Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 279–286
ISBN 978-83-60810-51-4

Predicting the presence of serious coronary artery disease based on 24 hour Holter ECG monitoring

Jan G. Bazan
Sylwia Buregwa-Czuma
Przemysław Wiktor Pardel
University of Rzeszów
Institute of Computer Science
ul. Dekerta 2, 35-030 Rzeszów, Poland
Email: bazan@univ.rzeszow.pl
sczuma@univ.rzeszow.pl
ppardel@univ.rzeszow.pl

Stanisława Bazan-Socha
Barbara Sokołowska
Jagiellonian University
Medical College
II Department of Internal Medicine
ul. Skawińska 8, 31-066 Kraków, Poland
Email: mmsocha@cyfr-kr.edu.pl
basiasok1@gmail.com

Abstract—**The purpose of this study was to evaluate the usefulness of classification methods in recognizing a cardiovascular pathology. Based on clinical and electrocardiographic (ECG) Holter data we propose a method for predicting a coronary stenosis demanding revascularization in patients with a diagnosis of a stable coronary heart disease. A possible solution of this problem has been set in a context of rough set theory and methods. The rough set theory introduced by Zdzisław Pawlak during the early 1980s provides a foundation for the construction of classifiers. From the rough set perspective, classifiers presented in the paper are based on a decision tree calculated on a basis of a local discretization method, related to the problem of reducts computation. We present a new modification of a tree building method which emphasizes the discernibility of objects belonging to decision classes indicated by human experts. The presented method may be used to assess the need for the coronary revascularization. The paper includes results of experiments that have been performed on medical data obtained from Second Department of Internal Medicine, Collegium Medicum, Jagiellonian University, Kraków, Poland.**

I. INTRODUCTION

CORONARY heart disease (CHD) is a major health problem worldwide and is one of the leading causes of high mortality rates in industrialized countries (see [15]). It is called angina, due to one of its main symptoms - chest pain, arising from ischemia of the heart muscle. The consequences depend largely on the number, degree and localization of artery stenosis. The current diagnostic standard of anatomic coronary vessel evaluation is invasive angiography which permits the determination of therapeutic plan and prognosis. In the case of unaltered coronary flow the pharmacological treatment is applied, otherwise revascularization is also needed. However the coronary angiography (coronarography) is a very sensitive method, it has its limitations. As an invasive investigation, it is relatively expensive, and it carries risks including a mortality rate of approximately 1 in 2000 (see [9]).

It would not be appropriate or practical to perform invasive investigations on all patients with a coronary heart disease diagnosis. Given the high incidence and prevalence of CHD, a non invasive test to reliably assess the coronary arteries would be clinically desirable.

We propose applying clinical data together with electrocardiographic (ECG) Holter recordings as prospective candidate data for coronary artery stenosis prediction. The proposed method helps determine the management of patients with stable angina, including those needing coronary intervention, without performing invasive diagnostic procedure like angiography. It could also work as a screening tool for all patients with CHD.

The presented subject employs classifier building for temporal data sets, where a *classifier* is an algorithm which enables us to forecast repeatedly on the basis of accumulated knowledge in new situations (see [2] for more details). There are many suitable methods for classification: e.g. classical and modern statistical techniques, neural networks, decision trees, decision rules and inductive logic programming (see [2] for more details). Classifiers were constructed also for temporal data (see [2], [11] for more details). In this paper, an approach for solving problems has been found in the context of rough set theory and its methods. Rough set theory introduced by Zdzisław Pawlak during the early 1980s provides the foundation also for the construction of temporal data set classifiers (see [2], [19], [4], [5]).

We present a method of classifier construction that is based on features which aggregate time points (see [2]). The patients are characterized by parameters (sensors), measured in time points for a period, called *a time window*. In the ECG recording context, the exemplary parameters are the number of QRS complexes, ST interval elevations and depressions or the total power of the HRV spectrum. The aggregation of time points is performed by special functions called *temporal patterns* (see [2], [20], [24]), that are tools for numerical characterization of values from a selected sensor during the whole time window. Temporal patterns are constructed on the basis of the domain knowledge by human experts. Next, a collection of attributes is constructed on the basis of temporal patterns. Having such attributes, a classifier is constructed approximating a temporal concept. In studied subjects, the temporal concept means the presence of coronary stenosis. The classification is performed using a decision tree that is calculated on the basis of the local discretization (see [18], [4]).

To illustrate the method and to verify the effectiveness of presented classifiers, we have performed several experiments with the data sets obtained from Second Department of Internal Medicine, Collegium Medicum, Jagiellonian University, Kraków, Poland (see Section V).

II. MEDICAL BACKGROUND

CHD refers to the narrowing (stenosis) or occlusion of the artery supplying blood to the heart, caused by plaque composed predominantly of cholesterol and fatty deposits within the vessel wall. The accumulation of plaque is a process known as atherosclerosis developing slowly over many years.

The diagnosis and assessment of angina involves a patient's medical history, physical examination, laboratory tests and specific cardiac investigations. Non-invasive investigations include a resting 12-lead ECG, ECG stress testing, resting two-dimensional and doppler echocardiography, and ECG Holter monitoring. Invasive techniques used in coronary anatomy assessment are: coronary arteriography and intravascular ultrasound. Using coronary arteriography the disease can be classified into one vessel, two vessel, three vessel, or left main coronary artery disease (LM CAD).

An electrocardiogram is a measurement of the electrical activity of the heart on the body's surface. Holter ECG monitoring is a continuous recording of the ECG, done over a period of 24 hours or more. Several electrodes are placed on the patient's chest and connected by wires to the recorder. The patient goes about his or her usual daily activities.

Modern Holter devices record data onto digital flash memory devices. When the recording of the ECG signal is completed, the data is uploaded into a computer system which then automatically analyzes the input. Commercial Holter software carries out an integrated automatic analysis process which counts i.a. ECG complexes, calculates summary statistics such as average, minimum and maximum heart rate and determines different kinds of heart beats and rhythms. It provides information about heart beat morphology, interval measurements, heart rate variability (HRV) and rhythm overview.

For many years, ECG analysis was adjusted and succeeded in methods for determining plentiful signal features in the background for the establishment of a diagnosis. The physicians had to gain experience in using these properties in diagnosing. Difficulties arise when there is a need to predict the necessity of revascularization. No distinct expert knowledge is available in this field, and the predictions based on non-invasive tests are conjectural. Due to the invasive character of revascularization which can exposure the patient to danger and is a costly procedure, it would be beneficial to know whether it's required in advance.

There are numerous systems analyzing ECG recordings. In our study the data was acquired using Aspel's HolCARD 24W application. Computer systems enable the processing and aggregation of data by means of existing signal analyzing methods. We took advantage of aggregated data to predict coronary angiography outcomes in regard to the necessity of revascularization. The data still remain temporal regardless data aggregation to points representing, e.g. an hour of ECG recording. The study answers the question whether the application of such inputs will be utilizable.

A. *Coronary arteriography*

Coronary arteriography is a diagnostic invasive procedure that requires the percutaneous insertion of a catheter into the vessels and heart. The catheter is introduced into the body through a vein or an artery, and its progress is monitored by x-ray. Injected dye (contrast medium, CM or IV dye) enables the evaluation of heart valves functioning, coronary blood flow and allows for the identification of the presence, localization and degree of stenosis in the coronary arteries.

Coronary arteriography is considered a relatively safe procedure, but in some patients complications arise. Most of them are minor, resulting in no long-term consequences and include nausea, vomiting, allergic skin rashes and mild arrhythmias. In patients with kidney dysfunction, the supply of an excessive quantity of CM may worsen kidney functioning. There may be bleeding at the catheter insertion site, which might develop into swelling. Less frequently, angiography may be associated with more serious complications. These include blood vessel damages, formation of blood clots, infections, arrhythmias, MI (myocardial infarction) or a stroke. The risk of major complications associated with coronary angiography is determined to be up to 2% (see [1]). Fatalities are extremely rare and may be caused by a perforation of the heart or surrounding vessels, arrhythmias, a heart attack or a severe allergic reaction to CM (see [14]).

Application of contrast medium during angiography additionally exposes patients to the adverse effects of supervention. Reactions to IV dye are relatively common, occurring in 1 to 12% of patients (see [8]). Most of these reactions are mild, and include a feeling of warmth, nausea and vomiting. Generally, these symptoms last for only a short period of time and do not require treatment. Moderate reactions, including severe vomiting, urticaria and swelling, occur in 1% of people receiving CM and frequently require treatment. Severe, life-threatening reactions, including anaphylaxis, are reported in 0,03 to 0,16% of patients, with an expected death rate of one to three per 100 000 contrast administrations (see [10]).

Diagnostic cardiac angiography plays an important role in the evaluation of patients with coronary heart disease. It is used to assess the presence and degree of coronary artery stenosis, heart valve and muscle dysfunction.

There are no routine noninvasive diagnostic procedures to assess coronary flow disturbances and when there is no opportunity to perform coronary angiography, alternative solutions to the problem are needed. The application of the proposed methods could select potential candidates for myocardial revascularization. We suggest using clinical data, derived from a patient's history, and laboratory test outcomes together with ECG Holter recordings as prospective candidate data for coronary artery stenosis prediction.

B. Management of angina

Once the diagnosis of CHD is made, it is important to define the treatment and the need for revascularization. Coronary arteriography in conjunction with a cardiovascular examination can appropriately select patients for coronary revascularization which means the restoration of the blood supply to ischemic myocardium. Modes of revascularization include thrombolysis with drugs, percutaneous coronary intervention (PCI) mainly by way of angioplasty, and coronary artery bypass grafting (CABG). PCI restores blood flow, usually with a balloon, inserted by a catheter through the peripheral artery, with or without stent placement. CABG refers to an "open heart" surgery where a peripheral vein is used to bypass the occlusion in the coronary artery.

Myocardial revascularization procedures require diagnosis which should indicate the localization, extent and severity of the disease, the presence and significance of the collateral circulation and the status of the left ventricular myocardium. For many years, the evaluation of the extent and severity of coronary artery disease has been mainly anatomical, carried out by a coronary angiography. However, this technique has methodological limitations and the interobserver variability is considerable. Intravascular ultrasounds (IVUS)have an indisputable advantage in determining lesion characteristics. But the only noninvasive technique that allows for quantitative assessment is a positron-emission tomography (PET), but it is highly complex and expensive, so its use is strictly limited.

III. PREDICTION OF CORONARY ATHEROSCLEROSIS PRESENCE

Forecasting coronary stenosis in patients without performing an angiography requires classifier construction, which on the basis of available knowledge assigns objects (patients) to defined decision classes. Considered decision classes are: *patients with unaltered arteries who do not need invasive treatment* (decision class: *NO*) and *patients with coronary stenosis who may need revascularization* (decision class: *YES*). Classification thus permits decision making about coronary stenosis and therapy management.

A. Temporal Concepts

The problem of forecasting coronary stenosis presence can be treated as an example of a concept approximation problem, where the term *concept* means *mental picture of a group of objects*. Such problems can often be modeled by systems of complex objects and their parts changing and interacting over time. The objects are usually linked by some dependencies, sometimes they can cooperate between themselves and are able to perform flexible complex autonomous actions (operations, changes). Such systems are identified as *complex dynamical systems* or *autonomous multiagent systems* (see [2] for more details). For example, in the problem of coronary stenosis prediction, a given patient can be treated as an investigated complex dynamical system, whilst diseases of this patient are treated as complex objects changing and interacting over time.

The concepts and methods of their approximation are usually useful tools for an efficient monitoring of a complex dynamic system (see [2]). Any concept can be understood as a way to represent some features of complex objects. An approximation of such concepts can be made using parameters (sensor values) registered for a given set of complex objects. However, a perception of composite features of complex objects requires observation of objects over a period called a *time window*. For construction of the features *temporal patterns* are used. In this paper, we consider temporal patterns as a numerical characterization of values of selected sensors from a time window (e.g., the minimal, maximal or mean value of a selected sensor, the initial and final values of a selected sensor, the deviation of selected sensor values).

One can see that any temporal pattern is determined directly by the values of some sensors. For example, in the case of coronary disease one can consider temporal patters such as minimal heart rate and estimated QT dispersion within a time window. We assume that any temporal pattern should to be defined by a human expert using domain knowledge accumulated for the given complex dynamical system.

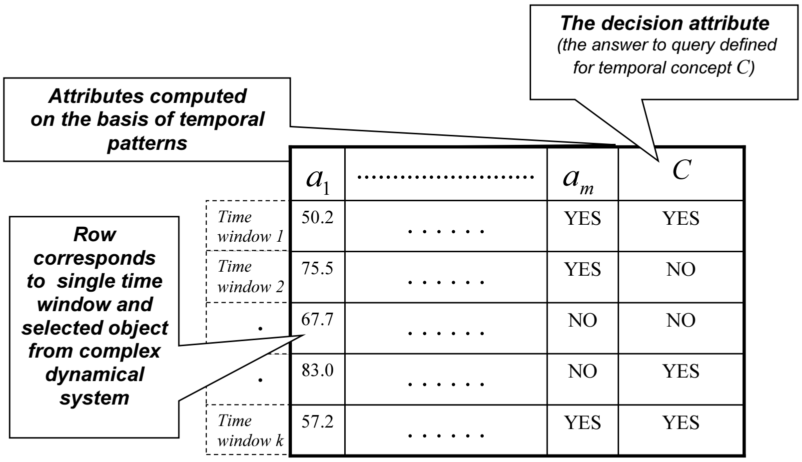
The temporal patterns can be apply for defining new features that can be used to approximate more complex concepts, that we call *temporal concepts*. We assume that temporal concepts are specified by a human expert. Temporal concepts are usually used in queries about the status of some objects in a particular temporal window. Answers to such queries can be of the form of Yes, No or *Does not concern*. For example, in the case of the main problem in this paper we define a complex concept by usaging the following query: "Was the stenosis of coronary artery detected for a given patient?".

B. Temporal pattern table

The approximation of temporal concepts can be defined by classifiers, which are usually constructed on the basis of decision tables. Hence, if we want to apply classifiers for the approximation of temporal concepts, we have to construct a suitable decision table called a *temporal pattern table* (PT) (see Figure 1).

A temporal pattern table is constructed from a table T consisting of registered information about objects (patients) occurring in a complex dynamical system. Any row of table T represents information about the parameters of a single object registered in a time window. Such a table can be treated as a data set accumulated from the observations of the behavior of a complex dynamical system. Assume, for example, that we want to approximate temporal concept C using table (data set) T. Initially, we construct a temporal pattern table PT as follows.

- Construct table PT with the same objects that are contained in table T.
- Any condition attribute of table PT is computed using temporal patterns defined by a human expert for the



The decision attribute *(the answer to query defined for temporal concept C)*

Attributes computed on the basis of temporal patterns

Row corresponds to single time window and selected object from complex dynamical system

	a_1	....	a_m	C
Time window 1	50.2	......	YES	YES
Time window 2	75.5	......	YES	NO
.	67.7	......	NO	NO
.	83.0	......	NO	YES
Time window k	57.2	......	YES	YES

Fig. 1. The scheme of the temporal pattern table (PT)

approximation of concept C.

- Values of the decision attribute (the characteristic function of concept C) are proposed by the human expert.

We assume that any temporal pattern is given by a formula defined by an expert. In a more advanced approach, the classifiers for condition attributes related to temporal patterns should be constructed.

C. Classifier construction

Next, we can construct a classifier for table PT that can approximate temporal concept C. The most popular method for classifiers construction is based on learning rules from examples (see, e.g., [19], [4], [5], [2]). Unfortunately, the decision rules constructed in this way can often be inappropriate to classify unseen cases. For instance, if we have a decision table where the number of values is high for some attributes, then there is a very low chance that a new object will be recognized by the rules generated directly from this table, because the attribute value vector of a new object will not match any of these rules. Therefore, some discretization strategies are built for decision tables with such numeric attributes to obtain a higher quality classifiers. This problem is intensively studied and we consider discretization methods developed by Hung S. Nguyen (see [18], [4] for more details). These methods are based on rough set techniques and boolean reasoning.

In this paper we use the local strategy of discretization (see [4]). One of the most important notion of this strategy is the notion of *a cut*. Formally, the cut is a pair (a, c) defined for a given *decision table* $\mathbf{A} = (U, A \cup \{d\})$ in Pawlak's sense (see [19]), where $a \in A$ (A is a set of attributes or columns in the data set) and c, defines a partition of V_a into *left-hand-side* and *right-hand-side interval* (V_a is a set of values of the attribute a). In other words, any cut (a, c) is associated with a new binary attribute (feature) $f_{(a,c)} : U \to \{0, 1\}$ such that for any $u \in U$:

$$f_{(a,c)}(u) = \begin{cases} 0 & \text{if } a(u) < c \\ 1 & \text{otherwise} \end{cases} \tag{1}$$

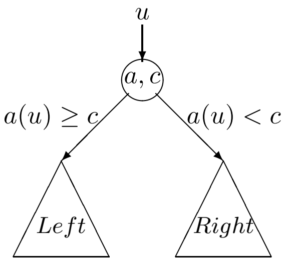


Fig. 2. The decision tree used in local discretization

Moreover, any cut (a, c) defines two templates, where a template is understood as a description of some set of objects. The first template defined by a cut (a, c) is a formula $T = (a(u) < c)$, while the second pattern defined by a cut (a, c) is a formula $\neg T = (a(u) \geq c)$.

The quality of a given cut can be computed as a number of objects pairs discerned by this cut and belonging to different decision classes. Such an approach was used in [3] and the classifier constructed through use of this method, will be called here as the *RSH-classic* classifier. However, in this paper a new method of a cut quality computation is introduced. It is based on special weights obtained for pairs of patient on the basic of domain knowledge (see Section IV). This method allowed to significantly improve the results of our experiments relatively to the results from [3] (see Section V).

The quality of cuts may be computed for any subset of a given set of objects. In the local strategy of discretization, after finding the best cut and dividing the object set into two subsets of objects (matching to both templates mentioned above for a given cut), this procedure is repeated for each object set separately until some stop condition holds. In this paper, we assume that the division stops when all objects from the current set of objects belong to the same decision class. Hence, the local strategy can be realized by using a *decision tree* (see Figure 2).

The decision tree computed during local discretization can be treated as a classifier for the concept C represented by a decision attribute from a given decision table $\mathbf{A}$. Let u be a new object and $\mathbf{A}(T)$ be a subtable containing all objects matching the template T defined by the cut from the current node of a given decision tree (at the beginning of algorithm work T is the template defined by the cut from the root). We classify object u starting from the root of the tree as follows:

```
Algorithm Classification by decision tree (see [4])
Step 1 If $u$ matches template $T$ found for $\mathbf{A}$
           then: go to subtree related to $\mathbf{A}(T)$
           else: go to subtree related to $\mathbf{A}(\neg T)$.
Step 2 If $u$ is at the leaf of the tree then go to 3
           else: repeat 1-2 substituting $\mathbf{A}(T)$
           (or $\mathbf{A}(\neg T)$) for $\mathbf{A}$.
Step 3 Classify $u$ using decision value attached to the leaf
```

Figure 3 presents a decision tree computed for the problem of forecasting the presence of coronary stenosis on the basis of a medical data set (see Section V). A sample application of the tree would be the classification of real life objects. For example, for a patient with first average of QTc interval in time window equal to 380 ms and first QT standard devation in channel no 1 of ECG Holter equal 5.8, we flow from the root of the tree, down to the right subtree, as the patient suits a template $FIRST\_QTC1\_AVG < 451$. Then, in the next step we tread again along the right tree, which consists of one node, called a leaf, where we stop. The fitting path indicates that the coronary arteries of that patient are narrowed by atherosclerosis.

Our findings have clinical relevance. Generated cuts concern the QT interval which represents the duration of ventricular depolarization and subsequent repolarization. Prolongation of the QT interval reflects a delay in cardiac repolarization and is associated with the increased risk of potentially fatal ventricular arrhythmias. Because of the QT inverse relationship to heart rate, it is routinely corrected by means of various formulae to a less heart rate dependent value known as the QTc interval. The duration of QTc interval should normally be up to 450 ms for women, and 430 ms for men (see [22]).

The proposed method can predict coronary stenosis using the multi-step approach. In the first step, the ECG Holter data are preprocessed and consolidated with clinical data. The second step generates temporal patterns chosen by an expert. Using a decision tree with the local strategy of discretization, the third step approximates the studied concept. This step employs the transition weighting approach to locally estimate the biggest number of differentiated pairs of objects belonging to opposite classes. This ensures that received cuts produce more stable results. In the fourth step the performance of classification is performed.

IV. WEIGHTS OF CUTS RELEVANT TO NUMBER OF VESSELS INVOLVED IN CHD

The concept we learn is defined as the presence of coronary stenosis indicating a need for revascularization. We based the concept membership on angiographic data which divides patients into groups having single-vessel, double-vessel or triple-vessel disease depending on the number of coronary arteries involved. The anatomic stratification of CHD to one, two and three vessels provides useful prognostic information and is used in the selection of patients for revascularization. Triple-vessel disease carries worse prospects than a double-vessel, which is usually worse than single-vessel disease. Generally, patients with single or double-vessel disease can benefit from PCI. For patients with triple-vessel disease, or the presence of poor heart function, a CABG can often be a good alternative or a better treatment option.

We acknowledge patients with unaltered coronary flow as not belonging to the concept, while patients with one, two or three vessel disease as concept positive examples. However, this criterion is too simple and not limitless. Dissociation often exists between the coronarography and clinical manifestations. The group with one, two and three vessel disease is inhomogeneous. It distinguishes patients based only on the number of vessels, and treats a multidimensional problem as a one-dimensional. Some physiological facts are fundamental for correct interpretation of the study results. On the one hand in patients with angiographically confirmed stenosis, the collateral circulation and coronary self-regulation may compensate for the blood pressure decrease caused by stenosis in order to maintain constant coronary flow. In coronarography, the role of collateral circulation can be underestimated.

Some previous studies revealed that the ECG's may appear inconsistent with stenosis disclosed by arteriography. Sometimes the angiographic picture appears much more alarming than what is anticipated from the ECG. In patients with complete occlusion the resting ECG may be completely normal ([12]). It was especially evident in the case of right coronary artery when severe luminal reduction, or even complete occlusions, was often accompanied by normal or only slightly abnormal ECG. In certain rare cases it may be difficult to demonstrate and evaluate the presence of coronary stenosis, despite a technically satisfactory arteriogam (e.g. [7]). On the other hand, some cases of myocardial infarction with normal coronary arteries are reported (e.g. [6]). Many theories have been proposed explaining these inconsistencies, such as syndrome X, microvascular angina, and non-atherosclerotic myocardial ischemia. The presence of muscular bridges may also bring some uncertainty although in these cases during the diastolic phases, the artery appears normal.

Several studies reported that ECG changes were not good indicators of coronary artery involvement with 51.5% sensitivity in correctly detecting significant stenosis ([16]). All things considered, the ECG data is still attractive as a prospective candidate for predicting CAD because it is noninvasive, easy to use, a relatively inexpensive tool offering safety and patient convenience.

Given the limited ability of ECG to distinguish patients correctly, we made an attempt to emphasize the differences between groups using weights for cuts discerning pairs of objects. It is a way to make the border between positive and negative examples more exposed. We are interested in distinguishing patients with normal and narrowed (one, two or

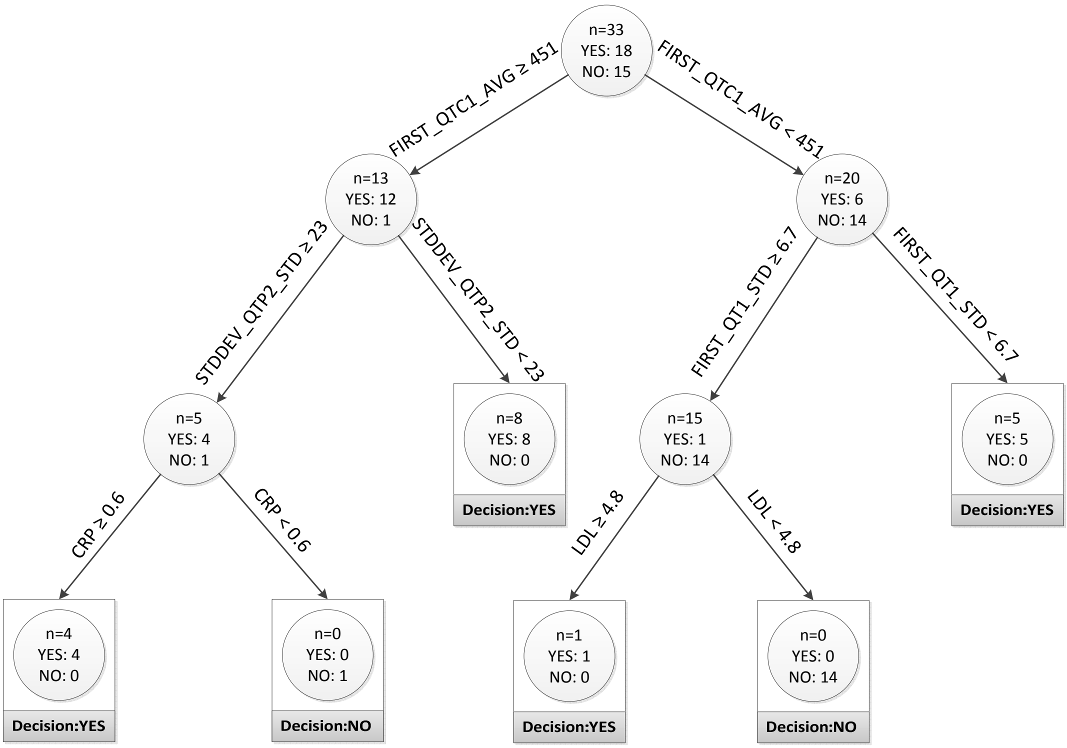


Fig. 3. The decision tree in CHD

TABLE I
DIVERSIFYING WEIGHTS RELEVANT TO NUMBER OF VESSELS INVOLVED IN CHD

	0	**1**	**2**	**3**
0	0	3	2	1
1	3	0	1	1
2	2	1	0	1
3	1	1	1	0

three) vessels, that‘s why we put the biggest weights between negative examples of the concept and the remainder. Class differences are the most subtle between patients with normal arteries and with only one vessel changed, compared with the divergence between normal and tree-vessel disease. The cut discriminating patients with normal and one vessel disease is assigned the biggest weight. Thereupon we propose weights as presented in table I, where the first row and the first column signify the number of affected arteries.

A classifier constructed by using the method based on the weights from Table I will be refered to from here as the *RSH-weights* classifier. It is easy to see, that the method of cuts weights computation used in the *RSH-classic* classifier (see Section III-C) can be treated as a special case of the *RSH-weights* method, that is, a case when there are only two values in the Table I (values: 0 or 1).

In the *RSH-classic* approach, the quality of a cut is a sum of the pairs of objects belonging to opposite decision classes, whereas in the *RSH-weights* method it is a summation of the weights of each pair of objects belonging to different decision classes, with 1, 2 or 3 vessel disease distinction.

V. EXPERIMENTS AND RESULTS

To verify the effectiveness of classifiers based on temporal patterns, we have implemented the algorithms from the library RSH-lib, which is an extension of the RSES-lib library forming the kernel of the RSES system [5].

The experiments have been performed using the medical data set obtained from Second Department of Internal Medicine, Collegium Medicum, Jagiellonian University, Kraków, Poland. The data was collected between 2006 and 2009. Two part 48-hour Holter ECG recordings were performed using Aspel's HolCARD 24W system. There was a coronary angiography after the first part of the Holter ECG (after first 24-hour recording). In the paper, we reported the results of the experiments performed for the first part of Holter ECG recordings. The data set includes a detailed description of clinical status (age, sex, diagnosis), coexistent diseases, pharmacological management, the laboratory tests outcomes (level of cholesterol, troponin I, LDL - low density lipoproteins) and various Holter-based indices such as: ST interval deviations, HRV, arrythmias or QT dispersion. Moreover, for Holter-based indices a data aggregation was performed resulting in points describing one hour of recording. Our group of 33 patients with normal rhythm, underwent coronary angiography and 24.2% of them required additional angioplasty, whereas 24.2% were qualified for CAGB.

The acquired data of patients was recorded as binary files. In

TABLE II
THE CONFUSION MATRIX

		Predicted	
		No	Yes
Actual	No	12	3
	Yes	1	17

TABLE III
RESULTS OF EXPERIMENTS FOR CORONARY STENOSIS IN CHD

Decision class	Accuracy	Precision
Yes	0.944	0.85
No	0.8	0.923
All classes (Yes + No)	0.879	-

the first step we verified the completeness of patient data (dates of tests, results, etc.) and a database of patients and medical data was created in the Infobright Community Edition (ICE) environment (see [13]). ICE is an open source software solution designed to deliver a scalable data warehouse, optimized for analytic queries (data volumes up to 50 TB, market-leading data compression, from 10:1 to over 40:1).

Data files for each test were imported into the database using an importer created in Java environment. In this process the text data was converted to the corresponding data formats (reviewing the various parameters we defined formats allowing its efficient storage in the database) to allow for storage of patient data without loss of information (such as float, integer, long). After internal preprocessing in the ICE environment (e.g., a data aggregation of Holter-based indices as was mentioned above) for further processing data have been imported into Java environment.

The aim of the conducted experiments was to check the effectiveness of the algorithm described in this paper in order to predict the stenosis in coronary arteries. Here we present the experimental results of presented method. For testing quality of classifiers we applied the leave-one-out (LOO) technique, that is usually employed when the size of a given data set is small. The LOO technique involves a single object from the original data set as the validation data, and the remaining observations as the training data. This is repeated such that each observation in the sample is used once as the validation data. The LOO provides a repeatable objects selection and uses the knowledge of all remaining objects in a learning process, and thus is comparable with other classification methods. Knowing however it is not free from disadvantages, we plan to use a bootstrap method.

As a measure of classification success (or failure) we used the following parameters which are well known from the literature: the accuracy, the coverage, the accuracy for positive examples (Sensitivity, SN), the coverage for positive examples, the precision for positive examples (Positive Predictive Value, PPV), the accuracy for negative examples (Specificity, SP), the coverage for negative examples and the precision for negative examples, also called Negative Predictive Value, NPV (see, e.g., [2]).

The numbers of objects correctly and incorrectly classified is contained in table II. Table III shows the results of applying the considered algorithm (*RSH-weights* classifier) for the concept related to the presence of coronary artery stenosis in patients with stable angina.

The method correctly identifies 94.4% of all patients with stenosis (SN) and 80% of those who did not have stenosis (SP). With PPV value equal 85%, a positive screen test is good at confirming coronary stenosis, however a negative result is also good as a screening tool at affirming that a patient does not have stenosis (NPV equal 92.3%).

In Table IV we gave the results of the experiments in applying other classification methods to our data. Those methods were developed in the following systems well known from the literature: WEKA [23], RSES [5], ROSE2 [21] (we used an early implementation of ModLEM algorithm [17] that is available in ROSE2), and our previous approach called *RSH-classic* classifier (see [3]). The coverage of all tested methods was equal 1.0 (every object was classified).

It should be mentioned that the results for WEKA and ROSE2 were generated using a set of standard parameters.

Experimental results showed that the presented method of stenosis prediction in coronary arteries gives good results and the results are comparable with results of other systems. Interestingly, the tests classified 94.4% of patients with narrowed vessels as stenosis presence, which is an expected property of the method, meaning that the number of false negatives should be the lowest.

A. Limitations of the study

The main limitation of the study was the size of the study population. However, our results can be applicable for a similar patient population. Another limitation of our study was using only the number of affected vessels as a determinant of CHD severity, as used in most of the studies in the literature. It is known that resting ECG has a limited value in determining coronary artery lesion characteristics.

VI. CONCLUSION

In the present work, clinical and ECG data were used to build a predictive model for the diagnosis of CHD. We believe that the method can be very useful to clinicians in managing patients with CHD. Patients with positive tests may be strongly considered for revascularization, even if other tests results indicate a moderate or weak risk range. For negative tests, the clinician may observe the patient continuing pharmacotherapy.

The most attractive aspect of the method is that it can be employed with easy to obtain clinical, laboratory, and electrocardiographic parameters. Estimating the coronary anatomy before angiography could be useful when deciding on diagnostic and therapeutic interventions.

The number of studies dealing with the relationship between electrocardiogram and the severity of CHD are limited. To the

TABLE IV
COMPARISON RESULTS OF ALTERNATIVE CLASSIFICATION SYSTEMS

Method	Accuracy			Precision	
	All classes	Yes	No	Yes	No
C4.5 (WEKA)	0.545	0.611	0.467	0.579	0.500
NaiveBayes (WEKA)	0.394	0.611	0.133	0.458	0.222
SVM (WEKA)	0.545	0.611	0.467	0.579	0.500
k-NN (WEKA)	0.667	0.833	0.467	0.652	0.700
RandomForest (WEKA)	0.515	0.722	0.267	0.542	0.444
Multilayer Perceptron (WEKA)	0.548	0.611	0.467	0.579	0.500
Global discretization + all rules (RSES)	0.667	0.611	0.733	0.733	0.611
Local discretization + all rules (RSES)	0.758	0.778	0.733	0.778	0.733
ModLEM (ROSE2)	0.576	0.556	0.600	0.625	0.529
RSH-classic	0.758	0.778	0.733	0.778	0.733
RSH-weights	0.879	0.944	0.8	0.85	0.923

best of our knowledge, no study investigating the relationship between the presence of coronary artery stenosis and ECG Holter monitoring has been demonstrated until now. The proposed work is very important for medical practitioners who treat patients with CHD in every day practice. The prediction of coronary arteries stenosis may help in better and tailored CHD management and treatment.

Further investigation is needed to assess whether proposed method leads to a meaningful change in clinical outcomes and may be used as a more routine, screening test for stenosis prediction. To that end we submitted an application to The National Science Centre in Poland for the research grant, which could enable the investigations on more numerous study population. Moreover, it would be of great benefit for patients, to develop a method predicting the precise number of narrowed arteries.

ACKNOWLEDGMENT

This work was supported by the grant N N516 077837 from the Ministry of Science and Higher Education of the Republic of Poland, the Polish National Science Centre (NCN) grant 2011/01/B/ST6/03867 and by the Polish National Centre for Research and Development (NCBiR) grant No. SP/I/1/77065/10 in frame of the strategic scientific research and experimental development program: "Interdisciplinary System for Interactive Scientific and Scientific-Technical Information".

REFERENCES

[1] ACC/AHA Guidelines for Coronary Angiography: Executive Summary and Recommendations. A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines, *Circulation,* 99, 1999, pp. 2345–2357.

[2] J. G. Bazan, "Hierarchical classifiers for complex spatio-temporal concepts", *Transactions on Rough Sets,* IX, LNCS 5390, 2008, pp. 474–750.

[3] J. G. Bazan, S. Bazan-Socha, S. Buregwa-Czuma, P. W. Pardel, B. Sokolowska, "Prediction of coronary arteriosclerosis in stable coronary heart disease", Proceedings of the Fourteen Conference of Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'2012), Catania, Italy, July 9-13, 2012, Springer-Verlag, Lecture Notes in Artificial Intelligence (accepted).

[4] J. G. Bazan, H. S. Nguyen, S. H. Nguyen, P. Synak, J. Wróblewski, "Rough set algorithms in classification problems", in: L. Polkowski, T. Y. Lin, S. Tsumoto (Eds.), "Rough Set Methods and Applications: New Developments in Knowledge Discovery in Information Systems," *Springer-Verlag/Physica-Verlag, Heidelberg, Germany, Studies in Fuzziness and Soft Computing,* vol. 56, 2000, pp. 49–88.

[5] J. G. Bazan, M. Szczuka, "The Rough Set Exploration System", *Transactions on Rough Sets,* III, LNCS 3400, 2005, pp. 37–56.

[6] R. M. Bossarte, M. J. Brown, R. L. Jones, "Myocardial Infarction with Normal Coronary Arteries: A Role for MRI?", *Clinical Chemistry,* vol. 5, 2007, pp. 995–996.

[7] R. Bugiardini, L. Badimon, P. Collins, R. Erbel, K. Fox, C. Hamm, F. Pinto, A. Rosengren, C. Stefanadis, L. Wallentin, F. Van de Werf, "Angina, "Normal" Coronary Angiography, and Vascular Dysfunction: Risk Assessment Strategies", *PLoS Medicine,* vol. 4, 2007, pp. 252–255.

[8] L.M. Canter, "Anaphylactoid reactions to radiocontrast media", *Allergy and Asthma Proceedings,* 26, 2005, pp. 199–203.

[9] S. Chaubey, S.W. Davies, N. Moat, "Invasive investigations and revascularisation", *British Medical Bulletin,* vol. 59, 2001, pp. 45–53.

[10] S.T. Cochran, "Anaphylactoid reactions to radiocontrast media", *Current Allergy and Asthma Reports,* 5, 2005, pp. 28–31.

[11] A. Douzal-Chouakria, C. Amblard, "Classification trees for time seriess", *Pattern Recognition,* vol. 45, issue 3, 2011, pp. 1076–1091.

[12] G. G. Gensini, C. Buonanno, "Coronary Arteriography : A Study of 100 Cases with Angiographically Proved Coronary Artery Disease", *Dis. Chest.,* vol. 54, 1968, pp. 90–99.

[13] The Infobright Community Edition (ICE), http://www.infobright.org/

[14] R.A. Lange, L.D. Hilliss, "Diagnostic Cardiac Catheterization", *Circulation,* 107, 2003, pp. e111–e113.

[15] J. Mackay, G.A. Mensah, "The Atlas of Heart Disease and Stroke", World Health Organization, 2004.

[16] S. Mahmoodzadeh, M. Moazenzadeh, H. Rashidinejad, M. Sheikhvatan, "Diagnostic performance of electrocardiography in the assessment of significant coronary artery disease and its anatomical size in comparison with coronary angiography", *J. Res. Med. Sci.,* vol. 16(6), 2011, pp. 750–755.

[17] K. Napierala, J. Stefanowski, "Argument Based Generalization of MODLEM Rule Induction Algorithm", *Proceedings of the RSCTC 2010, Springer-Verlag, Lecture Notes in Artificial Intelligence 6086,* 2010, pp. 138–147.

[18] H. S. Nguyen, "Approximate Boolean Reasoning: Foundations and Applications in Data Mining", *Transactions on Rough Sets,* **V**, LNCS 4100, 2006, pp. 334–506.

[19] Z. Pawlak, A. Skowron, "Rudiments of rough sets", *Information Sciences,* 177, 2007, pp. 3–27.

[20] J. F. Peters, "Classification of objects by means of features", in: D. Fogel, J. Mendel, X. Yao, T. Omori (Eds.), "Proc. 1st IEEE Symp. Foundations of Computational Intelligence," (FOCI'2007), Honolulu, Hawaii, 2007, pp. 1–8.

[21] The Rough Sets Data Explorer (ROSE2), http://idss.cs.put.poznan.pl/site/rose.html

[22] S. M. J. M. Straus, J. A. Kors, M. L. De Bruin, C. S. van der Hooft, A. Hofman, J. Heeringa, J. W. Deckers, J. H. Kingma, M. C. J. M. Sturkenboom, B. H. Ch. Stricker, J. C. M. Witteman, "Prolonged QTc Interval and Risk of Sudden Cardiac Death in a Population of Older Adults", *Journal of the American College of Cardiology,* vol. 47, no. 2, 2006, pp. 362–367.

[23] The Weka 3 - Data Mining Software in Java (WEKA), http://www.cs.waikato.ac.nz/ml/weka/

[24] M. Wolski, "Perception and classification. A note on near sets and rough sets", *Fundamenta Informaticae,* vol. 101, no. 1, 2010, pp. 143–155.