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

Jan G. Bazan
Sylwia Buregwa-Czuma
Przemysław 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 Sokolowska
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 cardiovascular pathology. Based on clinical and electrocardiographic (ECG) Holter data we propose the method for predicting coronary stenosis demanding revascularization in patients with diagnosis of stable coronary heart disease. An approach to solving this problem has been found in the context of rough set theory and methods. Rough set theory introduced by Zdzisław Pawlak during the early 1980s provides the foundation for the construction of classifiers. From the rough set perspective, classifiers presented in the paper are based on a decision tree calculated on the basis of the local discretization method. We present a new modification of tree building method which emphasizes the discernibility of objects belonging to decision classes indicated by human experts. Presented method may be used to assess the need for revascularization and in special circumstances, to confirm or reject the diagnosis of coronary artery disease. The paper includes results of experiments that have been performed on medical data obtained from Second Department of Internal Medicine, Collegium Medicum, Jagiellonian University, Krakow, Poland.**

I. Introduction

CORONARY heart disease (CHD) is a major health problem worldwide and is one of the leading cause of high mortality in industrialized countries (see [12]). It is called angina, owing to one of its main manifestation - chest pain, due to ischemia of the heart muscle. The consequences depend on the number, degree and localization of artery stenosis. The current diagnostic standard of coronary vessels anatomy evaluation is invasive angiography which permits determination of therapeutic plan and prognosis. In the case of unaltered coronary flow the pharmacological treatment is applied, otherwise there is also a need for revascularization. However the coronary angiography (coronarography) is a very sensitive method, it is not free of limitations. As invasive investigation it is relatively expensive, and carries risks including a mortality of approximately 1 in 2000 (see [7]).

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

We propose application of clinical data together with ECG Holter recordings as prospective candidate data for coronary artery stenosis prediction. The proposed method helps to determine the management of patients with stable angina, including the need for coronary intervention, without performing invasive diagnostic procedure that angiography is. It could also work as a screening tool for all patients with CHD.

The presented subject employs classifiers building for temporal data sets, where a *classifier* is an algorithm which enables us forecasting repeatedly on the basis of accumulated knowledge in new situations (see [1] for more details). There are many methods suitable for classification: e.g. classical and modern statistical techniques, neural networks, decision trees, decision rules and inductive logic programming (see [1] for more details). Classifiers were constructed also for temporal data (see [1], [8] for more details). In the paper, an approach to solving problem has been found in the context of rough set theory and methods. Rough set theory introduced by Zdzisław Pawlak during the early 1980s provides the foundation for the construction of classifiers also for temporal data sets (see [1], [16], [2], [4]).

We present a method of classifier construction that is based on features aggregating time points (see [1]). The patients are characterized by parameters (sensors), measured in time points for some period, called *a time window*. In the ECG recordings context, the exemplary parameters are number of QRS complexes, ST interval elevations and lowerings or total power of the HRV spectrum. The aggregation of time points is performed by special functions called *temporal patterns* (see [1]), that are numerical characterization of values of selected sensor from the whole time window. We assume that computing temporal pattern's value uses a formula defined by an expert. Having computed patterns, a classifier is constructed approximating a temporal concept. In studied subject, 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 [15], [2]).

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, Krakow, 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 on the arterial walls is a process known as atherosclerosis developing slowly over many years.

Diagnosis and assessment of angina involves patient's 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, ECG Holter monitoring. Invasive techniques used in coronary anatomy assessment are: coronary arteriography and intravascular ultrasound. In order to choose appropriate treatment, the patient's risk stratification is needed. Using coronary arteriography one can classify the disease 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 surface. It is depicted as graphic representation. Changes in normal ECG pattern can indicate several heart related conditions. 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 usual daily activities. ECG monitoring is performed to detect arrhythmias that occur irregularly or ischemia indicating CHD, to evaluate symptoms (e.g. chest pain, dizziness or syncopes) of possible heart disease or to monitor the effectiveness of therapy (medication or a pacemaker or automatic defibrillator).

Modern devices record data onto digital flash memory devices. When the recording of ECG signal is completed, the data are uploaded into a computer system which then automatically analyzes the input. 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. Advanced systems also perform QT segment and spectral analysis, ischemic burden evaluation, activity graph of patient or PQ segment analysis. Another possibility is the pacemaker detection and analysis.

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

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

1) 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 fluoroscopy. The function of the heart can be asses in several ways. Injected dye (contrast medium, CM or IV dye) enables to evaluate the heart valves function, the coronary blood flow and allows to identify the presence, localization and degree of stenosis in the coronary arteries. The pressures measured inside the heart is used to evaluate the pumping action of the heart.

Coronary arteriography is considered a relatively safe procedure, but in same patients complications arise. Most of them are minor, of no long-term consequence 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 function. There may be bleeding at the catheter insertion site, developing swelling. Less frequently, angiography may be associated with more serious complications. These include blood vessels damages, formation of blood clots, infections, arrhythmias, MI or a stroke. The risk of major complications associated with coronary angiography is determined to be up to 2% (see [**?**]). Fatalities are extremely rare and may be caused by perforation of the heart or surrounding vessels, arrhythmias, a heart attack or a severe allergic reaction to CM (see [**?**]).

Application of contrast medium during angiography additionally exposes patients to adverse effects supervention. Reactions to IV dye are relatively common, occurring in 1 to 12% of patients (see [**?**]). Most of these reactions are mild, and include a feeling of warmth, nausea and vomiting. Generally, these symptoms last only for 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 [**?**]).

Although reactions to CM have the same manifestations as anaphylactic reactions, these reactions are not allergic in nature, meaning that there are no allergic antibodies (immunoglobulin E - IgE) present in blood that cause allergic reactions. The pathogenesis probably involves direct cellular effects, enzyme induction and activation of the complement. CM rather directly releases histamine and other chemicals from mast cells. In addition, previous sensitization is not required and reactions to CM do not consistently recur in a given patient. There are no diagnostic tests to predict an adverse reaction to contrast medium.

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. The procedure can confirm or exclude suspicions derived from a patient's history, physical examination and evaluation by noninvasive methods such as ECG, chest x-ray, echocardiogram and stress testing. It can be also used to elucidate inadequacies or obscure conditions in patients whose clinical findings and noninvasive testing are indeterminate. Finally, catheterization can confirm suspected abnormalities in the patient for whom cardiac intervention or surgery is planned.

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. Application of proposed methods could select potential candidates for myocardial revascularization. We suggest clinical data, derived from patient's history, laboratory test outcomes together with ECG Holter recordings as prospective candidate data for coronary artery stenosis prediction.

A. Management of angina

Once the diagnosis of CHD is made, it is important to define treatment and the need for revascularization. Coronary arteriography in conjunction with a cardiovascular examination can appropriate select patients for coronary revascularization which means the restoration of 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 peripheral vein is used to bypass 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 coronary angiography. However, this technique has methodological limitations and interobserver variability is considerable. An indisputable advantage in determining lesion characteristics has intravascular ultrasound (IVUS). But the only noninvasive technique that allows quantitative assessment is 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 angiography requires construction of classifier, 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 atherosclerosis who may need angioplasty* (decision class: *YES*). Classification thus permits decision making about coronary stenosis and therapy management.

A. Complex 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 often can be modeled by systems of complex objects and their parts changing and interacting over time. The objects are usually linked by some dependencies, sometimes can cooperate between themselves and are able to perform flexible autonomous complex actions (operations, changes). Such systems are identified as *complex dynamical systems* or *autonomous multiagent systems* (see [1] 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.

Concepts and methods of their approximation are usually useful tools for an efficient monitoring of complex dynamic system (see [1]). Any concept can be understand as a way to represent some feature 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*. Such features are often represented by *temporal patterns*. 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, initial and final values of selected sensor, deviation of selected sensor values).

One can see that any temporal pattern is determined directly by values of some sensors. For example, in case of the 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 ought to be defined by a human expert using domain knowledge accumulated for the given complex dynamical system.

The temporal patterns can be treated as new features that can be used to approximate more complex concepts. We call them *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 Yes, No or *Does not concern*. For example, in case of the main problem from this paper we define complex concept with usage of 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 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 parameters of a single object registered in a time window. Such a table can be treated as a data set accumulated from observations of the behavior of a complex dynamical system. Assume, for example, that we want to approximate a 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 as contained in table T.
- Any condition attribute of table PT is computed using temporal patterns defined by a human expert for the approximation of concept C,
- Values of the decision attribute (the characteristic function of concept C) are proposed by the human expert.

We assume that for any temporal pattern a formula for computing its value is given by an expert. In more advanced approach the classifiers for the temporal patterns represented by condition attributes 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., [16], [2], [4], [1]). Unfortunately, the decision rules constructed in this way can often be not appropriate 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 is recognized by rules generated directly from this table, because the attribute value vector of a new object will not match any of these rules. Therefore for decision tables with such numeric attributes some discretization strategies are built to obtain a higher quality classifiers. This problem is intensively studied and we consider discretization methods developed by Hung S. Nguyen (see [15], [2] for more details). These methods are based on rough set techniques and boolean reasoning.

In this paper we use local strategy of discretization (see [2]). One of the most important notion of this strategy is the

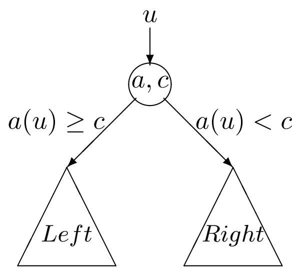

Fig. 2. The decision tree used in local discretization

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 [16]), 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 $u \in U$:

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

Moreover, any cut (a, c) defines two templates, where a template we understand 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 approach was described in [3] and the classifier constructed with usage of this method we call here as the *RSH-classic* classifier. However, in this paper we introduce a new method of a cut quality computation. It is based on special weights obtained for pairs of patient on the basic of domain knowledge (see Section IV). This method allowed significantly improve results of our experiments relatively to results from [3] (see Section V).

The quality of cuts may be computed for any subset of a given set of objects.

In 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 *decision tree* (see Figure 2).

The decision tree computed during local discretization can be treated as a classifier for the concept C represented by 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 to 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 [2])

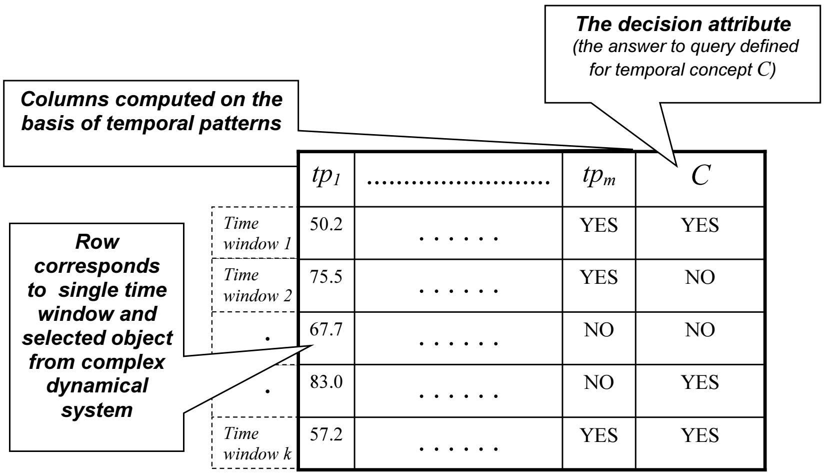

Fig. 1. The scheme of the temporal pattern table (PT)

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 coronary atherosclerosis presence on the basis of medical data set (see Section V). Sample application of the tree is classification of real life objects. For example, for a patient with first in time window average of QTc interval equal to 380 ms and first QT standard devation in channel no 1 of ECG Holter equal 5.8, we follow 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 a right tree, which consists of one node, called leaf, where we stop. The fitting path indicates that the coronary arteries of that patient are narrowed by atherosclerosis.

Generated cuts concern 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 of cardiac potentially fatal ventricular arrhythmias. Because of 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 proposed work is very important for medical practitioners who face patients with CHD in every day practice. Prediction of coronary arteries stenosis may help in better and tailored CHD management and treatment.

The proposed method can predict coronary stenosis using multi-step approach. In the first step, the ECG Holter data are aggregated using commercial Holter processing system and then together with clinical data are loaded to database. The second step generates temporal patterns chosen by expert. Using a decision tree with local strategy of discretization, the third step approximates the studied concept. The step employs 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.

Especially of interest is that tests classified as stenosis presence 93% of patients with altered vessels, what is expected property of the method.

IV. Weights of cuts relevant to number of vessels involved in CHD

The concept we learn is defined as coronary stenosis presence indicating revascularization need. We base the concept membership on angiographic data which divides patients among having single-vessel, double-vessel or triple-vessel disease depending on 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 double one, which is usually worse than single-vessel disease. Generally, patients with single or double-vessel disease can benefit from PCI. With triple-vessel disease, or the presence of poor heart function, 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 free of limitations. The dissociation often exists between coronarography and clinical manifestations. The group with one, two and three vessel disease is inhomogeneous. Distinguishing patients based only on the number of vessels 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

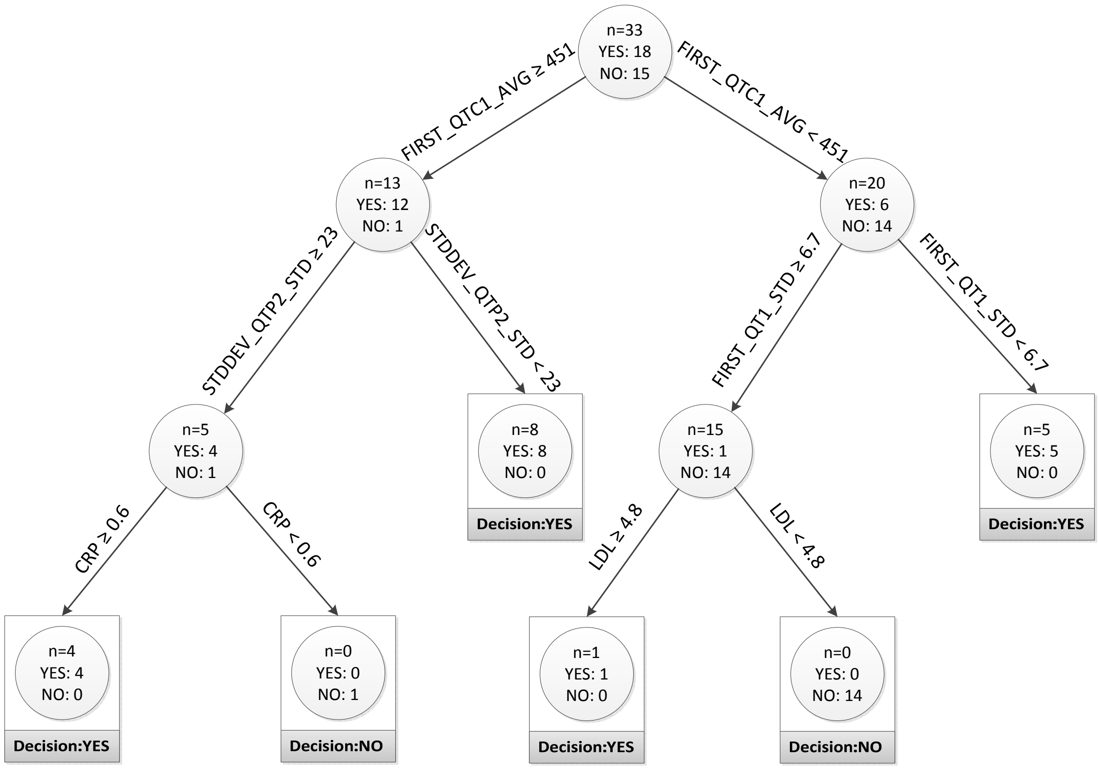

Fig. 3. The decision tree in CHD

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 ([9]). It was especially evident in 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. [6]). On the other hand, some cases of myocardial infarction with normal coronary arteries are reported (e.g. [5]. Many theories have been proposed explaining these inconsistency, such as syndrome X, microvascular angina, and non-atherosclerotic myocardial ischemia. The presence of muscular bridges may 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 arteries involvement with 51.5% sensitivity in correctly detecting significant stenosis ([13]). All things considered, the ECG data is still attractive as a prospective candidate for predicting CAD because it is noninvasive, easy to use, 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 number of discerning pairs of objects. It‘s a way to make border between positive and negative examples more exposed. We are interested in distinguishing patients with normal and narrowed (one, two or three) vessels, that‘s why we put the biggest weights between negative examples of the concept and the remainder. Class differences are most subtle between patients with normal arteries and with only one vessel changed, compared with divergence between normal and tree-vessel disease. The cut discriminating patients with normal and one vessel disease obtains the biggest weight. Thereupon we propose weights as presented in table I, where the first row and the first column signify number of affected arteries.

TABLE I
TRANSITION/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

A classifier constructed with usage of the method based on weights from Table I we call 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 a special case of *RSH-weights* method, that is, a case when there are only two values in the Table I (values: 0 or 1).

V. Experiments and results

To verify the effectiveness of classifiers based on behavioral 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 [4].

The experiments have been performed on the medical data set obtained from Second Department of Internal Medicine, Collegium Medicum, Jagiellonian University, Krakow, Poland. The data were collected between 2006 and 2009. Two part 48-hour Holter ECG recordings were performed using Aspel's HolCARD 24W system. There was coronary angiography after first part of Holter ECG (after first 24-hour recording). In the paper, we report results of experiments performed for the first part of Holter ECG recordings. The data set includes a detail 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.

All data were imported into Infobright Community Edition (ICE) environment (see [11]). ICE is an open source software solution designed to deliver a scalable data warehouse, optimized for analytic queries (data volumes up to 50TB, market-leading data compression, from 10:1 to over 40:1). 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 conducted experiments was to check the effectiveness of the algorithm described in this paper in order to predict atherosclerosis in coronary arteries. Here we present the experimental results of presented method. For testing quality of classifiers we applied 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. As a measure of classification success (or failure) we use the following parameters well known from literature: the accuracy, the coverage, the accuracy for positive examples (Sensitivity, SN or recall), 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.*, [1]).

TABLE II
Results of experiments for coronary stenosis in CHD

Decision class	Accuracy	Coverage	Precision
Yes	0.944	1.0	0.85
No	0.8	1.0	0.923
All classes (Yes + No)	0.879	1.0	-

Table II shows the results of applying the considered algorithm (*RSH-weights* classifier) for the concept related to presence of coronary atherosclerosis in patients with stable angina.

The method correctly identifies 94.4% of all patients with stenosis (SN), that's why a negative result would suggest the absence of disease. 80% of those who did not have stenosis (SP) were correctly identified, so a positive result means a high probability of the presence of disease. 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 = 92.3%).

In Table III we give the results of experiments in applying other classification methods to our data. Those methods were developed in the following systems well known from literature: WEKA [18], RSES [4], ROSE2 [17] (we used an early implementation of ModLEM algorithm [14] 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).

Experimental results showed that the presented method of atherosclerosis prediction in coronary arteries gives good results and the results are comparable with results of another systems.

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 the number of affected vessels for the severity of CHD, as used in most of the studies in the literature. It is well known that coronary angiography has a limited value in determining lesion characteristics.

VI. Conclusion

In the present work, clinical and ECG data were used to build predictive model for the diagnosis of CHD. Further investigation is needed to assess whether proposed method leads to a meaningful change in clinical outcome. Whether to use this method as a more routine, screening test for stenosis prediction, requires clinical verification.

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 moderate or weak risk range. For negative tests, the clinician may observe patient continuing pharmacotherapy. Further diagnostic techniques, such as stress

TABLE III
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

echocardiography or nuclear stress testing may help to delineate which patients have higher degrees of CHD.

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 CAD are limited. To the 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.

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 the strategic scientific research and experimental development program: “Interdisciplinary System for Interactive Scientific and Scientific-Technical Information”.

REFERENCES

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

[2] 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.

[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, M. Szczuka, “The Rough Set Exploration System”, *Transactions on Rough Sets,* III, LNCS 3400, 2005, pp. 37–56.

[5] 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.

[6] 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.

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

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

[9] 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.

[10] “Guidelines. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology”, *European Heart Journal,* 17, 1996, pp. 354–381.

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

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

[13] 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.

[14] 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.

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

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

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

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