*Journal of Medical Systems, Vol. 26, No. 5, October 2002 (© 2002)*

# Does Size Really Matter—Using a Decision Tree Approach for Comparison of Three Different Databases from the Medical Field of Acute Appendicitis

**Milan Zorman,[1,3] Hans-Peter Eich,[2] Bruno Stiglic,[1] Christian Ohmann,[2] and Mitja Lenic[1]**

*Decision trees have been successfully used for years in many medical decision making applications. Transparent representation of acquired knowledge and fast algorithms made decision trees one of the most often used symbolic machine learning approaches. This paper concentrates on the problem of separating acute appendicitis, which is a special problem of acute abdominal pain, from other diseases that cause acute abdominal pain by use of an decision tree approach. Early and accurate diagnosing of acute appendicitis is still a difficult and challenging problem in everyday clinical routine. An important factor in the error rate is poor discrimination between acute appendicitis and other diseases that cause acute abdominal pain. This error rate is still high, despite considerable improvements in history-taking and clinical examination, computer-aided decision-support, and special investigation such as ultrasound. We investigated three databases of different size with cases of acute abdominal pain to complete this task as successful as possible. The results show that the size of the database does not necessary directly influence the success of the decision tree built on it. Surprisingly we got the best results from the decision trees built on the smallest and the biggest database, where the database with medium size (relative to the other two) was not so successful. Despite this we were able to produce decision tree classifiers that were capable of producing correct decisions on test data sets with accuracy up to 84%, sensitivity to acute appendicitis up to 90%, and specificity up to 80% on the same test set.*

**KEY WORDS:** acute appendicitis; decision trees; acute abdominal pain; machine learning; medical informatics.

[1]Laboratory for System Design, Faculty of Electrical Engineering and Computer Science, University of Maribor, Slovenia.
[2]Theoretical Surgery Unit, Department of General and Trauma Surgery, Heinrich-Heine University, Düsseldorf, Germany.
[3]To whom correspondence should be addressed; e-mail: milan.zorman@uni-mb.si.



0148-5598/02/1000-0465/0 © 2002 Plenum Publishing Corporation

## INTRODUCTION

Decision support systems that help physicians are becoming a very important part of medical decision making. They are based on different models and the best of them are providing an explanation together with an accurate, reliable, and quick response. One of the most popular among machine learning approaches are decision trees. For years they have been successfully used in many medical decision making applications. Transparent representation of acquired knowledge and fast algorithms made decision trees what they are today: one of the most often used symbolic machine learning approaches.[1] Decision trees have been already used successfully in medicine, but as in traditional statistics, some hard real world problems cannot be solved successfully using the traditional way of induction.[2] One of the hardest problems is the diagnostic of acute appendicitis (AAP), which is a special problem of acute abdominal pain. The early and accurate diagnosis of acute appendicitis is still a difficult and challenging problem in everyday clinical routine. Of major concern are the perforation rate (up to 20%) and negative appendectomy rate (up to 30%).[3,4] An important factor in the error rate is poor discrimination between acute appendicitis and other diseases that cause acute abdominal pain. This error rate is still high, despite considerable improvements in history-taking and clinical examination, computer-aided decision-support, and special investigation such as ultrasound.

Different types of automatic knowledge acquisition tools such as decision trees[5] and neural networks[6] have been already evaluated on databases with cases of acute abdominal pain. This clinical problem seems to be well suited for inductive learning systems since a standardized terminology has been defined. Agreed definitions, criteria, and minimum data sets have been laid down by the World Organization of Gastroenterology.[7]

This paper concentrates on the problem of separating acute appendicitis from other diseases that cause acute abdominal pain by use of an improved decision tree approach. In addition three different large databases with cases of acute abdominal pain have been investigated to find out the influence of different database characteristics such as size, prevalence of appendicitis, reliability of data collection, etc. to the accuracy of decision support tools.

## METHODS

### Decision Trees

Inductive inference is a process of moving from concrete examples to general models, where the goal is to learn how to classify (predict) objects by analyzing a set of instances (already solved cases) whose classes (predictions) are known. Instances are typically represented as attribute-value vectors. Learning input consists of a set of such vectors, each belonging to a known class (prediction), and the output consists of a mapping from attribute values to classes (predictions). This mapping should accurately classify/predict both the given instances and other unseen instances.

A decision tree[8,9] is a formalism for expressing mappings from attribute values to classes (predictions) and consists of tests or attribute nodes linked to two or more

subtrees and leafs or decision nodes labeled with a class which represents the decision. Because of the very simple representation of accumulated knowledge they also give us the explanation of the decision, and that is essential in medical applications.

The tool we used is called MtDeciT2.0. It not only follows the same principles as many other decision tree building tools but also implements different extensions.[10,11] One of those extensions is called dynamic discretization of continuous attributes, which was used in our experiments with success.

For the purpose of performing objective tests, we used a basic type of decision tree algorithm, which is implemented in MtDeciT2.0 among other machine learning approaches. We performed tests with all three training sets using different discretization methods: from simple equidistant discretization, to threshold and both dynamic types of discretization. We also varied other parameters such as prepruning criteria and type of heuristic function (entropy, gain, or gain ratio). The values for those parameters that had an important influence to the tree's accuracy, sensitivity, and specificity, are also listed for each best individual decision tree.

### *Dynamic Discretization of Continuous Attributes*

Because of the nature of decision trees, all numeric attributes must be mapped into a set of discrete values. In MtDeciT 2.0 tool we implemented an algorithm for finding subintervals,[11] where we consider the distribution of training objects and there are more than two subintervals possible. The approach is called dynamic discretization of continuous attributes since the subintervals are determined dynamically during the process of building the decision tree. This technique first splits the interval into many subintervals, so that every training object's value has its own subinterval. In the second step it merges together smaller subintervals that are labelled with the same outcome into larger subintervals (see Fig. 1).

Here is a more detailed algorithm for dynamic discretization:

1. Set the values for constants $z$ and percentage of discretization tolerance. Here $z$ represents a preset value, usually greater then 2. Percentage of discretization tolerance can take any value between 0% and 50% and it represents the upper percentage of all training objects, contained in subintervals marked with different outcome then their own.
2. Split the interval of the continuous attribute into many subintervals, so that every training object's value lays its own subinterval. Threshold candidates are all values $t_i$, (Eq. (1)) that lay between $a_i$ in $a_{i+1}$, where $a_i$ in $a_{i+1}$ are two neighbor values of the continuous attribute:

$$t_i = \frac{a_{i+1} - a_i}{2} \tag{1}$$

3. For each subinterval count the number of training objects that belong to each class. It is possible that two training objects which belong to different classes have the same value of the attribute in question. Set the dominant class in each subinterval.
4. Merge together smaller subintervals that are labelled with the same class into larger subintervals.

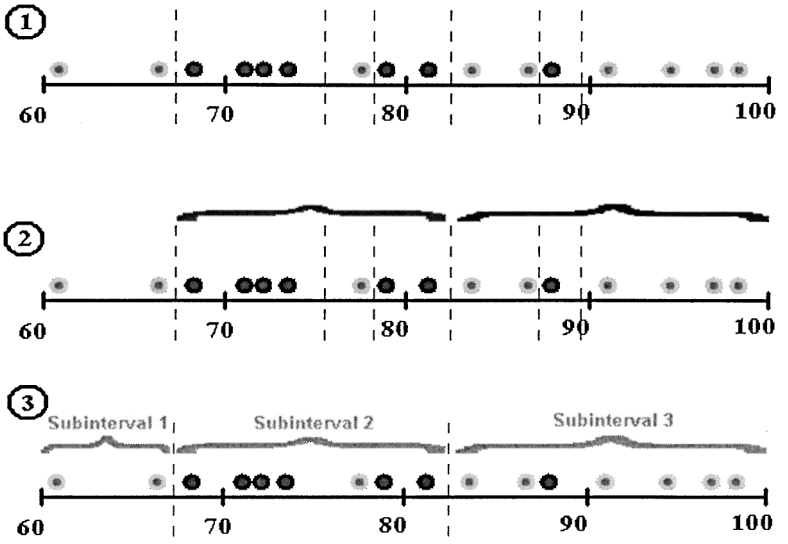


**Fig. 1.** Dynamic discretization of a continuous attribute, which has values between 60 and 100.

5. Determine triplets of neighbor subintervals ($[A_0)[A_1)[A_2)$), where the outer ("strong") two subintervals $A_0$ and $A_2$ are labelled with the same class. Each of the outer intervals $A_0$ and $A_2$ must contain at least as many training objects as the inner ("weak") subinterval $A_1$. $A_0$ and $A_2$ must have together at least $z$ -times as many training objects as $A_1$. Find the triplet where the ratio between the "strong" subintervals and the "weak" subinterval is the greatest. If merging these three subintervals into one subinterval and labelling it with the dominant class does not cause the overflow of percentage of discretization tolerance, then we merge the three subintervals together.
6. Repeat the procedure in point 5 until one of the following conditions is met:
   - There are no more triplets left
   - We exceeded the percentage of discretization tolerance
   - There are less then four subintervals left.

In comparison to other approaches the dynamic discretization returns more "natural" subintervals, which results in better and smaller decision trees.

In general we differentiate between two types of dynamic discretization:

- General dynamic discretization and
- Nodal dynamic discretization.

General dynamic discretization uses all available training objects for the definition of subintervals. That is why we perform the general dynamic discretization before we start building the decision tree. All the subintervals of all attributes are memorized so as to be used later in the process of building of the decision tree. Nodal dynamic discretization performs the definition of subintervals for all continuous attributes that are available in the current node of the decision tree. Only those training objects

**Table I.** Parameters for Training

| Parameters |
|---|
| Sex |
| Age |
| Progress of pain |
| Duration of pain |
| Type of pain |
| Severity of pain |
| Location of pain now |
| Location of pain at onset |
| Previously similar complains |
| Previous abdominal operation on appendix |
| Distended abdomen |
| Tenderness |
| Severity of tenderness |
| Movement of abdominal wall |
| Rigidity |
| Rectal tenderness |
| Rebound tenderness |
| Leukocytes |

that came in the current node are used for setting the subintervals of the continuous attributes.

## Data Collection

To compare the databases we concentrated on 18 parameters from history-taking and clinical examination, which could be identified in all three databases (Table I). Only clinical parameters with a missing value rate of less than 10% were included. Since we were focusing on the problem of separating acute appendicitis (class: "appendicitis") from other diseases that cause acute abdominal pain, these other diagnoses fall into one common class (class: "other diseases"). The clinical parameters in the investigation are sex, age, progress of pain, duration of pain, type of pain, severity of pain, location of pain now, location of pain at onset, previously similar complains, previous abdominal operation on appendix, distended abdomen, tenderness, severity of tenderness, movement of abdominal wall, rigidity, rectal tenderness, rebound tenderness, and leukocytes.

### *Databases of Acute Abdominal Pain*

1. AAP I ($n = 1254$): This prospective clinical database of AAP was built-up in the framework of a Concerted Action of the European Community (COMAC-BME-European Community Concerted Action on Objective Medical Decision Making in Patients with Acute Abdominal Pain; project leader: F. T. de Dombal (Leeds, UK).[12] The data came from six surgical departments in Germany, which participated in the study. Included in the study were all patients with acute abdominal pain of less than 1 week duration. A structured and standardized history and clinical examination were performed in every patient and the data were documented prospectively using a form

suitable for computer use. This form was based on the original abdominal pain chart of the World Organization of Gastroenterology (OMGE). Terminology and definitions were taken from the European Community Concerted Action.[12] Final diagnosis was based on operative findings, special investigations, and the course of the disease during hospital stay. In cases of patients with nonspecific abdominal pain, data from readmission and telephone interviews were used. The prevalence of appendicitis in this database is 16.8% ($n = 211$).

2. AAP II ($n = 2286$): This prospective database was built-up during the German MEDWIS project A70 "Expert System for Acute Abdominal Pain," (project leader: C. Ohmann).[13] Data came from 14 centres in Germany. Included in the study were all patients with acute abdominal pain of less than 1 week duration. For data collection a computer program with revised and enhanced forms of AAP I[12] was used. The final diagnosis was based on diagnosis at discharge. The prevalence of appendicitis in this database is 22.7% ($n = 519$). This data set contained a lot of special (more complicated) cases where patients were sent from ordinary hospitals for treatment in the university hospitals.
3. AAP III ($n = 4020$): This prospective database was built-up during an Concerted Action funded by the European Commission during the COPERNICUS programme no.: 555 (project leader: C. Ohmann): "Information Technology for the Quality Assurance in Acute Abdominal Pain." Data was collected in 16 centres from Central and Eastern Europe. For data collection the computer program developed in the MEDWIS programme was used. Medical terminology was translated into 10 different languages, so that the participating centres could be provided with national versions of the software.[14] The final diagnosis was based on the diagnosis at discharge. The prevalence of appendicitis in this database is 40.5% ($n = 1628$).

In all the three databases we additionally filtered out the cases (objects) for which more than 90% of parameters were not known. As a result of this action, the number of cases in the AAP I reduced for 3 objects (from 1254 to 1251), the number of cases in the AAP II reduced for 7 cases (from 2286 to 2279), and the number of cases in the AAP III remained the same.

### *Training Sets*

For the training purposes we decided not to use training objects with more than 10 missing values. By that we did our best to increase the quality of knowledge stored in the decision trees. Let us call the data sets which contained objects with no more than 10 missing values as "cleaned data sets."

During our preliminary tests we found out that the percentage of appendicitis cases in all the three data sets was substantially lower than 50% and therefore influenced the decision trees in such way that they learned more about the other diseases than about appendicitis. In order to improve the power of classifiers we reduced the number of objects in the sets by removing the objects classified as "other diagnosis" that had the most missing values. Let us call such data sets as "reduced data sets."

For each data set we built two training sets. For the first training set (marked as "Training set 50:50" in Tables II–IV) we took approximately 2/3rd of the cleaned data set. The remaining 1/3rd of the data set was saved for the testing purposes as the test set. Then we reduced the 2/3rd training set so that it contained approximately the same number of appendicitis cases and cases marked as other diagnosis.

The second training set was our reduced data set (marked as "Full set 50:50" in Tables II–IV)—the original data set with an approximate ratio of half of objects classified as appendicitis cases and the other half classified as other diagnoses.

The number of training objects in the "training sets 50:50" was 274 for AAP I (137 classified as appendicitis, 137 classified as other diagnoses), 763 for AAP II (363 classified as appendicitis, 400 classified as other diagnoses), and 2186 for AAP III (1086 classified as appendicitis, 1100 classified as other diagnoses).

The number of training objects in the "full sets 50:50" was 422 for AAP I (211 classified as appendicitis, 211 classified as other diagnoses), 1119 for AAP II (519 classified as appendicitis, 600 classified as other diagnoses), and 3330 for AAP III (1628 classified as appendicitis, 1702 classified as other diagnoses).

### *Test Sets*

Similar as for training sets, we also built two test sets for each data set. The first test set (marked in Tables II–IV as "Test set") was the remaining 1/3rd of the "cleaned data set," which meant that we did not have the approximate 50:50 ratio of the appendicitis and other diagnoses in the test set. The same was true for the second test set, which was actually a "cleaned data set" (marked in Tables II–IV as "Full set").

The number of test objects in "test sets" was 414 for AAP I (74 classified as appendicitis, 340 classified as other diagnoses), 731 for AAP II (156 classified as appendicitis, 575 classified as other diagnoses), and 1340 for AAP III (542 classified as appendicitis, 798 classified as other diagnoses).

The number of test objects in "full sets" was 1251 for AAP I (211 classified as appendicitis, 1040 classified as other diagnoses), 2279 for AAP II (519 classified as appendicitis, 1760 classified as other diagnoses), and 4020 for AAP III (1628 classified as appendicitis, 2329 classified as other diagnoses).

## RESULTS

For each AAP data set we built two types of decision trees: one for each type of training set. We tested each of those decision trees on each possible test set, except on its own full set (for decision trees built on training set) and its own full and testing set (for decision trees built on reduced data set). Reason for latter was that training sets contained also a few objects that were in the test sets and the results would not be objective.

Each part of Tables II–IV at intersection between the training and test set contains cells with the following data: number of nodes in the decision tree, settings for the decision tree (prepruning percentage, type of discretization technique), overall

| Size of the decision tree | Settings for the decision tree |
|---|---|
| Overall accuracy | |
| Sensitivity to Appendicitis | Specificity |

**Fig. 2.** Cell map.

accuracy, sensitivity to appendicitis, and specificity (see cell map in Fig. 2). Types of discretization techniques used in Tables II–IV, are quartiles (Q) and dynamic discretization in the current node with different settings (DC40-2).[11]

In Table II gives the best results of the decision trees built on the AAP I data set and tested on each possible data set, except on the full AAP I set. It is interesting that the best accuracy was not achieved on the AAP I test set, but on the AAP III test and full sets.

In Table III gives the best results of the decision trees built on the AAP II data set. Using the training set we got the best results on the AAP II test set. To our surprise the best results for the decision tree built with full AAP II set were achieved on the AAP I test and full sets. Despite that, no decision trees built on AAP II could be described as useful, since in the majority of case the best accuracy hardly exceeded 50%.

The results of testing decision trees, built on AAP III training and full sets (Table IV), indicate that this data set is capable of providing more knowledge as the AAP II data set (Table III). The highest accuracy of the decision tree built on AAP III training and full sets (Table II) matched the highest accuracy of the one built on the AAP I training and full sets. But if we look at the average accuracy, AAP I (Table II) still gave a bit better results.

From the results of comparison of different data sets we can see that the best average accuracy has been achieved by the decision tree, built on the small reduced data set AAP I (marked as "Full set 50:50" in Table II), followed closely by the decision trees built on the large AAP III data set (Table IV), which is, considering the data set background and sizes, quite a surprise to us.

## DISCUSSION

By knowing the background and methods used to collect those three data sets, we did not expect the decision trees built on medium data set AAP II to present themselves in such a negative sense. The worse results were achieved on AAP II test sets and with the decision trees that were built on the training sets of AAP II data set. The only reason for this which arise at the moment is that AAP II contains a large number of special cases and that the decision tree learning method does not exploit the training set as it should. The overall accuracy of the remaining comparisons between AAP I and AAP III is so high that some of those decision trees could be of a practical use to clinicians.

**Table II.** Results of the Decision Trees Built on the AAP I Data Set

| Small training set (AAP I) | Test Sets | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Small (AAP I) | | Medium (AAP II) | | | Large (AAP III) | | |
| | Test set | | Test set | | Full set | Test set | | Full set |
| Training set 50:50 | 69 Nodes 25%DC40-2 | | 69 Nodes 25%DC40-2 | | 69 Nodes 25%DC40-2 | 69 Nodes 25%DC40-2 | | 69 Nodes 25%DC40-2 |
| | 73.67% | | 56.77% | | 54.76% | 75.52% | | 75.67% |
| | 74.32% | 73.53% | 42.95% | 60.52% | 39.11% 59.38% | 65.87% | 82.08% | 64.99% 82.94% |
| Full set 50:50 | | | 23 Nodes 30%DC40-2 | | 23 Nodes 30%DC40-2 | 23 Nodes 30%DC40-2 | | 23 Nodes 30%DC40-2 |
| | | | 55.81% | | 52.87% | 82.31% | | 81.99% |
| | | | 49.36% | 57.57% | 42.77% 55.85% | 82.29% | 82.33% | 81.88% 82.07% |

**Table III.** Results of the Decision Trees Built on the AAP II Data Set

| Medium training set (AAP II) | Test Sets | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Small (AAP I) | | | | Medium (AAP II) | | Large (AAP III) | | | |
| | Test set | | Full set | | Test set | | Test set | | Full set | |
| Training set 50:50 | 464 Nodes 30%Q | | 464 Nodes 30%Q | | 47 Nodes 40%DC40-2 | | 464 Nodes 30%Q | | 464 Nodes 30%Q | |
| | 42.75% | | 43.73% | | 52.39% | | 46.79% | | 43.73% | |
| | 43.24% | 42.65% | 38.86% | 44.71% | 51.28% | 52.70% | 39.30% | 51.88% | 36.12% | 48.91% |
| Full set 50:50 | 162 Nodes 40%DC40-2 | | 162 Nodes 40%DC40-2 | | | | 162 Nodes 40%DC40-2 | | 162 Nodes 40%DC40-2 | |
| | 68.84% | | 69.30% | | | | 56.79% | | 55.47% | |
| | 54.05% | 72.06% | 44.08% | 74.42% | | | 32.84% | 73.06% | 31.27% | 71.95% |

**Table IV.** Results of the Decision Trees Built on the AAP III Data Set

| Large training set (AAP III) | Test Sets | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Small (AAP I) | | | | Medium (AAP II) | | | | Large (AAP III) | |
| | Test set | | Full set | | Test set | | Full set | | Test set | |
| Training set 50:50 | 48 Nodes | 40%DC40-2 | 48 Nodes | 40%DC40-2 | 48 Nodes | 40%DC40-2 | 48 Nodes | 40%DC40-2 | 48 Nodes | 40%DC40-2 |
| | 64.73% | | 63.39% | | 53.21% | | 50.59% | | 83.81% | |
| | 86.49% | 60.0% | 88.63% | 58.27% | 52.56% | 53.39% | 45.86% | 51.99% | 89.48% | 79.95% |
| Full set 50:50 | 43 Nodes | 40%DC40-2 | 43 Nodes | 40%DC40-2 | 43 Nodes | 40%DC40-2 | 43 Nodes | 40%DC40-2 | | |
| | 66.67% | | 65.87% | | 54.17% | | 51.73% | | | |
| | 86.49% | 62.35% | 88.15% | 61.35% | 51.92% | 54.78% | 45.09% | 53.69% | | |

The accuracy we achieved during our experiments is substantially higher than the accuracy other authors reported on approaches like neural networks[15] or case based reasoning.[16]

Nevertheless the overall results are good and the impression that even more knowledge can be extracted from the three data sets makes us plan new experiments with different training set combination and different machine learning approaches.

## CONCLUSION

The presented results show that we are on the right way to solve the acute appendicitis problem with the use of machine learning techniques, but many problems still remain to be solved. In our experiments we have shown that the gathering data from different types of sources can substantially influence the performance of classifiers, even though the methods for data gathering were almost the same. We have also shown, that larger size of the training sets does not necessary guarantee the higher accuracy in comparison to smaller training sets.

Currently we are testing improved machine learning methods such as genetic decision trees and hybrid neural decision trees on all the three AAP databases and preliminary results seem to be even a bit better than results presented in this paper.

## ACKNOWLEDGMENTS

The work was supported in the framework of a bilateral scientific-technological cooperation between Slovenia (Ministry of Science and Technology, Project No.: L2-1640-0796-99) and Germany (Deutsches Zentrum für Luft- und Raumfahrt e.V., Project No.: SVN 99/022).

## REFERENCES

1. af Klercker, T., Effect of pruning of a decision-tree for the ear, nose and throat realm in primary health care based on case-notes. *J. Med. Syst.* 20(4):215–226, 1996.
2. Zorman, M., Podgorelec, V., Kokol, P., Peterson, M., and Lane, J., Decision tree's induction strategies evaluated on a hard real world problem. *13th IEEE Symposium on Computer-Based Medical Systems* 22–24 June 2000, Houston, Texas, USA, pp. 19–24, 2000.
3. Andersson, R. E., Hungander, A., and Thulin, J. G., Diagnostic accuracy and perforation rate in appendicitis: Association with age and sex of the patient and with appendectomy rate. *European J. Surg.* 158:37–41, 1992.
4. Blind, P. J., and Dahlgren, S. T., The continuing challenge of the negative appendix. *Acta Chir. Scand.* 152:623–627, 1986.
5. Ohmann, C., Moustakis, V., Yang, Q., and Lang, K., Evaluation of automatic knowledge acquisition techniques in the diagnosis of acute abdominal pain. *Artif. Intell. Med.* 8:23–36, 1996.
6. Pesonen, E., Eskelinen, M., and Juhola, M., Comparison of different neural network algorithms in the diagnosis of acute appendicitis. *Int. J. Bio-Med. Comput.* 40:227–233, 1996.
7. de Dombal, F. T., *Diagnosis of Acute Abdominal Pain*, Churchhill Livingstone, Edinburgh, 1991, pp. 105–106.
8. Quinlan, J. R., *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Mateo, CA, 1993.
9. Russel, S. J., Norvig, P., *et al.*: *Artificial Intelligence: A Modern Approach*, Prentice-Hall, Englewood Cliffs, 1995, pp. 525–562.

10. Zorman, M., Hleb, Š., and Šprogar, M., Advanced tool for building decision trees MtDeciT 2.0. In Kokol, P., Welzer-Družovec, T., and Arabnia, H. R. (eds.), *International Conference on Artificial Intelligence*, June 28–July 1, 1999, Las Vegas, Nevada, USA. Book 1, pp. 315–318.
11. Zorman, M., and Kokol, P., Dynamic discretization of continuous attributes for building decision trees. In Fyfe, C. (ed.), *Proceedings of the Second ICSC Symposium on Engineering of Intelligent Systems*, EIS 2000, June 27–30, 2000, University of Paisley, Scotland, U.K. Academic Press, Wetaskiwin; Zürich, pp. 252–257, 2000.
12. de Dombal, F. T. , de Baere, H., van Elk, P. J., Fingerhut, A., Henriques, J., Lavelle, S. M., Malizia, G., Ohmann, C., Pera, C., Sitter, H., and Tsiftsis, D., Objective medical decision making in acute abdoinal pain. In Benken, J. E. W., and Thevin, V. (eds.), *Advances in Biomedical Engineering*, IOS Press, 1993, pp. 65–87.
13. Ohmann, C., Platen, C., Belenky, G., Franke, C., Otterbeck, R., Lang, K., *et al*.: Expertensystem zur Unterstützung von Diagnosestellung und Therapiewahl bei akuten Bauchschmerzen. *Informatik, Biometrie und Epidemiologie in Medizin und Biologie* 26(3):262–274, 1995.
14. Ohmann, C., Eich, H. P., and Sippel, H., A data dictionary approach to multilingual documentation and decision support for the diagnosis of acute abdominal pain (COPERNICUS 555, An European Concerted Action). *Medinfo* 9(Pt 1):462–466, 1998.
15. Pesonen, E., Ohmann, C., Eskelinen, M., and Juhola, M., Increasing the accuracy of the acute appendicitis of a LVQ neural network by the use of larger neighborhoods. *Methods Inf. Med.* 37(1):59–63, 1996.
16. Puppe, B., Ohmann, C., Goos, K., Puppe, F., and Mootz, O., Evaluating four diagnostic methods with acute abdominal pain cases. *Methods Inf. Med.* 34:361–368, 1995.