TEKA. COMMISSION OF MOTORIZATION AND ENERGETICS IN AGRICULTURE – 2012, Vol. 12, No.4, 171-175

The prediction of myocardial infarction consequences as a result of vectorcardiography research using «Decision trees» data mining algorithm

Evelina Musayeva[1], Inna Belaya[2]

[1] Volodymyr Dahl East-Ukrainian National University, Lugansk, Ukraine
[2]State institution "Lugansk State Medical University", Lugansk, Ukraine

Summary. The article represents the possibility of the new MTM-SKM cardiodiagnostic complex use in the acute myocardial infarction diagnostics. The prognostic criteria for adverse prognosis in acute coronary pathology can be a small group of vectorcardiogram indicators, found with the help of "Decision trees" data mining algorithm.
Key words: vectorcardiography, "Decision tree" algorithm, prediction.

INTRODUCTION

Widespread coronary heart disease has become the most important health issue of our time. Today the largest part of mortality structure belongs to the diseases associated with the circulatory system [Musayeva, Belaya 2011]. At approximately 5% of patients hospitalised with acute chest pain acute coronary syndrome is impossible to recognize, which causes the development of myocardial infarction (MI); the mortality rate, however, remains relatively high in the world, as well as in Ukraine. Therefore, it is especially important to find early MI diagnostics methods. In this regard, the new hardware diagnostic methods together with information technologies can detect cardiac abnormalities in cases where conventional methods of laboratory and instrumental diagnostics can not provide any comprehensive information for early diagnostics and the possibility to predict the disease development [Parkhomenko 2010, Tseluyko 2009, Kopitsa 2011, Yang 2011, Ghista, Acharya, Nagenthiran 2010, David Strauss, Charles Olson, Katherine C. 2009] .

In recent years, MTM-SKM polygraph has been used to study the electromotive force of the heart (EMF) changes in patients with acute myocardial infarction period at Luhansk Clinical Multihospital № 1, that combines electrical (ECG) and vektorkardiography (VCG). The device is designed by Severodonetsk Scientific Production Association "Microterm" together with Volodymyr Dahl East-Ukrainian National University under the guidance of professor B. Yu Dobrin. This method allows to determine the features of the EMF distribution over the surface of the heart in real time except for the "dead zones", and to increase the studied sections of the route up to 3000 times.

OBJECTS AND PROBLEMS

VCG analysis in patients with acute myocardial infarction revealed certain patterns of changes of the heart EMF. Firstly, there is a sudden spatial displacement of QRS loop, especially its initial part, from the infarction localization in the opposite direction. Separate areas or large areas of myocardium, affected by necrobiosis, do not give their portion of normal electrical power and therefore oppositely directed unbalanced electrical forces of intact areas "pull" QRS loop in the opposite direction from the lesions. Secondly, together with acute coronary pathology the direction of QRS loop route changes as the indicator of abnormal spread of excitation wave over myocardium. The process of myocardium impulse conduction can be disrupted in the form of additional QRS loop nodes, as well as in the concentration of time stamps (the reflection of speed indicators changes). Intersections, convexities, and additional QRS loop poles are formed. In addition, the reduction in the total QRS loop area is typical, that becomes the most distinct in the case of left ventricle aneurysm. Thirdly, unclosed QRS and T loops, as well as ST ECG intervals displacement indicate the unbalanced electrical forces presence in the phase of depolarization/repolarization transition. The positive pole of electrical forces of damage vector is directed to the necrobiosis lesion and is also referred to as ST vector. T loop in acute myocardial infarction may have a normal size, but the decrease of its deviation angle is possible with its direction to the initial part of QRS rout and even further. T loop shape resembles a horseshoe [David Strauss, Charles Olson, Katherine C. 2009, Belaya 2010, Belaya 2011].

Information, received after VCG, enables to confirm the focal myocardial changes and to get the additional information about heart EMF for more objective assessment of myocardium pathological changes [Belaya 2010, Belaya 2011]. The use of data mining can provide a narrow group of VCG parameters that determine an adverse outcome, and to represent consistent pattern in analytical form that will enable the prognosis in patients with atypical acute coronary pathology, uncertain ECG dynamics and questionable laboratory data.

THE DECISION OF THE TASK

To implement this task, Decision Trees data mining algorithm can be used, which is one of the most popular methods in classification and prediction problems solving. Sometimes this data mining method is also called Decision Rules, Classification and Regression Trees. Decision Trees allow to create classification models in cases, where it is difficult for an analyst to formalize knowledge [Berestneva, Dobrianskaya, Muratova 2005, Panchenko 2004, Zubov, Ulshin, Gorbunov 2010, Ulshin, Klimchuk 2010].

There are basic concepts of the Decision Trees theory. One of the first concepts is an object that is observed. The sign of the independent variable is an attribute concept. The dependent variable or attribute that defines the class of an object (the death in this case) is called class label. Node is the internal node of the tree, the check node; leaf is the end node of the tree, the decision node and test is a condition in the node [Akobir Shahidi].

Decision Trees are a way of representing the rules in hierarchical, sequential structure, where each object has a single node that gives the solution.

The rule is a logical construction, presented in the form of "if ... then ...".

All problems solved by the Decision Trees algorithm can be grouped into three types. Description, classification and regression [Murthy 1997]. In this paper we solve the problem of classification, the grouping of objects into one of known classes. Class in this particular case is a logical variable that attributes an object to the "survivors" class or "dead" class.

Learning set T (50 patients) is given, that contains objects (examples), each of them is characterized m attributes (QRS loop clockwise rotation; loops location in the coordinate system; the direction of main vector in the coordinate system; the distinctive route look; the maximum vector size and the loops area; the speed of excitation spread through QRS, P and T loops; the presence of unclosed QRS and T loops; angular divergence of QRS-T and QRS-P loops). One of the attributes indicates that an object belongs to a particular class - dead or survivors.)

The idea of decision trees building from a set T, first proposed by Hunt, will be given according to R. Quinlan [Ross Quinlan 1993].

Let $\{C_1, C_2, ..., C_k\}$ represent the classes (class label value), then there are three situations:

1. The set T contains one or more examples of the same class C_k. Then, the decision tree for T is a leaf that defines class C_k;

2. The set T does not contain any examples, it is empty. Then it is a leaf again, and the class associated with this leaf, is selected from another set, different from T, for example, from the set associated with the parent;

3. The set T contains examples that belong to different classes. In this case, the set T must be divided into subsets. To do this, one of the signs is chosen, that has two or more distinct values $O_1, O_2, ..., O_n$. T is divided into subsets $T_1, T_2, ..., T_n$, with each subset T_i containing all the examples that have value O_i for the chosen sign. This procedure will be recursively conducted until an end set consists of examples of the same class.

The above procedure is the basis for many modern decision trees algorithms; this method is also known as Divide and Conquer.

As all objects have been previously attributed to the known classes, the process of decision tree building is called Supervised Learning.

The algorithm of decision tree building does not require a user to select the input attributes (independent variables). The algorithm input can use all existing attributes, the algorithm will choose more substantial among them, and only they will be used to build a tree. In comparison with neural networks, for example, it is much easier to work, because neural networks selection of incoming attributes significantly affects the studies [Barseghyan, Kupriyanov, Stepanenko, Holod 2004, Paklin 2012].

Deductor Studio Academic software is used as a tool of data analysis. It is a complete analytical platform.

Thus, with the help of this software the decision tree has been constructed and rules has been formed. About 50 patients, which have been examined earlier, were chosen as an initial data sample.

The results are interpreted with the available visualizers. For a start we will analyze the contingency table (table 1).

Table 1. Contingency table

	Classified		
Factually	False	True	In total
False	36	2	38
True	1	11	12
In total	37	13	50

The examples that have been correctly recognized are located diagonally, in other cells the examples that have been attributed to another class are located. In this case, the tree has correctly classified 47 cases and three cases have been misclassified. We can say that the result has an accuracy of 94% (fig. 1).

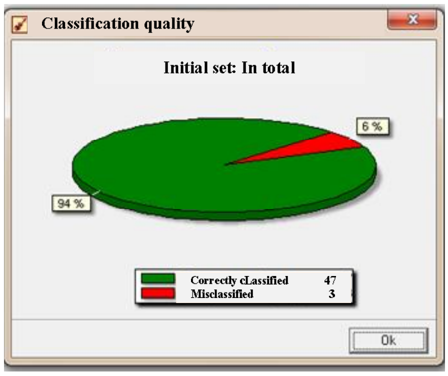


Fig. 1. Classification quality

Move to the main visualizers for this algorithm – “Decision Tree” and “Rules” (fig. 2, 3). Apparently, the decision tree is not very complicated, most factors have been cut off, that is, their influence on the fact of death in patients is minimal or absent. This visualizer allows to see the examples that were grouped to a particular node, and the information on the node.

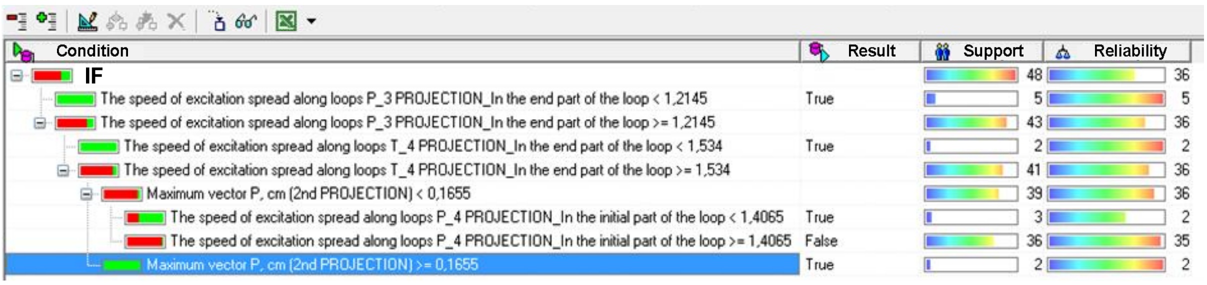
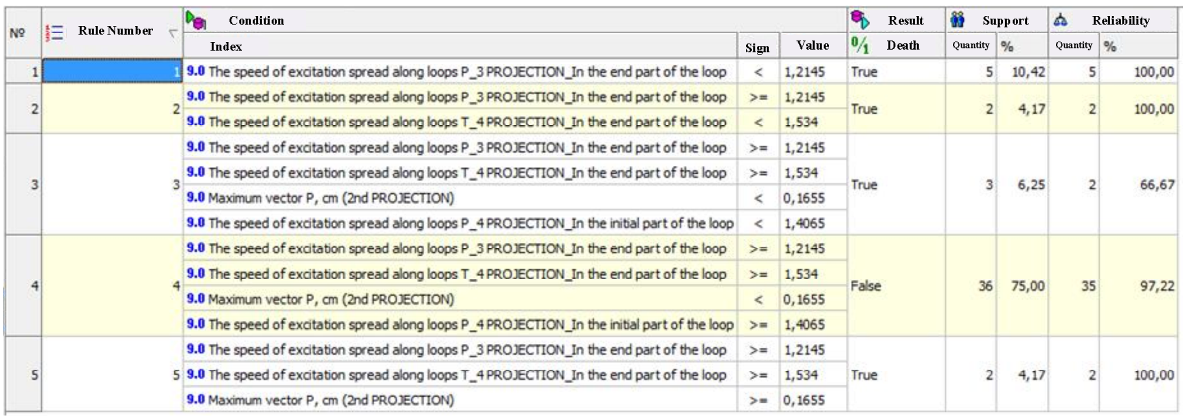


Fig. 2. “Decision tree” tab



№	Rule Number	Condition: Index	Sign	Value	Result: Death	Support: Quantity	Support: %	Reliability: Quantity	Reliability: %
1	1	9.0 The speed of excitation spread along loops P_3 PROJECTION_In the end part of the loop	<	1,2145	True	5	10,42	5	100,00
2	2	9.0 The speed of excitation spread along loops P_3 PROJECTION_In the end part of the loop	>=	1,2145	True	2	4,17	2	100,00
		9.0 The speed of excitation spread along loops T_4 PROJECTION_In the end part of the loop	<	1,534					
3	3	9.0 The speed of excitation spread along loops P_3 PROJECTION_In the end part of the loop	>=	1,2145	True	3	6,25	2	66,67
		9.0 The speed of excitation spread along loops T_4 PROJECTION_In the end part of the loop	>=	1,534					
		9.0 Maximum vector P, cm (2nd PROJECTION)	<	0,1655					
		9.0 The speed of excitation spread along loops P_4 PROJECTION_In the initial part of the loop	<	1,4065					
4	4	9.0 The speed of excitation spread along loops P_3 PROJECTION_In the end part of the loop	>=	1,2145	False	36	75,00	35	97,22
		9.0 The speed of excitation spread along loops T_4 PROJECTION_In the end part of the loop	>=	1,534					
		9.0 Maximum vector P, cm (2nd PROJECTION)	<	0,1655					
		9.0 The speed of excitation spread along loops P_4 PROJECTION_In the initial part of the loop	>=	1,4065					
5	5	9.0 The speed of excitation spread along loops P_3 PROJECTION_In the end part of the loop	>=	1,2145	True	2	4,17	2	100,00
		9.0 The speed of excitation spread along loops T_4 PROJECTION_In the end part of the loop	>=	1,534					
		9.0 Maximum vector P, cm (2nd PROJECTION)	>=	0,1655					

Fig. 3. “Rules” tab

Target attribute: Death			
№	Number	Attribute	Significance, %
1	63	The speed of excitation spread along loops P_3 PROJECTION_In the end part of the loop	38,456
2	55	Maximum vector P, cm (2nd PROJECTION)	22,552
3	64	The speed of excitation spread along loops P_4 PROJECTION_In the initial part of the loop	19,974
4	35	The speed of excitation spread along loops T_4 PROJECTION_In the end part of the loop	19,017
5	51	678	0,000
6	50	677	0,000

Fig. 4. "The Relevant attributes" tab

It is easier to see the importance of factors or attributes in the Relevant Attributes visualizer (fig 4).

This visualizer demonstrates how much the output field depends on each of the input factors. The more the significance of the attribute is, the larger contribution it makes into the classification. In this case, the largest contribution is the speed of excitation spread over R loop (mV/c) in BA3 in the end part of the loop.

Rules visualizer (fig. 3) represents the complete list of the rules, according to which a patient can be attributed to a particular class.

The data is presented in a table. The fields of the table are [BaseGroup™ Labs Company, 1995-2012]:

- rule number;
- condition that uniquely identifies belonging to a group;
- consequence – is a patient dead or not;
- support - the number and percentage of the original sample of examples that meet this condition;
- reliability - the percentage of correctly recognized examples that meet this condition, the total number of examples that meet this condition.

This table data analysis allows to assert, up 97%, what exactly influences the fact of death, what is the value of this influence (support) and what is the rule accuracy. In this case, it is obvious that of the entire list of rules the most reliable is rule number 4.

IF

The speed of excitation spread over R loop (mV/s) Ba3 in the end part of the loop> = 1.2145

AND

the speed of excitation spread over T loop (mV/s) BA4 in the end part of the loop> = 1.534

AND

the maximum vector value of P loop in mV (cm) Ba2 <0.1655

AND

the speed of excitation spread in R loop (mV/s) BA4 in the initial part of the loop> = 1.4065

THEN

Death = False (belongs to the survivors)

CONCLUSIONS

Thus, the prognostic significant set of vectorcardiography indicators of favourable course of acute myocardial infarction with 97% accuracy, found with "Decision Trees" algorithm, are:

1. The speed of excitation spread over P loop in the end part in BA3 > = 1,2145 mV / s.
2. The speed of excitation spread over P loop in the initial part in BA4 > = 1,4065 mV / s.
3. The speed of excitation spread over T loop in the end part in BA4> = 1,534 mV / s.
4. The value of P loop maximum vector in Ba2 <0.1655 cm

Failure of at least one criteria or group of criteria means absolutely unfavourable outcome. In this case, further research is needed.

REFERENCES

1. **Akobir Shahidi**, Decision Trees Internet Resource - general principles. Access mode. http://www.basegroup.ru/library/analysis/tree/descriptio/
2. **Barseghyan A., Kupriyanov M., StepanenkoV., Holod I., 2004.:** Methods and models of data analysis: OLAP and Data Mining. St. Petersburg.: BHV-Petersburg, 336.
3. BaseGroup™ Labs Company, 1995-2012.: Internet resource. Analyst's Guide Deductor 5.2. Download mode www.basegroup.ru/download/guide_analyst_5.2.0.pdf
4. **Belaya I., 2010.:** The value of Policardiography in myocardial infarction complications prediction. Ukrainian therapeutic magazine, № 3, p. 101-106.
5. **Belaya I., 2011.:** System approach to the acute period of myocardial infarction diagnosis. Russian cardiological magazine, № 4 (90), p. 46-52.
6. **Belaya I., 2011.:** Vector-electrocardiographic parallels in the diagnosis of myocardial infarction. Heart and blood vessels, № 3, p. 105-117.
7. **Berestneva O., Dobrianskaya R., Muratova Ye., 2005.:** Application of Data Mining technology in labor outcome prediction. Reports of the XII All-Russian

Conference Mathematical Methods for Pattern Recognition MMRO-12, Moscow.

8. **David G. Strauss, Charles W. Olson, Katherine C., 2009.:** Wu et al. Vectorcardiogram synthesized from the 12-lead electrocardiogram to image ischemia. J. of Electrocardiol, Vol. 42, № 2, p. 190-197
9. **Ghista D., Acharya U., Nagenthiran T., 2010.:** Frontal plane vectorcardiograms: theory and graphics visualization of cardiac health status. J. Med. Syst., Vol. 34, № 4, p. 445-458.
10. **J. Ross Quinlan. 1993.**: C4.5: Programs for Machine learning. Morgan Kaufmann Publishers.
11. **Kopitsa N., 2011.:** Fast identification and prognosis in patients with acute coronary syndrome. Ukraine therapeutic magazine, № 1, p. 102-106
12. **Musayeva E., Belaya I., 2011.:** Analysis of existing methods of acute coronary syndrome prediction. Journal of Volodymyr Dahl East-Ukrainian National University, № 2 (156), S. 360-364.
13. **Paklin N., 2012.:** Business Intelligence: From Data to Knowledge: Tutorial - 2nd ed., Corrected, St. Petersburg.: Peter, 704.
14. **Panchenko M., 2004.:** Development of mathematical, algorithmic and program software for decision making in various fields based on the Decision Trees method tree: dissertation for the scientific degree of Candidate of Technical Sciences: spec. 05.13.06 "Automated systems and advanced information technologies", Kyiv, 16 p.
15. **Parkhomenko A., 2010.:** The problem of myocardial infarction in 2009: Making new discoveries, do not forget about available opportunities. Health of Ukraine, № 4 (233), p. 16.
16. **S.Murthy, 1997.:** Automatic construction of decision trees from data: A Multi-disciplinary survey.
17. **Tseluyko V., 2009.:** A complex clinical case. How to make the correct diagnosis? Health of Ukraine, № 17 (222), p. 18.
18. **Ulshin V., Klimchuk S., 2010.:** Decision support system for bridge granes diagnostics on the basis of cases. TEKA Commission of Motorization and Power Industry in Agriculture, V. XD, p. 5-14.
19. **Yang H., 2011.:** Multiscale recurrence quantification analysis of spatial cardiac vectorcardiogram signals. IEEE Trans. Biomed. Eng., Vol. 58, Issue 2.- p. 339-347.
20. **Zubov D., Ulshin V., Gorbunov A., 2010.:** Analogue complexing algorithm's usage in Data Mining. TEKA Commission of Motorization and Power Industry in Agriculture, V. XD, p. 81-90.

ПРОГНОЗИРОВАНИЕ ИСХОДОВ ИНФАРКТА МИОКАРДА ПО РЕЗУЛЬТАТАМ ВЕКТОРКАРДИОГРАФИЧЕСКОГО ИССЛЕДОВАНИЯ ПРИ ПОМОЩИ АЛГОРИТМА DATA MINING «ДЕРЕВО РЕШЕНИЙ»

Эвелина Мусаева, Инна Белая

Аннотация. В статье представлены возможности использования нового кардиодиагностического комплекса МТМ-СКМ в диагностике острого инфаркта миокарда. В качестве прогностических критериев неблагоприятного течения острой коронарной патологии может быть использована выделенная с помощью алгоритма Data Mining «Деревья решений» узкая группа показателей векторкардиограммы.

Ключевые слова: векторкардиография, алгоритм «Деревья решений», прогноз