

Dariusz JANKOWSKI,<sup>1</sup> Konrad JACKOWSKI<sup>1</sup>

## **AN IMPROVED MEDICAL DIAGNOSING OF ACUTE ABDOMINAL PAIN WITH DECISION TREE**

In medical decision making (e.g., classification) we expect that decision will be made effectively and reliably. Decision making systems with their ability to learn automatically seem to be very appropriate for performing such tasks. Decision trees provide high classification accuracy with simple representation of gathered knowledge. Those advantages cause that decision trees have been widely used in different areas of medical decision making. In this paper we present characteristic of univariate and multivariate decision tree. We apply those classifiers to the problem of acute abdominal pain diagnosis.

### **1. INTRODUCTION**

Decision is usually made as a combination of knowledge on the problem, experiences from solving similar cases (result of recent researches) and personal judgment. We expected that the decision will be made effectively and reliably. Therefore, it is very important to “support” the decision providing an explanation of how and why the suggested decision is chosen.

One of the significant techniques supporting us are decision support systems (DSS). A properly designed DSS is an interactive software-based system intended to help decision makers compile useful information from a combination of raw data. One can specify three fundamental components of DSS [18]: the database (or knowledge base), the model (i.e., the algorithm of decision making), and the user interface. Generally, the benefits of decision support system are: high quality decision, cost savings, and reduction of the labour involved in decision-making process. Centralizing data simplifies their processing and allows a decision maker to explore a problem more thoroughly. More about DSS benefits one can find in [5, 7].

The use of machine learning methods for the building of predictive and descriptive models has become widely accepted in medical decision support systems [1, 11, 16, 17]. One of the most common machine learning technique use in DSS is classification. The input data for classification task is a collection of records. Each record, also known as an instance or example, has several attributes which describe selected properties of the instance. There is one distinguished attribute called the class label. In general, classifier generates a concise meaningful description for each class in terms of the attributes.

---

<sup>1</sup> Wroclaw University of Technology, Faculty of Electronics

The model is then used to predict class labels of unknown objects - as shown in Figure 1

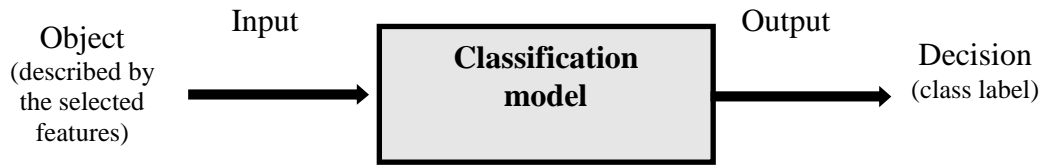


Fig.1 Classification as the task of mapping an input feature set into its class label.

In this paper we present two medical decision support systems based on two optional models of decision tree which can be used to learn the relationship between symptoms and diseases in the problem of the acute abdominal pain diagnosis

The remainder of this paper is organized as follows: In Section 2 the review of decision tree algorithm is presented. Section 3 describes the acute abdominal pain problem. In Section 4 the results of investigations are presented and discussed. Finally, Section 5 contains final remarks.

## 2. DECISION TREE

Various models including decision trees, decision rules, logistic regression, artificial neural networks and SVMs have been tested in the analysis of medical data [8, 9]. However, a rule-based classifiers, such as decision trees, have some advantages i.e., they provide human readable rules of classification which are easy to interpret, construction of decision tree is fast, and they exhibit high accuracy of classification.

The tree consists of elements called nodes and branches. There are three types of nodes: a root node, internal nodes and terminal nodes (leaf). The root and the internal nodes perform tests on the attributes. The branches outgoing from the nodes represent outcome of the test. Each leaf node holds a class label i.e. the final decision.

The decision tree classifier has two phases:

- Growth or build phase
- Pruning phase

In the first phase a decision tree is built by selecting the best test attribute as the root of the decision tree. The procedure exploits instances available in the database (the learning set). Based on the test results, entire learning set is split into two subsets. Then, the same procedure is operated recursively on each branch to induce the remaining levels of the decision tree until all instances in a leaf belong to the same class.

There are many measures that can be used to determine the best way to generate the test in the node and split the records. These measure are defined in terms of the class distribution of the records before and after the split. Developed measures often based on the degree of impurity of the child nodes. The smaller degree of impurity, the more skewed the class distribution. For example, a node with class distribution (0,1) has zero impurity, whereas a node with uniform class distribution (0.5, 0.5) has the highest impurity. The most common impurity measure are: Gini Index, Information Gain, and Gain Ratio [2].

Decision-tree algorithms can create over-complex trees that do not generalize the data well. This phenomenon is called overfitting [12]. Mechanisms such as pruning are necessary to avoid this problem. The pruning phase may be done only on the fully grown tree. The tree is pruned by cutting off their “insignificant” nodes or even subtrees (sections of a nodes that may be based on noisy or erroneous data).

The frequently used decision tree induction algorithms use univariate partitioning methods, which are attractive because they are straightforward to implement (since only one feature is analysed at each node) and the derives decision trees that are relatively easy to understand. Beside univariate partitioning methods there are also some successful partitioning methods which do not partition the search space axis-

parallel based on only one attribute at a time but form oblique partition boundaries based on a combination of attributes, which produces remarkably small trees.

On Figure 2 we have the example of data set that cannot be partitioned optimally using test condition involving single attributes.

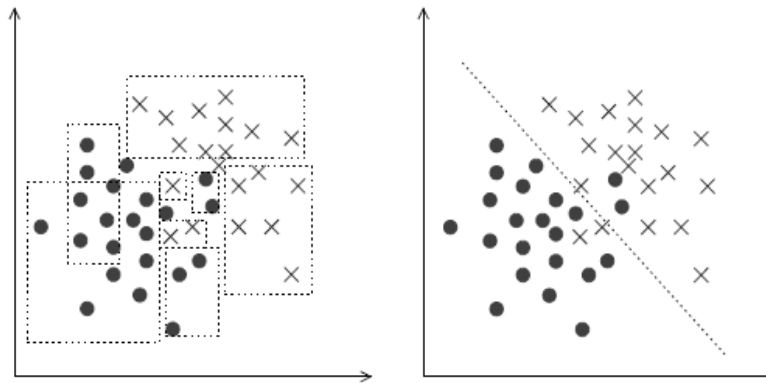


Fig.2 Example of data set with (i) axis-parallel and (ii) oblique decision boundaries.

Most widely used methods for decision tree induction are: CART [3] for univariate decision trees and OC1 [13] for multivariate decision trees. The experiments are conducted on the above two algorithms.

### 2.1. CLASSIFICATION AND REGRESSION TREES (CART)

CART [3] algorithm learns decision trees by binary recursive partitioning of objects in the nodes. The term binary implies that the group of patients examined in the node, can be split into one of two subsequent groups. Thus, each node (called parent node) has two outgoing branches and is connected with two child nodes. To find the best possible criterion of the division, the algorithm checks all attributes (which described the patient state) and create simple test formula based on selected attribute. The formula has a form of decision rule which examines the value of the attribute (e.g., IF attribute is grater then  $\alpha$  THEN the patient belongs to group A; ELSE the patient belongs to group B). The term recursive refers to the fact that the entire process is then repeated over and over again for each child node until one of the stopping conditions is met: (1) all object in the child node belong to the same class; (2) the number of observations in the node is less than predefined threshold. Terminal nodes are called leafs and they represent class labels.

### 2.2. OBLIQUE CLASSIFIER 1 (OC1)

Oblique Classifier 1 [13] is algorithm for generating multivariate decision trees. The main difference between OC1 and the CART is that the first one uses linear combinations of the features at each non-leaf node for testing (which divides the attribute space with hyperplanes). The initial hyperplane at each node in decision tree is chosen randomly. Even if such a randomly placed hyperplane has very poor location, it is usually improved greatly in the first few perturbations. An important problem in searching for the best hyperplane is that algorithm may fall into local minima. Authors have implemented two methods of dealing with this problem: perturbing an hyperplane in a random direction, and re-running the perturbation algorithm with additional initial hyperplanes.

## 3. ACUTE ABDOMINAL PAIN DIAGNOSIS

Abdominal pain (or stomach ache) is a common symptom associated with transient disorders or serious disease. Diagnosing the cause of abdominal pain can be difficult, because many diseases can

cause this symptom. The pain may frequently be associated with nausea and vomiting, abdominal distention, fever and signs of shock. Acute abdominal pain (AAP) can be defined as severe, persistent abdominal pain of sudden onset that is likely to require surgical intervention to treat its cause.

The goal of the evaluation of the patient with acute abdominal pain is early, efficient, and accurate diagnosis. The most important elements in making accurate early diagnosis are the history and physical examination. A careful description of the chronology, location, intensity, and character of the pain as well as aggravating and alleviating factors, also other symptoms usually allows an accurate diagnosis to be made. A thorough physical examination can confirm a diagnosis that was suspected during the history. In some cases, the diagnosis is obscure despite an exhaustive evaluation. In most settings in which the patient's clinical status is stable, repetitive examination over time eliminates diagnostic uncertainty. In such situation, admission to the hospital for serial abdominal examinations is necessary.

The goals of treatment of patients with abdominal pain are identification and cure of the responsible underlying disease.

The application of machine learning methods to the problem of AAP diagnosis has been the subject of quite few studies. In [14] authors applied Bayesian methods on a clinical data set of AAP patients. Neural networks algorithm was used in [15] by Pesonen et al. Different large databases with cases of AAP and decision trees were investigated in [19].

## 4. EXPERIMENTAL ANALYSIS AND RESULTS

The aim of the presented study is presentation and comparison of two Clinical Decision Support System based on univariate and multivariate decision tree for the acute abdominal pain diagnosis.

We used CART and OC1 algorithm and tested them on AAP medical data set. The experiments were carried out in C++ and own software created in Matlab environment. To ensure the highest reliability of the results, all experiments were carried out using ten-fold cross-validation t-paired test [6].

In ten-fold cross-validation, the learning data set is randomly partitioned into ten subsets. One subset is retained as the test set for testing the model, and the remaining nine subsets are used as training data. The whole process is then repeated ten times (the folds), with each of the ten subsets used exactly once as the validation set. The advantage of this method over repeated random choosing objects is that all observations are used for both training and validation but never at the same time. This assures that a similar distribution of outcomes is present in each of the ten subsets of data. Additionally, statistical t-paired test allows reliable comparison of results obtained by two competing methods. It checks if the difference in their performance is statistically significant.

### 4.1. EXPERIMENTAL DATA

We have used real medical data set collected from patients the Clinic of Surgery, Wroclaw Medical Academy. It was originally used in [4, 10], where a more detailed description of the data can be found, in particular list of all features and possible values. There are eight diagnostic groups (the classes):

- |  |                                  |
|--|----------------------------------|
| 1) Cholecystitis (CHO)                       | 5) Appendicitis (APP)            |
| 2) Pancreatitis (PAN)                        | 6) Diverticulitis (DIV)          |
| 3) Non-specific Abdominal Pain (NAP)         | 7) Small-bowel obstruction (SBO) |
| 4) Rare Disorders of "Acute Abdominal" (RDA) | 8) Perforated Peptic Ulcer (PPU) |

NAP is a group in which were placed all patients that did not belong to one of the other groups. In this meaning this is not a diagnostic group. The data consist of 476 records and 31 attributes each.

Table 1 presents the number of examples that belong to each diagnostic group training set, and testing set. The training set is used to train the models. The test used to assess the model, does not overlapped with the training data and.

Table 1. Data distribution.

	CHO	PAN	NAP	RDA	APP	DIV	SBO	PPU	Total
<i>Training Set</i>	113	13	24	23	44	26	126	13	382
<i>Test Set</i>	28	4	5	5	11	6	31	4	94
<i>Total</i>	141	17	29	28	55	32	157	17	476

#### 4.2. ACCURACY COMPARISON

The algorithms growing the tree try to perfectly classify the training examples, which not always give good results. Both univariate and multivariate algorithms can produce trees that overfit the data set. This occurs due to two different causes. Firstly the data set may contain noise. If we learn all examples we will also learn the noise, which reduces our performance over the test set. Secondly our training set may not be a good representative (big enough) of the data set. Taking into account the above we decide analyse both pruned and unpruned version received trees and compare their accuracy as the classification quality measure.

In carried experiments we have used two different splitting criteria: Gini Index and Information Gain for OC1. Gini Index it is the default criteria used for classification trees.

The following tables show accuracy evaluation on our AAP data set. Table 2 presents the mean accuracy, and the variance of ten-fold cross validation t-paired test. (one sided test with probability of incorrect rejecting the null hypothesis of 0.05).

Table 2. Classifier accuracy.

<b>Classifier</b>	<b>Mean Accuracy [%]</b>	<b>Variance</b>	<b>Statistically better than</b>
CART-Gini* -Up**	84.71	26.72	OC1-Gini-Up, OC1-Gini-Pr, OC1-Info-Up, OC1-Info-Pr,
OC1-Gini-Up	75.47	8.79	OC1-Gini-Pr
OC1-Info-Up	78.59	56.51	OC1-Gini-Pr
CART-Gini-Pr	83.41	42.74	OC1-Gini-Up, OC1-Gini-Pr, OC1-Info-Pr
OC1-Gini-Pr	70.72	45.15	--
OC1-Info-Pr	78.35	37.08	OC1-Gini-Pr

\* Split criterion: Gini - Gini Index, Info - Information Gain

\*\* Version of decision tree: Pr - Pruned; Up - Unpruned

The following observation can be made:

1. In almost all cases CART algorithms statistically outperform all version of OC1 algorithms.
2. There is no statistical difference between pruned and unpruned version of CART algorithm.
3. Contrary to our expectation, pruned versions of OC1 algorithm do not outperform respective unpruned algorithm.
4. There is no statistical difference between accuracy for OC algorithms which exploit Gini Index and Information Gain as the splitting criterion.

Conclusion that can be drawn based on observed facts is that more sophisticated OC1 algorithm, which can adjust very effectively to samples gathered in learning set, does not generalize properly the problem. Its performance measured on testing set is significantly lower comparing to CART algorithms. Surprising weak performance of pruned trees might result from skewed class distribution problem, i.e. cutting off branches related with poorly represented classes (which may be interpreted as noise or

outliers) might disturb decision system and elevate misclassification level. The last conclusion is that for given diagnosis problem there is no significant difference between Gini Index and Information Gain. Both of he can be used according to user preference with the same result.

4.3. NODESIZE AND LEARNING TIME COMPARISON

The second experiment had two objectives: (1) comparison of computational effort required for creating the DSS based on CART and OC1 with pruning procedure active and inactive; (2) comparison of complexity of generated trees. Table 3 presents mean values of ten-fold cross-validation for: the processing time (in seconds), tree depth, and number of leaves needed to build a decision tree.

Table 3. Tree size and execution time.

Classifier	Unpruned tree			Pruned tree		
	Max. tree depth	Leaves number	Time (sec)	Max. tree depth	Leaves number	Time (sec)
CART-Gini*	9	33	0.7	7	17	0.7
OC1-Gini	12	36	65	8	16	67
OC1-Info	8	33	71	6	13	72

\* Split criterion: Gini - Gini Index, Info - Information Gain

Table 3 clearly reveals that:

1. As it was expected, there is a difference in sizes between pruned and unpruned trees in all cases. The first ones have always significantly smaller tree depth and number of leafs what is the natural result of the main purpose of pruning procedure.
2. CART algorithms always created larger trees comparing to OC1. This is the consequence of simplified node test formula which cannot effectively split the group of instances featuring more complex distribution (see Figure 2).
3. OC1 methods needs approximately ten times more time to build the tree than CART. This is due to high complexity of attribute subset selection for multivariate node test formula in OC1.

5. CONCLUSIONS

In the paper we presented Decision Support System dedicated to supporting process of acute abdominal pain diagnosis. The core of the system exploits decision trees algorithm. In order to select the best model, two version of the algorithm were implemented and tested with different splitting criterion. In addition pruning procedure which is recommended to use in order to prevent overfitting was also tested. In this study we used as well the statistical paired t-test method for reliable validation of the algorithms quality comparison.

Results of experimental analysis performed on empirical data collected in cooperation with Wroclaw Medical Academy, shows that the best performance in term of accuracy of diagnosis was achieved by CART algorithm without pruning. Apparently more sophisticated OC1 algorithm was not able fit to the data set effectively and in consequence its accuracy was not acceptable, although, as it had been expected, the complexity of the decision tree generated by OC1 was relatively lower comparing to CART trees. The overall accuracy of the recommended CART algorithm without pruning of nearly 85% of correct classification makes the system interesting proposition for implementation in practice. Apart from aforementioned drawbacks of OC1, the last factor that makes it less useable in practice was the long time required for training, which is roughly 10 times longer than time required by CART.

Before using our DSS in real clinical situations, those system must be tested extensively, in order to ensure that he is conforming the clinical guidelines on which based.

In the future, we aim to develop a better model that provides easy to understand rules with optimal accuracy.

**BIBLIOGRAPHY**

- [1] BELLAZZI R., ZUPAN B., Predictive data mining in clinical medicine: current issues and guidelines, *International Journal of Medical Informatics*, 2008, 77, 2, pp. 81–97.
- [2] BRAMER M., *Principles of Data Mining*, 2007, Springer.
- [3] BREIMAN L., et al., *Classification and Regression Trees*, Wadsworth, 1984.
- [4] BURDUK R., WOZNIAK M., Different decision tree induction strategies for a medical decision problem, *Central European Journal of Medicine*, 2012, 7, 2, pp. 183–193.
- [5] CHI L., et al., *Organizational Decision Support Systems: Parameters and Benefits. Handbook of Decision Support Systems*, 2007.
- [6] DIETTERICH T.G., Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms, *Neural Computation*, 1998, 10, 7, pp. 1895–1923.
- [7] HOLSAPPLE C., ERP plans and decision-support benefits. *Decision Support Systems*, 2005, 38, 4, pp. 575–590.
- [8] IGOR K., et al., Application of Machine Learning to Medical Diagnosis, *Journal of Orthopaedic Trauma*, 2011, 25 Suppl 3, 99–100.
- [9] KONONENKO I., Machine learning for medical diagnosis: history, state of the art and perspective, *Neural Networks*, 2001, 23, 1, pp. 1–25.
- [10] KURZYŃSKI M.W., Diagnosis of acute abdominal pain using a three-stage classifier, *Computers in Biology and Medicine*, 1987, 17, 1, pp. 19–27.
- [11] MATHENY M.E., OHNO-MACHADO L., Generation of knowledge for clinical decision support: Statistical and machine learning techniques, *Clinical Decision Support*, R.A. Greenes et al., eds. Elsevier, Academic Press, 2007, pp. 227–248.
- [12] MITCHELL T.M., *Machine Learning*, McGraw-Hill, 1997.
- [13] MURTHY S.K., et al., OC1: A Randomized Induction of Oblique Decision Trees, *National Conference on Artificial Intelligence*, 1993, pp. 322–327.
- [14] OHMANN C., et al., Evaluation of automatic knowledge acquisition techniques in the diagnosis of acute abdominal pain, *Acute Abdominal Pain Study Group, Artificial Intelligence in Medicine*, 1996, 8, 1, pp. 23–36.
- [15] PESONEN E., et al., Comparison of different neural network algorithms in the diagnosis of acute appendicitis, *International journal of biomedical computing*, 1996, 40, 3, pp. 227–233.
- [16] SHORTLIFFE E.H., *Computer-Based Medical Consultations: MYCIN*, Elsevier, 1976.
- [17] SOTOS J.G., MYCIN and NEOMYCIN: two approaches to generating explanations in rule-based expert systems, *Aviation space and environmental medicine*, 1990, 61, 10, pp. 950–954.
- [18] SPRAGUE R.H., CARLSON E.D., *Building Effective Decision Support Systems*, Prentice Hall, 1982.
- [19] ZORMAN M., et al., Comparison of three databases with a decision tree approach in the medical field of acute appendicitis, *Studies In Health Technology And Informatics*, 2001, 84, Pt 2, pp. 1414–1418.

