Cent. Eur. J. Med. • 7(2) • 2012 • 183-193
DOI: 10.2478/s11536-011-0142-x

VERSITA

Central European Journal of **Medicine**

Different Decision Tree Induction Strategies for a Medical Decision Problem

Research Article

Robert Burduk*, Michal Wozniak

Department of Systems and Computer Networks,
Faculty of Electronics, Wroclaw University of Technology
Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland

Received 5 October 2011; Accepted 12 December 2011

Abstract: The paper presents a comparative study of selected recognition methods for the medical decision problem –acute abdominal pain diagnosis. We consider if it is worth using expert knowledge and learning set at the same time. The article shows two groups of decision tree approaches to the problem under consideration. The first does not use expert knowledge and generates classifier only on the basis of learning set. The second approach utilizes expert knowledge for specifying the decision tree structure and learning set for determining mode of decision making in each node based on Bayes decision theory. All classifiers are evaluated on the basis of computer experiments.

Keywords: *Acute abdominal pain • Univariate and multivariate decision trees • Bayes decision theory • Multistage classifier • Medical decision support systems*

1. Introduction

Medical diagnosis is a very important and attractive area for implementing decision support systems. About 11% of expert systems are dedicated to medically aided diagnosis, and about 21% of papers connected with application of those methods are illustrated by medical cases [1]. One of the first and well-known expert systems dedicated to medical aided diagnosis is MYCIN [2], considered by many researchers as the exemplar of expert system. In a 2001 article, Sims et al [3], describe the medical decision support system as *software that is designed to be a direct aid to clinical decision-making in which the characteristics of an individual patient are matched to a computerized clinical knowledge base, and patient-specific assessments or recommendations are then presented to the clinician and/or the patient for a decision*. Thus, the concept of a clinical decision support system is not new. There are many papers that describe these systems: some present reviews of working medical decision support software [4–6], whereas others are relate the problems involved in choosing the best method of classification for the particular medical task [7–12] or are cluster analyses based on a hierarchical dendrogram [13,14].

In many cases, these system might utilize different kinds of learning materials. On one hand, we might obtain rules or any other kind of knowledge from human experts; on the other hand, we might generalize the knowledge on the basis of learning sets. Of note, the learning examples are delivered from databases, and each instance is labeled by a human expert or on the basis of crucial examination results. We should ask ourselves whether it necessary to use all available learning materials during the decision support design project. This paper presents a particular medical task – diagnosis of acute abdominal pain – where for the construction of multistage classifier we might use a learning set and/or human expert's knowledge as well. For the problem under consideration, the expert proposed the structure of a decision tree and specified which attributes had to be tested in each node.

The paper describes our research into qualities of classifiers based on a Bayesian approach that utilizes

* *E-mail: robert.burduk@pwr.wroc.pl*

the schema of decision given by the expert and learning set for decision making in each node of the tree. We compare results of the those experiments to the quality of classifiers obtained via several machine learning procedures that do not use expert knowledge during the learning; we are particularly interested in comparing these two main approaches based on the univariate and multivariate trees. Additionally, we will discuss whether the qualities of the obtained classifiers could be improved by modifying the classification strategy for the Bayesian classifier.

The content of the work is as follows. Section 2 introduces the basis for a Bayesian approach, the idea of top down induction of the decision tree, and shortly introduces the methods used during computer experiments. In the next section, we describe a mathematical model of the acute abdominal pain decision problem. We then present conditions and results of the experimental investigations of the proposed algorithms. The last section concludes the paper.

2. The multistage recognition task

The basic idea involved in a multistage approach is to break up a complex decision into several simpler classifications [15]. Decision tree and hierarchical classifiers are two possible approaches to the multistage pattern recognition. Hierarchical classifiers are a special type of multistage classifiers that allow the rejection of class labels at intermediate stages. The synthesis of a hierarchical classifier is a complex problem. It involves specification of the following components [16]:

1. design of a decision tree structure;
2. selection of features used at each terminal node of a decision tree;
3. choice of decision rules for performing the classification.

We now summarize the two approaches. The first uses a given decision tree structure and a set of features for each tree's node. This method focuses its attention on the decision rules construction based on Bayesian approach for each node. The second group uses different methods of tree induction, such as univariate or multivariate trees.

2.1. Bayesian hierarchical classifier

Among different concepts and methods of using "uncertain" information in the pattern recognition, Bayes decision theory is efficient and attractive from the theoretical point of view. This approach consists of an assumption [17] that the feature vector $x=(x^{(1)},x^{(2)},\ldots,x^{(d)})$ (describing the object under recognition) and the class label $j\in\{1,2\ldots,M\}$ are the realization of the pair of random variables *X*, *J*. For medical use, *X* describes the result of a patient's examinations and *J* denotes the patient's state. The random variable *J* is described by the prior probability p_j, where

$$p_j=P(J=j) \tag{1}$$

X has probability density function

$$f(X=x|J=j)=f_j(x) \tag{2}$$

for each *j*, the conditional density function. These parameters can be used to enumerate posterior probability according to Bayes' formulae:

$$p(j|x)=\frac{p_j f_j(x)}{\sum_{k=1}^{M} p_k f_k(x)} \tag{3}$$

The formalisation of the recognition in the case under consideration implies the setting of an optimal Bayes decision algorithm $\Psi(x)$, which minimizes the expected value of the so-called loss function that describes the cost of the wrong classification [17]. For the well-known 0-1 loss function, the classifier assures the lowest value of the probability of the misclassification, and the decision rule chooses the classes for which the posterior probability achieved the largest values [18]

$$\Psi(x)=\arg\max_{i\in 1,\ldots,M} p(i|x) \tag{4}$$

In an actual situation, the *a priori* probabilities and the conditional density functions are usually unknown. Furthermore, we often have no reason to decide that the prior probability is different for each of the decisions. Instead, we can use the expert rules and/or the learning set for the constructing decision algorithms.

The design of a decision tree structure in our approach to the hierarchical classifier is based on human expert knowledge. In our consideration the decision rules are based on the probabilistic approach. The Bayes hierarchical classifier uses the Bayes theorem to design a classifier in each intermediate node.

The Bayesian hierarchical classifier consists of a sequence of actions (see Figure 1). These actions are the simple classification tasks executed in individual nodes of the decision tree. Some specific features are measured on every level of the decision tree. At the first stage features x_0 are measured, at the second stage features x_1 are measured, and so on. Every set of features comes from the whole vector of features. In every node of the decision tree, the classification is executed according to the Bayes rule. The decisions $i_1,i_2,\ldots,i_N$ are the results of recognition in the suitable node of the tree. At the last (N-th) stage, the decision made i_N indicates a single class. This class is the result of the Bayesian hierarchical classifier.

In our task of classification, the number of classes is equal to NC. The logic of making the decision is represented using the decision tree. The terminal nodes are labeled with the number of the classes from the M={1,2,...,NC}. The nonterminal nodes are labeled by the numbers of 0 , NC+1, NC+2… reserving 0 for the root-node.

We introduce the notation for the received model of the multistage recognition [19]:

M(n) – the set of numbers of nodes, which distance from the root is n, n=0,1,2,...,N.

In particular M(0)={0}, M(*N*)=M,

$\bar{M}(n)$ – the set of the interior node numbers (non terminal),

$M_i \subseteq M(N)$ – the set of class numbers attainable from the i-th node ($i \in \bar{M}$),

M^i – the set of numbers of the immediate descendant nodes i ($i \in \bar{M}$),

m_i – number of immediate predecessor of the i-th node ($i \neq 0$).

The Bayes hierarchical classifier is an example of the probabilistic model of pattern recognition. In this model, the class of the pattern being recognised as $j_N \in M(N)$ is the realization of random variable J_N and observed features x are realizations of random variable X .

Our target now is to calculate the so-called multistage recognition strategy $\pi_N = \{\psi_i\}_{i \in \bar{M}}$, that is, the set of recognition algorithms in the form

$$\Psi_i : X_i \to M^i, \quad i \in \bar{M} \tag{5}$$

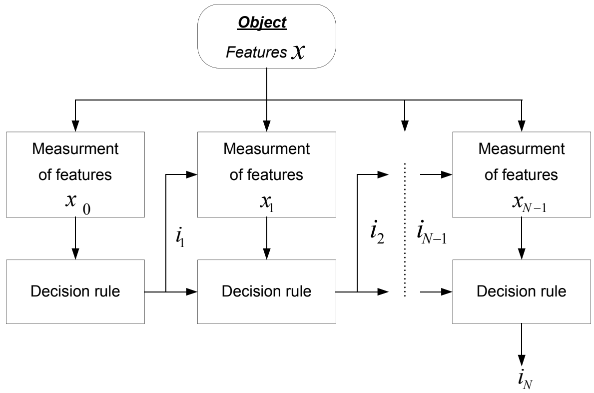
Equation (5) is the decision rule (recognition algorithm) used at the *i*-th node, which maps the observation subspace to the set of immediate descendant nodes of the *i*-th node.

The strategy of the decision tree classifier represents the logic of making the decision. We favour two possible strategies:the first is the locally optimal strategy, which minimizes the misclassification for particular nodes of a tree. Its decision rules are mutually independent. There are no relationships between nodes in logic expressed using the misclassification. The recognition algorithm at the *n*-th stage is

$$\bar{\Psi}_i(x_i) = \arg\max_{k \in M^i} p(k) f_k(x_i) \tag{6}$$

The second, globally optimal strategy, minimizes the mean probability of misclassification. The decision rules are mutually dependent through the empirical probability of the correct classification. The recognition algorithm at the *n*-th stage is as follows:

Figure 1. Bayesian hierarchical classifier



$$\Psi_i^*(x_i) = \arg\max_{k \in M^i} Pc(k) p(k) f_k(x_i) \tag{7}$$

where $Pc(k)$ is the empirical probability of the correct classification at the next stages if at the n-th stage decision i_{n+1} is made.

As we mention below, in practice the unknown probabilistic characteristics (values of the *a priori* probabilities and probability density functions) are replaced by their estimators obtained via parametric or nonparametric approaches [17].

2.2. Decision tree induction

Eecision tree induction algorithms have been under development for several years [20,21]. From the mathematical point of view, they propose a way to estimate discrete functions that could be adapted to classification tasks. From the practical point of view, the decision trees achieve satisfying results in many actual decision tasks. Among different methods of tree training, the *top down decision tree induction* concept is frequently used. Algorithms based on the aforementioned idea train a tree from a root node to leaf using a splitting attribute's choosing measure. The most famous representative of an algorithm family using the aforementioned concept is ID3, developed by Quinlan [22]. ID3 uses the *information gain* measure to decide which attribute should be tested in a given node. The proposed measure evaluates the homogeneity of subsets of a training set (according to the given class labels) that were obtained on the basis of the original set split using the chosen attribute values. The pseudocode of ID3 is presented in the Appendix.

The descendants of IDs improve its main features. The main disadvantage of *information gain* is that it prefers

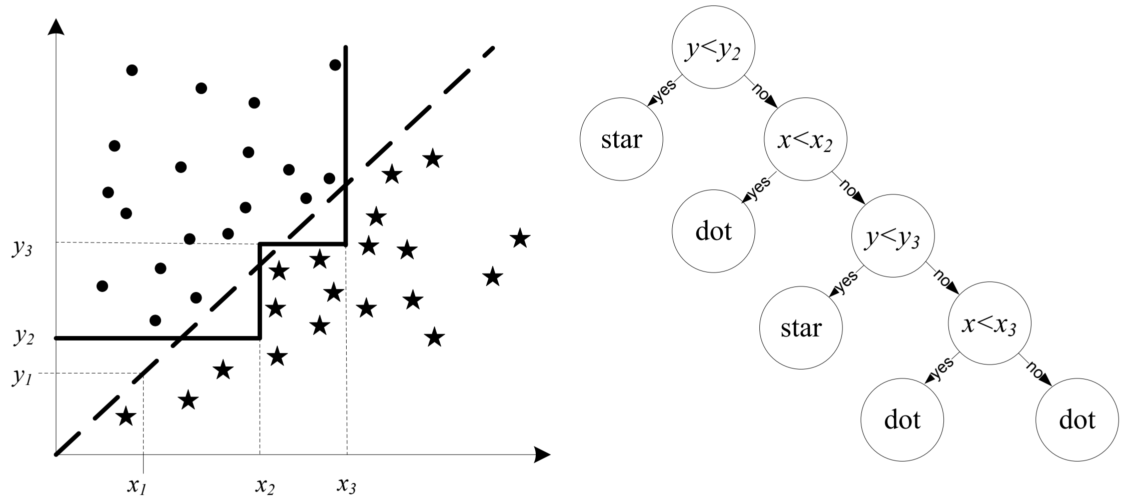


Figure 2. (left) Two class recognition problem and liner classifier boundary (dashed line) and univariate decision tree classifier boundary (solid line); (right) univariate decision tree structure for the problem presented in the left picture

features with high number of values. C4.5 [23] uses another attribute measure, the *information ratio*. Both measures are based on information theory that uses Shannon's entropy as a measure; many other measures are proposed, such as the Gini metric used by CART or the χ^2 statistic [24], to enumerate only a few. The aforementioned algorithms also propose methods that protect tree classifiers against overtraining such as reduce-error pruning or rule post pruning, and show how to deal with continuous attributes, how to handle with missing attribute values and attributes with weights, how to reduce computational complexity, and how to use algorithm in the distributed computing environments.

Figure 2 presents an exemplary decision problem.

Let us note that the classifier depicted by the dashed line separates classes well. It works according to a very simple decision rule

$$\text{if } y > \frac{y_1}{x_1}x \text{ then "dot" else "star"} \qquad (8)$$

This classifier could be interpreted as a multivariate decision tree with one node. The univariate decision tree prefers the more complex separating curve depicted by the solid line. Multivariate decision trees usually offer simpler structures than univariate decision trees. We can also assume that a complex structure is more susceptible to overtraining. Moreover, the univariate decision tree induction uses "greedy" search methods. As a criterion, local discrimination power measures like the aforementioned *information gain* or Gini index are used. It means that they do not guarantee finding an optimal classifier. Additionally, we can suppose that the univariate tree is usually not the best solution because in [25] the author proved that for a set of attributes, the best pair could consist of two attributes which differ from the two best individual ones.

There are several propositions for multivariate decision tree training. Some of them suggest using classifiers in each node, e.g. LMDT uses a linear classifier [26], in [27] the author proposes using a Bayesian classifier. In [28], the authors propose a LMT algorithm that uses an ID3 algorithm for discrete features and then linear regression for the others. Other approaches use traditional or heuristic feature selection methods in each node [29,30].

3. Model of acute abdominal pain diagnosis

The first mathematical model of acute abdominal pain (APP) was reported in [31]. We simplified that model in cooperation with the experts from the Clinic of Surgery, Wroclaw Medical Academy, who regard the stated problem of diagnosis as very useful.

It leads to the following classification of the AAP:

1. cholecystitis
2. pancreatitis
3. non-specyfic abdominal pain
4. rare disorders of "acute abdominal"
5. appendicitis
6. divercitulitis
7. small-bowel obstruction
8. perforated peptic ulcer

Although the set of symptoms necessary to assess the existing APP correctly is relatively wide, in practice for diagnosis it results in 31 (non-continuous) examinations that are used; these are presented in the Appendix. Since the abdominal area contains many different organs, it is divided in smaller areas [32]. One division method uses one median sagittal plane and one transverse plane that passes through the umbilicus at right angles. This method divides the abdomen into four left and right upper, left and right lower quadrants. For our study, we used the more precise description of abdominal pain location (Figure 3).

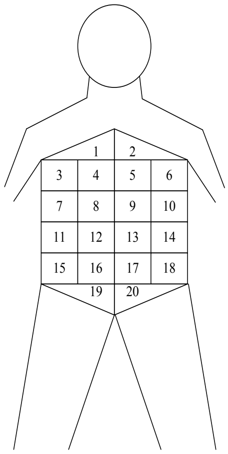


Figure 3. Encoding of the location of abdominal pain

The expert-physicians provided the decision tree depicted in Figure 4 [31]. The number of leafs are the numbers of the diagnosis presented above. The numbers in the nodes correspond with the following diagnosis:

9. acute enteropathy
10. acute disorders of the digestive system
11. others

4. Experimental investigation

The aim of the experiment is to compare the errors of Bayesian classifiers with the quality of classifiers obtained via machine learning algorithms. Additionally, we would like to compare qualities of the univariate with multivariate trees.

The following classifiers and fusion methods were chosen:

1. Multistage classifiers used heuristic decision tree and Bayesian classifier in each node according the local optimal strategy. For these classifiers the estimators of the conditional probability density function were obtained via k_n -Nearest Neighbor [18].
2. Multistage classifiers used heuristic decision tree and Bayesian classifier in each node according the global optimal strategy. For these classifiers the estimators of the conditional probability density function were obtained via k_n -Nearest Neighbor.

Classifiers based on univariate decision tree produced by

3. The C4.5 algorithm [23]
4. The BFTree method, which uses binary split of attribute values [33]
5. The LADTree algorithm, which generates a multiclass alternating tree [34]

Classifiers based on the multivariate decision tree were obtained via:

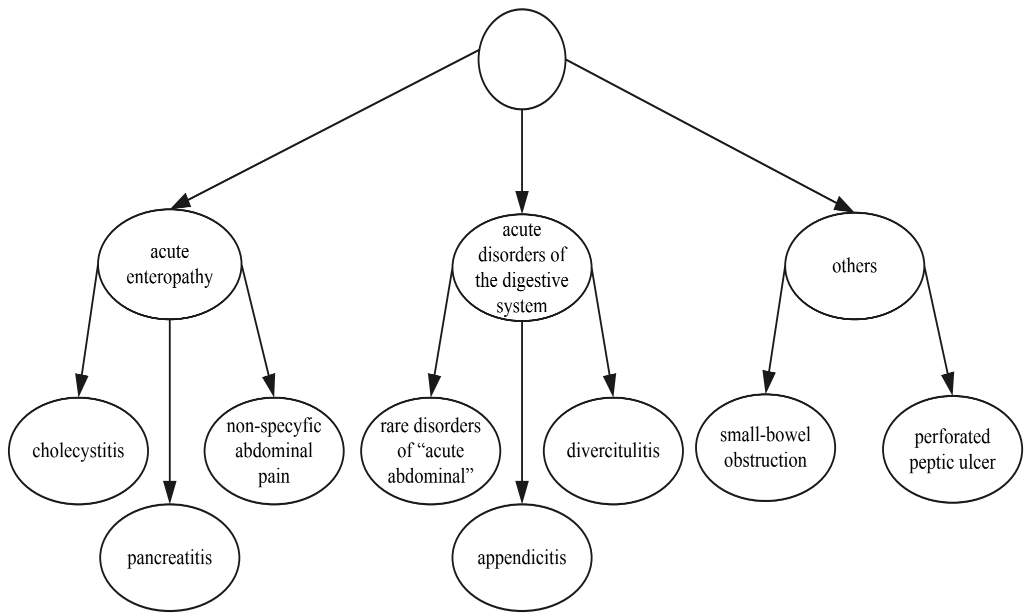


Figure 4. Heuristic classifier for the APP diagnosis problem

6. The LMT method, which returns the decision tree with logistic regression function at the leaves [28,35]
7. FT algorithm, which could use the logistic regression function in the inner nodes [28,36]
8. NBT method, which returns the tree with naive Bayes classifiers in the leaves [37]

The set-up of experiment was as follows:

1. Dataset was collected at the Clinic of Surgery, Wroclaw Medical Academy and consists of 476 elements.
2. All experiments were carried out in WEKA environment [38] and own software created in Matlab environment and PRTools toolbox [39].
3. Errors of the classifiers were estimated using the ten fold cross validation method [40].

The results of experiments are presented in Table 1. Apart from classifier error, time generation (only for generated trees) and tree size are shown.

4. Computer experiment evaluation

The following conclusions could be drawn from the experiment:

1. The quality of multistage classifiers based on Bayes decision theory are similar. The global strategy gave a slightly better result (ca 2%) than the local one. However, we must note that the computational complexity of the global strategy is higher than the local one. For practical implementation, we need to decide if we are willing to bear higher computational costs.
2. We must notice that the qualities of the best multistage classifier based on Bayes decision theory and boosted C4.5 classifier are similar. This observation leads to the conclusion that for this medical case, we could give the expert knowledge about the shape of a decision tree.
3. The multivariate decision trees (LMT, FB, NBTree) achieved much better results than the classifiers based on the univariate decision trees. Additionally, the aforementioned trees outperformed the quality of heuristic trees.
4. We could expect that the multivariate decision trees are not susceptible to overtraining because they usually use smaller trees than the univariate trees.
5. The main disadvantage of the multivariate decision tree training is the time-consuming (especially in comparison with the univariate tree training time), but we can use a smaller number of attributes (lower cost of diagnosis) for making decisions on the similar level. Generating cheap diagnosis-aiding computer tools is actually the problem of research into the so-called cost-sensitive methods [41].
6. Experts revised the structures of classifiers produced automatically and confirmed that most of rules were correct and that the heuristic tree is possibly oversimplified.
7. A similar decision support system is described in [7]; this system uses a different model of acute abdominal pain diagnosis, therefore the results of our tests are not comparable with those in that study. It is worth mentioning that the authors used the C4.5 method and that the frequency of the obtained classifier was ca 57%.

Class number	Locally optimal strategy K=5	Globally optimal strategy K=5	Globally optimal strategy K=7	C4.5	BFTree	LADTree	LMT	FT	NBTree
1	91%	95%	95%	79,43%	88,70%	80,90%	96,50%	95,00%	97,90%
2	52%	60%	62%	52,94%	52,90%	41,20%	64,70%	58,80%	82,40%
3	100%	100%	100%	89,66%	96,60%	89,70%	100,00%	100,00%	96,60%
4	86%	86%	86%	75,00%	67,90%	71,40%	100,00%	100,00%	100,00%
5	96%	96%	96%	89,09%	83,60%	94,50%	100,00%	96,40%	100,00%
6	81%	82%	85%	84,38%	65,60%	62,50%	93,80%	93,80%	96,90%
7	99%	99%	99%	91,08%	95,50%	96,20%	100,00%	99,40%	99,40%
8	94%	94%	94%	70,59%	58,80%	82,40%	100,00%	1,00%	100,00%
Average	92,72%	94,02%	94,82%	83,82%	85,71%	84,90%	97,30%	96,01%	**98,11%**
tree generating [s]				0,02	2,07	2,38	85,97	4,24	30,36
Size [nodes]	4	4	4	102	51	31	1	1	41

Table 1. Frequencies of correct classifications, tree generating time and size of trees.

5. Final remarks

The recognition methods based on of the compound, hierarchical Bayesian approach, and inductive learning are presented. The classifiers generated by those algorithms were applied to the medical decision problem (recognition of Acute Abdominal Pain) and can be used to assist clinicians to in diagnosis. Our empirical results for the inductive learning classifiers and heuristic ones demonstrate the effectiveness of the proposed concepts in such computer-aided medical diagnosis problems. The advantages of the proposed methods make them attractive for a wide range of applications in medicine, which might significantly improve the quality of the care that the clinicians can give to their patients.

Acknowledgments

This work is supported by The Polish National Science Center under the grant which is realizing in years 2011-2014.

Reference

[1] Liebowitz J. [ed], The Handbook of Applied Expert Systems, CRC Press, 1998

[2] Shortliffe E., MYCIN: Computer-based Medical Consultations, New York: American Elsivier, 1975

[3] Sim I., et al., Clinical Decision Support Systems for the Practice of Evidence-based Medicine, Journal of the American Medical Informatics Association, 2001, 8(6), 527–534

[4] Kaplan B., Evaluating informatics applications - clinical decision support systems literature review, International Journal of Medical Informatics, 2001, 64, 15-37

[5] Mextaxiotis K., Samouilidis J.E., Expert systems in medicine: academic illusion or real power? Information Management & Security, 2000, 75-79

[6] Wozniak M., Two-Stage Classifier for Diagnosis of Hypertension Type, LNCS, 2006, 4345, 433-440

[7] Eich H.P., Ohmann C., Lang K., Decision support in acute abdominal pain using an expert system for different knowledge bases, Proceedings of the 10th IEEE Symposium on Computer-Based Medical Systems, 1997, 2-7

[8] Karapandzic V.M., Matic M.D., Pesko P.M., Rankovic V.I. Milicic B.R., Risk assessment in coronary patients undergoing abdominal nonvascular surgery, Central European Journal of Medicine, 2009, 4(4), 459-466

[9] Polat K., Güneşa S., The effect to diagnostic accuracy of decision tree classifier of fuzzy and k-NN based weighted pre-processing methods to diagnosis of erythemato-squamous diseases, Digital Signal Processing, 2006, 16(6), 922-930

[10] Tang T.I., Zheng G., Huang Y., Shu G., A Comparative Study of Medical Data Classification Methods Based on Decision Tree and System Reconstruction Analysis, Industrial Engineering and Management Systems, 2005, 4(1), 102-108

[11] Übeyli E.D., Comparison of different classification algorithms in clinical decision-making Expert Systems, 2007, 24(1), 17-31

[12] Zhang Q., Wang J., Wang J.,Biweekly CHOP therapy improves therapeutic effect in the non-GCB subtype of diffuse large B-cell lymphoma, Central European Journal of Medicine, 2007, 2(4), 488-498

[13] Nikolov M., Simeonova P., Simeonov V., Chemometrics as an option to assess clinical data from diabetes mellitus type 2 patients, Central European Journal of Medicine, 2009, 4(4), 433-443

[14] Papaioannou A., Karamanis G., Rigas I., Spanos T., Roupa Z.,Determination and modelling of clinical laboratory data of healthy individuals and patients with end-stage renal failure, Central European Journal of Medicine, 2009, 4(1), 37-48

[15] Safavian, S.R., Landgrebe, D., A survey of decision tree classifier methodology. IEEE Trans. Systems, Man Cyber., 1991, 21 (3), 660-674

[16] Mui J., Fu K.S., Automated classification of nucleated blood cells using a binary tree classifier, IEEE Trans. Pattern Anal. Mach. Intell., 1980, PAMI-2, 429-443

[17] Duda R.O., Hart P.E., Stork D.G., Pattern Classification, John Wiley and Sons, 2000

[18] Devijver P. A., Kittler J., Pattern Recognition: A Statistical Approach, Prentice Hall, London, 1982

[19] Burduk R., Kurzyński M., Two-stage binary classifier with fuzzy-valued loss function, Pattern Analysis and Applications, 2006, 9(4), 353-358

[20] Alpaydin E., Introduction to Machine Learning. Second edition, The MIT Press, Cambridge, MA, USA, London, UK, 2010

[21] Mitchell T.M., Machine Learning, McGraw-Hill Comp., Inc, New York, 1997

[22] Quinlan J.R., Induction on Decision Tree, Machine Learning, 1986, 1, 81-106

[23] Quinlan J.R., C4.5: Program for Machine Learning, Morgan Kaufman, San Mateo, CA, 1993

[24] Breiman L., Friedman J.H., Olshen R.A., Stone C.J., Classification and Decision trees, Belmont, CA, Wadsworth, 1984

[25] Cover T.M., The Best Two Independent Measurements are Not the Two Best, IEEE Transactions on Systems, Man and Cybernetics, 1974, SMC-4(1), 116-117

[26] Brodley C.E., Utgoff P.E., Multivariate Decision Trees, Machine Learning, 1995, 19(1), 45–77

[27] Kurzynski M., The optimal strategy of a tree classifier, Pattern Recognition, 1983, 16(1), 81-87

[28] Landwehr N., Hall M., Frank E., Logistic model trees, LNCS, 2003, 2837, 241-252

[29] Blum A.L., Langley P., Selection of Relevant Features and Examples in Machine Learning, Artificial Intelligence, 1997, 97(1–2), 245–271

[30] Dash M., Liu H., Feature Selection for Classification, Intelligent Data Analysis, 1997, 1(1–4), 131–156

[31] Kurzynski M., Diagnosis of acute abdominal pain using three-stage classifier, Computers in Biology and Medicine, 1987, 17(1), 19-27

[32] Townsend C.M., Beauchamp R.D., Evers B.M., Mattox K.L., Sabiston Textbook of Surgery, 17th ed. St. Louis, Mo: WB Saunders, 2004

[33] Shi H., Best-first decision tree learning. Hamilton, NZ, 2007

[34] Holmes G., Pfahringer B., Kirkby R., Frank E., Hall E., Multiclass alternating decision trees. Proceedings of ECML, 2001, 161-172

[35] Sumner M., Frank E., Hall M., Speeding up Logistic Model Tree Induction, Proc. of the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases, 2005, 675-683.

[36] Gama J., Functional Trees, Machine Learning, 2004, 55(3), 219-250

[37] Kohavi R., A study of cross-validation and bootstrap for accuracy estimation and model selection, Proceedings of the 14-th International Joint Conference on Artificial Intelligence, San Mateo, 1995, 1137–1143

[38] Witten I.H., Frank E., Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann Pub, 2000

[39] van der Heijden F., Duin R.P.W., de Ridder D., Tax D.M.J, Classification, parameter estimation and state estimation - an engineering approach using Matlab, John Wiley and Sons, 2004

[40] Kohavi R., Scaling Up the Accuracy of Naive-Bayes Classifiers: A Decision-Tree Hybrid, Proc of the Second International Conference on Knoledge Discovery and Data Mining, 1996, 202-207

[41] Turney P.D., Cost-sensitive classification: Empirical evaluation of a hybrid genetic decision tree induction algorithm, Journal Artificial Intelligence Research, 1995, 2, 369–409

Appendix

Pseudocode of ID3 algorithm

```
function ID3(examples, target_concept, attributes)

examples – learning set

target_concept

attributes – list of available attributes

Create a Root node for tree

IF all examples belongs to the same class

      THEN return the single node tree Root with this class's label and
return.

IF set of attributes is empty

      THEN return the single node tree Root with label = most common value
      of label in the set of examples and return

Choose "the best" attribute A from the set of attributes.

FOR EACH possible value vi of attribute

      1. Add new tree branch bellow Root, corresponding to the test A=vi.

      2. Let Evi be the subset of set of examples that has value vi for A.

      3. IF Evi is empty

            THEN below these new branches add a leaf node with label = most
            common value of label in the set of examples

            ELSE below this new branch add a new subtree and call ID3(Evi,
            target_concept, attributes – {A}).

END

RETURN Root
```

Clinical feature description

no	attribute	possible values
1	sex	1 – men, 2 women
2	age	1 – less than 20 years 2 – 20-30 years 3 – 31-40 years 4 – 41-50 years 5 – more than 50 years
3	pain location on the beginning	Location encoding according to the map presented in Figure 3 0 – absence of pain 1 – areas no 1 or 3 or 4 2 – areas no 2 or 5 or 6 3 – areas no from 1 to 10 4 – areas no 1 or 3 or 4 or 7 or 8 or 9 or 11 or 12 or 15 or 16 or 19 5 – areas no 2 or 5 or 6 or 9 or 10 or 13 or 14 or 17 or 18 or 20 6 – areas no 8 or 9 or 12 or 13 7 – each location 8 – areas no 15 or 16 or 19 9 – areas no 17 or 18 or 20 10 – areas no from 11 to 20
4	pain location on present	The same as in the point 3
5	pain intensity	1 – absence or mild 2 – moderate 3 – strong
6	aggravating factors	1 – no factors 2 – breathing 3 – cough 4 – body movement
7	relieving factors	1 – no factors 2 –vomiting 3 – position of body
8	pain progression	1 – outgoing 2 – stable 3 – intensifying
9	pain duration	1 – less than 12 hours 2 – 12-24 hours 3 – 24-48 hours 4 – more than 48 hours
10	pain type on the beginning	1 – broken 2 – stable 3 – colic
11	pain type at present	The same as in the point 3
12	nausea and vomiting	1 – absent 2 – nausea without vomiting 3 – nausea with vomiting
13	appetite	1 – decreased 2 – normal 3 – increased
14	bowel movement	1 – diarrhea 2 – correct 3 – constipation
15	urinate	1 – normal 2 – dysuria
16	previous indigestion	1 – no 2 – yes
17	jaundice	1 – no 2 – yes
18	previous surgery (abdominal)	1 – no 2 - yes
19	drugs	1 – no 2 – yes
20	mood	1 – stimulated/suffered 2 – normal 3 – apathetic/sleepy

21	skin's color	1 – pale 2 – normal 3 – red skin (face)
22	temperature	1 – less than 36,5oC 2 – 36,6 oC-37,0 oC 3 – 37,1 oC-37,5 4 – 37,6 oC-38,0 oC 5 – 38,1 oC-39,0 oC 6 – more than 39,0 oC
23	pulse	1 – less than 60 bits per minute 2 – 61-70 bits per minute 3 – 71-80 bits per minute 4 – 81-90 bits per minute 5 – 91-100 bits per minute 6 – 101-110 bits per minute 7 – 111-120 bits per minute 8 – more than 120 bits per minute
24	respiratory movements of abdomen	1 – normal 2 – absent
25	flatulence	1 – no 2 – yes
26	tenderness (location)	The same as in the point 3
27	Blumberg's sign	1 – negative 2 – positive
28	muscle's defence	1 – no 2 – yes
29	increased tension of abdominal	1 – no 2 – yes
30	swellings	1 – no 2 – yes
31	Murphy's sign	1 – negative 2 – positive