# Determining Risk Factors for Survival after LMCA Stenosis with Intelligent Data Analysis

P Povalej[1], V Kanic[3], P Kokol[1,2]

[1]Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia
[2]Faculty of Health Sciences, University of Maribor, Maribor, Slovenia
[3]Teaching Hospital Maribor, Dep of Internal Medicine, Div of Cardiology and Angiology, Maribor, Slovenia

## Abstract

*Coronary artery disease is one of the most frequent causes of premature deaths in Slovenia and also in most countries in the world. A "gold standard" for treatment of left main coronary artery (LMCA) stenosis is still a surgical therapy; however Percutanueous Transluminal Coronary Angioplasty (PTCA) is much simpler for the patients and gives comparable short-term and mid-term results to surgical therapy. PTCA of LMCA stenosis is safe and technically demanding but long-term clinical outcomes are not yet defined.*

*In this paper we present an intelligent data analysis method for inducing a decision tree that was able to outline some anticipated and also some relatively unexpected but useful risk factors for survival after PTCA.*

## 1. Introduction

Coronary artery disease is the narrowing or obstruction of the vessels that supply blood and oxygen to the heart muscle that is caused by fatty deposits on the walls of arteries. The consequence of gradual loading of fatty deposits is reduced flow of blood and oxygen to the heart. When the blood flow is significantly reduced, some form of medical treatment becomes necessary.

One of the most common non-surgical treatment for opening obstructed coronary arteries is Percutanueous Transluminal Coronary Angioplasty (PTCA). The procedure is technically demanding and relatively safe for the patients.

In this research we focus on determining the most important risk factors for survival after PTCA of left main coronary artery (LMCA) stenosis, which is a relatively infrequent but important cause of symptomatic coronary artery disease. Multiple studies have found LMCA stenosis to be an independent indicator of increased morbidity and mortality rates among patients with coronary artery disease [1]. The main reason for the research is the fact that stenting of unprotected left main coronary artery (LMCA) stenosis is often performed but a long-term safety of this therapy is not yet established. Limited information is available regarding risk factors that might lead to cardiac events (death, myocardial infarction, unstable angina, left ventricular failure, and life-threatening ventricular arrhythmias) after surgery [2].

We performed this study in patients with unprotected LMCA stenosis/occlusion that were followed for approximately one year after the procedure. An intelligent data analysis method was used for identifying the risk factors that might predict long term survival.

Nowadays, many real-world medical problems are being handled with tools for automatic intelligent data analysis. Various methods such as neural networks, decision trees, genetic algorithms, hybrid approaches, etc. have been developed and evaluated on different medical databases [3-6]. However, from physician's point of view the ability to track and evaluate every step in the decision making process is the most important factor for trusting the decisions gained with machine learning methods. Therefore from the medical point of view the symbolic representation of the extracted knowledge is crucial. That fact exactly puts the use of decision tree knowledge representation on the first place in medical intelligent data analysis.

Decision trees namely provide a very powerful feature – the possibility of explaining the decision in an easy and humanly understandable way. Just by looking at the decision tree's structure, we (and also a physician) can tell which attributes are more important, and which are not. Therefore the decision trees have been often used for decision-making and knowledge extraction in various medical databases.

The paper is outlined as follows: the next section provides more detailed description of the methods used for intelligent data analysis. The novelty among used approaches is a multimethod approach, which in general produces a population of qualitative solutions represented with hybrids of different methods dynamically combined in non-predefined order [5]. For the reasons exposed above in this research only the methods, which can represent the extracted knowledge in a form of decision trees, are used. The database is presented in Section 3 and in Section 4 we present the results. The paper concludes with discussion and some final remarks.

## 2. Methods

Historically different approaches for knowledge extraction evolved [7], such as symbolic approaches and computational learning theory. Among them we can find many classical approaches, like decision trees, rules, rough-sets, case based reasoning, neural networks, support vector machines, different fuzzy methodologies, ensemble methods [8], but they all have some advantages and limitations. Evolutionary approaches (EA) are also a good alternative, because they are not inherently limited to local solutions [9]. Recently, taking into account the limitations of classical approaches many researchers focused their research on hybrid approaches, following the assumption that only the synergetic combination of single models can unleash their full power [10].

Current studies show that the selection of appropriate method for data analysis can be crucial for the success. Therefore, for a given problem, different methods should be tried to increase the quality of extracted knowledge. According to the previous paragraph a logical step would also be to combine different methods into one more complex methodology in order to overcome the limitations of a single method. We noticed that almost all attempts to combine different methods use loose coupling approach where the methods work almost independent of each other. Therefore a lot of “luck” and trying is needed to unify them into a “team”. Thus we decided to design a new approach that enables tight tangling of single methods. This approach is called a multimethod approach [4]. Opposed to the conventional hybrids our idea is to dynamically combine and apply different methods in not predefined order in the manner to solve a single problem or the decomposition of that problem.

Multimethod approach introduces the idea of a population of different intelligent systems - individuals that can produce multiple comparable good solutions, which are incrementally improved using the EA approach. In order to enable knowledge sharing between different methods the support for transformation between each individual method is provided. Initial population of intelligent systems is generated using different methods. In each generation different operations appropriate for individual knowledge representation are applied to improve existing and also to create new intelligent systems. That enables incremental refinement of extracted knowledge, with different views on a given problem. For example, using different induction methods such as different purity measures can be simply combined into a decision trees. As long as the knowledge representation is the same, a combination of different methods is not a big obstacle. The main problem is how to combine methods that use different knowledge representations (for example neural networks and decision trees). In such cases we provide two alternatives: (1) to convert one knowledge representation into another, using different already known methods or (2) to combine both knowledge representations into a single intelligent system.

The first alternative requires implementation of the knowledge conversion (for example conversion of a neural network into a decision tree). Such conversions are not perfect and some of the knowledge is normally lost, but conversions can produce a different aspect on a presented problem that can lead to better results.

The second alternative requires some cut-points where knowledge representations can be merged. In a decision tree internal nodes or decision leafs represent such cut points, i.e. a condition can be replaced by another intelligent system (for example support vector machine - SVM). We call such trees the hybrid decision trees.

### 2.1. Decision trees

One of the main advantages of using decision trees, in compare with other methods of machine learning, is very simple and clear representation of the path to acquired decision. Inducing a decision tree is a form of machine learning, where we extract knowledge from a set of examples (objects) and present it in a 2-dimensional form of a decision tree [6].

A decision tree is inducted on a training set, which consists of training objects. Every training object is completely described by a set of attributes (object properties) and class (decision, outcome). Attributes can be numeric or discrete, but numeric attributes are not suitable for learning a tree. Therefore they must be mapped into a discrete space.

There are two types of nodes in a decision tree: internal and external nodes. Each internal node (non-terminal node) contains a test of a specific attribute value. External nodes (terminal nodes, decision nodes, leaves) are labeled with a class, which represents a decision. Nodes are connected with edges (links). Edges are labeled with different outcomes of a test performed on an attribute in a source node.

For testing a decision tree a testing set is used. Testing set consists of testing objects described with the same attributes as training objects except that testing objects are not included in training set.

The results are described with specificity, sensitivity and total accuracy. Specificity is defined as the number of correctly classified children with normal cholesterol level divided by the number of all children with normal cholesterol level. Sensitivity is the number of correctly classified children with abnormal cholesterol level divided by the number of all children with abnormal cholesterol level. The overall quality of a decision tree is described with total accuracy.

## 3. The study

The study included 38 unselected patients with unprotected LMCA stenosis/occlusion. 24 (63.2 %) of the patients had acute coronary syndrome (11 with ST elevation myocardial infarction (STEMI) and 13 with non-ST elevation myocardial infarction) and 14 (36.8%) patients with angina pectoris. The average age of the patients was 67.63 ±12.76. The youngest patient was 38 and the oldest 86 years old. Procedures were successful in all cases. 30 (75%) patients had multi vessel disease. Patients were observed on average for 12.6±11.6 months. 2 (5.2 %) patients with STEMI died three days after procedure. There was 1 pre–stent restenosis in unprotected LMCA and one in-stent restenosis in bare metal stent not in LMCA. No stroke or intracranial haemorrhage was observed. We had one complication – major hemorrhage at puncture site which needed transfusion. No surgical therapy was needed in the time of observation.

The prevalence of cardiac events (such as: death, myocardial infarction, unstable angina, left ventricular failure, and life-threatening ventricular arrhythmias) in first months after surgery was 23.7%. 4-8 months after surgery the patients had control angiography. Statistically most significant risk factors are: STEMI shock ($p<0.01$), the position of stenosis ($p=0.033$), intra-aortic balloon pump (IABP) ($p<0.01$), diabetes ($p<0.01$), killip ($p<0.01$) and increase of troponin ($p<0.01$).

## 4. Results

Since the goal of intelligent data analysis with decision trees was to determine important risk factors and the number of patients was relatively small, we used the whole database for training a decision tree.

The most interesting decision tree showed some interesting facts about survival after LMCA stenosis [Figure 1]. It is very interesting that the most known risk factors such as the position of stenosis and the presence of acute coronary syndrome (ACS) did not appear in the decision tree. However, diabetes proved to be very important risk factor that was not expected. Patients with diabetes have a high risk for not surviving almost irrespective of ACS or STEMI shock. On the other hand, non-diabetic patients that don't need Intra-aortic balloon pump (IABP) all survived.

For non-diabetic patients with inserted IABP it is very important survival factor that they don't have low total cholesterol level (under 3.66). It is interesting that the decision tree correctly showed that low cholesterol level is in this situation a bad factor for survival.

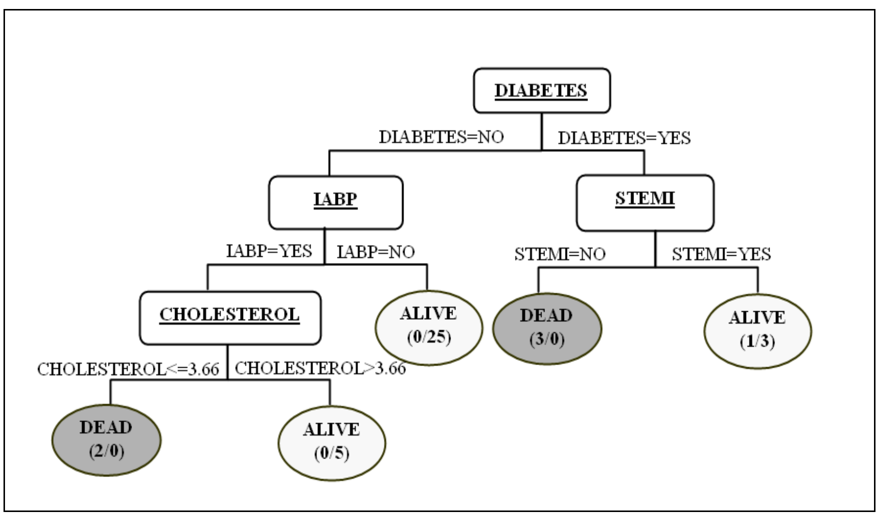


Figure 1. The decision tree for determining survival factors after LMCA stenosis. The leafs in the tree are represented with circle marked with decision (dead or alive after the observation period). The first number in each decision leaf in the bracket represents the number of all patients that did not survive and the second number is the number of survived patients.

## 5. Discussion and conclusions

The induced decision tree showed that for long term survival after LMCA it is very important that the patient is not diabetic and that his/hers cholesterol level is not too low.

The risk factors that have been discovered with intelligent data analysis have also been confirmed with statistical analysis. The influence of diabetes on prognosis after PTCA is statistically significant ($p<0.01$; $\chi^2=8.464$). The prognosis for non-diabetic patients is better than for diabetic patients.

The influence of IABP on non-diabetic patients was investigated with $\chi^2$ test. Insertion of IABP is significantly important for long-term survival of non-diabetic patients ($p<0.042$; $\chi^2=7.619$).

The decision tree outlined the level of patient's total cholesterol as an important factor for log-term survival of non-diabetic patients with IABP because patients with low cholesterol level (<3.66 mmol/L) all die and patients with cholesterol level higher than 3.66 mmol/L all survive. This conclusion is very interesting; however we have to consider that it was obtained on relatively small number of patients. Therefore it would be interesting to examine its reliability on larger dataset.

The diabetic patients are considered in the separate branch of the decision tree. As we already established, diabetes is significantly important risk factor. It is interesting that ST elevation myocardial infarction (STEMI) after PTCA is not significantly important for survival of patients with diabetes. Again we have to deal with very small number of patients and consequently the reliability has to be reconsidered.

We can conclude that the application of intelligent data analysis method for determining possible risk factors for long-term survival after PTCA LMCA stenosis was very successful. Some interesting conclusions were obtained that deserve additional considerations and examinations.

## References


[1] Takaro T, Pifarre R, Fish R. Veterans Administration Cooperative Study of medical versus surgical treatment for stable angina—progress report. Section 3. Left main coronary artery disease. Prog Cardiovasc Dis 1985;28:229–34.

[2] Sciagra R, Tebbe U, Vogt A, Wiegand V, Kreuzer H, Neuhaus KL. Occlusion of the common trunk of the left coronary artery. Physiopathological features and clinical findings. G Ital Cardiol 1986;16:516–21.

[3] Kokol P, Završnik J, Zorman M, Malčić I, Kancler K. Participative Design, Decision Trees, Automatic Learning and Medical Decision Making. In: Brender J. Medical Informatics Europe'96. Ios Press. 1996:34:501-505.

[4] Povalej P, Lenič M, Molan Štiglic M, Skerbinjek-Kavalar M, Završnik J, Kokol P. A journey through clinical applications of multimethod decision trees. Lect. notes comput. sci., 249-253.

[5] Lenič M, Kokol P. Combining classifiers with multimethod approach. In: Ajith A. et al. Second international conference on Hybrid Intelligent Systems, Soft computing systems: design, management and applications. IOS Press. 2002:374-383.

[6] Quinlan J.R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publ. 1993.

[7] Thrun S., Pratt L. Learning to Learn. Kluwer Academic Publishers. 1998.

[8] Dietterich T.G. Ensemble Methods in Machine Learning. In: First International Workshop on Multiple Classifier Systems. Lecture Notes in Computer Science. Springer Verlag. 2000:1-15.

[9] Goldberg D.E. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison Wesley. Reading MA. 1989.

[10] Iglesias C.J. The Role of Hybrid Systems in Intelligent Data Management. In: The Case of Fuzzy/neural Hybrids, Control Engineering Practice. 1996:4:6: 839-845.



Address for correspondence

petra.povalej@uni-mb.si, vojko.kanic@sb-mb.si, kokol@uni-mb.si