I):

- Group 1 – patients with no further cardiac episode during follow-up;
- Group 2 – patients who died during follow-up period;
- Group 3 – patients with further cardiovascular hospitalization during follow-up.

TABLE I.
Characteristics of each study group

	All	Group 1	Group 2	Group 3
Number of patients	73	53	5	15
Age [years]	57.36 ± 8.29	57.94 ± 7.90	64.80 ± 9.42	53.89 ± 7.75
Females	24.66 %	24.49 %	40.00 %	21.05 %

B. Results

Table II presents the results of the risk stratification model based on logical conjunction. Cut-off point was defined as cardiac events leading to hospitalization or death. Optimization was performed for parameters recorded 6 weeks after heart attack.

Table III presents the results of the risk stratification model based on decision tree. Cut-off point was defined as cardiac events leading to hospitalization or death. Optimization was performed for parameters recorded 6 weeks after heart attack.

TABLE II.
Results of the risk stratification model based on logical conjunction

Model	Objective function
SDNN < 105.00 ms	0.56
TS < 2.60 ms	1.24
DC < 3.41 ms	1.24
SDNN < 183.00 ms; TS < 2.60 ms	1.24
SDNN < 131.00 ms; DC < 4.33 ms	1.28
TS < 2.60 ms; DC < 3.74 ms	1.62
SDNN < 183.00 ms; DC < 3.41 ms; QRS > 80.24 ms	1.51
SDNN < 183.00 ms; TS < 2.60 ms; DC < 3.74 ms	1.62
TS < 2.60 ms; DC < 3.74 ms; QRS > 72.54 ms	1.74

TABLE III.
Results of the risk stratification model based on decision tree

Model	Objective function
SDNN = 88.10 ms; TS(L) < 1.81 ms; TS(R) < 2.20 ms	15.25
DC = 3.14 ms; TS(L) < 2.20 ms; TS(R) < 1.81 ms	20.67
TS = 2.12 ms; DC(L) < 2.12 ms; DC(R) < 3.14 ms	2.63

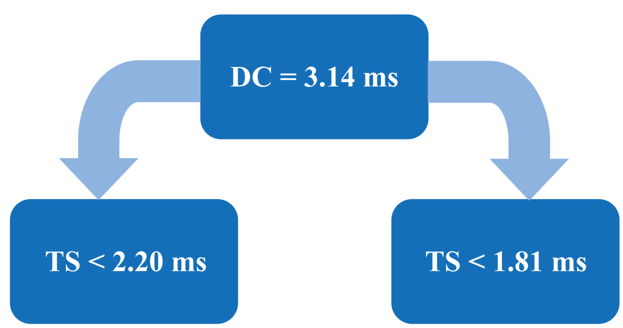

Figure 10. Model based on decision tree (calculated for parameters: DC, TS).

VI. Conclusions and Future Plans

Obtained results confirm the validity of the proposed method. Models based on the increased number of parameters appear to be better (although a little in this case) than one-parameter models. The method also allows you to choose parameters with the best predictive properties in a fully automated way.

Acknowledgment

Authors would like to thank the following employees of the Department of Cardiology, Medical University of Lodz: Professor Małgorzata Kurpesa MD, PhD; Urszula Cieślik-Guerra MD, without whose commitment and support this project would have no chance of realization.

References

[1] J. Szymborski, “Zdrowie publiczne i polityka ludnościowa.” [red.] Warszawa : Rządowa Rada Ludnościowa, 2012.

[2] A. Budaj, “Prewencja wtórna schorzeń sercowo-naczyniowych jako stały element kompleksowej rehabilitacji kardiologicznej.” 10/2008, Warszawa 2008

[3] H. Huikuri, T. Makikallio, K.E.J. Airaksinen, R. Mitrani, A. Castellanos, R.J. Myerburg, “Measurement of heart rate variability: a clinical tool or a research toy?” *J. Am. Coll. Cardiol.* 1999; 34: 1878–1883

[4] 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. Circulation, 1996; 93: 1043–1065.

[5] A. Bauer, P. Barthel, R. Schneider, K. Ulm, A. Muller, A. Joeinig, R. Stich, A. Kiviniemi, K. Hnatkova, H. Huikuri, A. Schömig, M. Malik, G. Schmidt, „Improved Stratification of Autonomic Regulation for risk prediction in post-infarction patients with preserved left ventricular function (ISAR-Risk)” European Heart Journal (2009) 30, 576–583

[6] G. Schmidt, M. Malik, P. Barthel, R. Schneider, K. Ulm, L. Rolnitzky, A.J. Camm, J.T. Bigger Jr, A. Schömig, „Heart-rate turbulence after ventricular premature beats as a predictor of mortality after acute myocardial infarction”; Lancet Vol. 353 April 24, 1999, pp. 1390-1396

[7] A. Bauer, J.W. Kantelhardt, P. Barthel et al., “Deceleration capacity of heart rate as a predictor of mortality after myocardial infarction”, Cohort study. Lancet, 2006; 367: 1674–1681

[8] D. Hand, H. Mannila i P. Smyth, Eksploracja danych, WNT, Warszawa 2005

[9] J. Pearl, “Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.” Morgan Kaufmann Publishers, San Mateo, 1988

[10] W. Tylman, „Detecting Computer Intrusions with Bayesian Networks”, Lecture Notes in Computer Science Volume 5788, 2009, ss. 82-91

[11] W. Tylman, G. Anders, „Application of probabilistic networks for decision support in power system analysis”, Energy 2006 Vol.31 Issue 14 ss. 2538-2553

Marek Kamiński was born in Łódź, Poland, on July 18, 1978. He received the M.Sc. and Ph.D. degrees from the Lodz University of Technology in 2002 and 2006, respectively. Currently he is with the Department of Microelectronics and Computer Science, Lodz University of Technology.

Rafał Kotas was born in Łódź, Poland, on August 31, 1985. He received the M.Sc. degree from the Lodz University of Technology in 2009 after defending his thesis entitled: "Data Glove Controlled by Microprocessor System". He is a Ph.D. student of the Department of Microelectronics and Computer Science at Lodz University of Technology since September 2009.

Piotr Mazur was born in Słupsk, Poland, on July 21, 1980. He received the M.Sc. degree from the Lodz University of Technology in 2005, respectively. Currently he is with the Department of Microelectronics and Computer Science, Lodz University of Technology.

Bartosz Sakowicz was born in Łódź, Poland, on November 23, 1976. He received the M.Sc. and Ph.D. degrees from the Lodz University of Technology in 2001 and 2007, respectively. Currently he is with the Department of Microelectronics and Computer Science, Lodz University of Technology.

Andrzej Napieralski received the M.Sc. and Ph.D. degrees from the Lodz University of Technology (TUL) in 1973 and 1977, respectively, and a D.Sc. degree in electronics from the Warsaw University of Technology (Poland) and in microelectronics from the Université de Paul Sabatié (France) in 1989. Since 1996 he has been the Director of the Department of Microelectronics and Computer Science. Between 2002 and 2008 he held a position of the Vice-President of TUL. He is an author or co-author of over 920 publications and editor of 20 conference proceedings and 12 scientific Journals. He supervised 48 PhD theses; six of them received the price of the Prime Minister of Poland. In 2008 he received the Degree of Honorary Doctor of Yaroslaw the Wise Novgorod State University (Russia).