

## Mining Knowledge in Medical Image Databases

Petra Perner

IBaI, Arno-Nitzsche-Str. 45, 04277 Leipzig, Germany

e-mail: [ibaiperner@aol.com](mailto:ibaiperner@aol.com)

### ABSTRACT

Availability of digital data within picture archiving and communication systems raises a possibility of health care and research enhancement associated with manipulation, processing and handling of data by computers. That is the basis for computer-assisted radiology development. Further development of computer-assisted radiology is associated with the use of new intelligent capabilities such as multimedia support and data mining in order to discover the relevant knowledge for diagnosis. In this paper, we present our work on data mining in medical picture archiving systems. We use decision tree induction in order to learn the knowledge for computer-assisted image analysis. We are applying our method to interpretation of x-ray images for lung cancer diagnosis. We are describing our methodology on how to perform data mining on picture archiving systems and our tool for data mining. Results are given. The method has shown very good results so that we are going on to apply it to other medical image diagnosis tasks such as lymph node diagnosis in MRI and investigation of breast MRI.

**Keywords:** Data Mining, Decision Tree Induction, Medical Image Data Base, Medical Image Diagnosis, Knowledge Discovery

### 1. Introduction

Radiology departments are at the center of a massive change in technology. The ubiquitous radiographic film that has been the basis of image management for almost 100 years is being displaced by new digital imaging modalities such as 1. computed tomography (CT); 2. magnetic resonance (MR); 3. nuclear medicine (NM); 4. ultrasound (US); 5. digital radiography (DF); 6. computed radiography (CR) using storage phosphor imaging plates or film digitizers, 7. digital angiography (DA); 8. MR spectroscopy (MRS); 9. electron emission radiography (EMR).

These digital modalities are continuously refined and new digital applications are being developed. The scientific prognosis is that about 80 of patient imaging examination in radiology department will be performed using digital imaging modalities already at the end of this century<sup>1</sup>.

Digital image management systems are under development now to handle images in digital form. These systems are termed Picture Archiving and Communication Systems (PACS)<sup>2,3</sup>. The PACS are based on the integration of different technologies that form a system for image acquisition, storage, transmission, processing and display of images for their analysis and further diagnosis. The main objective of such systems is to provide a more efficient and cost-effective means of examining, storing, and retrieving diagnostic images. These systems must supply the user with easy, fast, reliable access to images and associated diagnostic information.

Availability of digital data within the PACS raises a possibilities of health care and research enhancements associated with manipulation, processing and handling of data by computers. That is a basis for computer-assisted radiology development. However, that will only work if the systems are carefully designed so that they supply sufficient data for the development of decision support systems. Often this aspect has not been considered when implementing a radiology information system.

In this paper, we present our work on data mining for picture archiving systems in medicine. We explain our methodology for performing data mining in picture archiving systems. The experiment described here has provided the methodology for other medical image diagnosis experiments. The experiment has also led to a variety of other data mining applications in medical image diagnosis such as lymph node diagnosis in MRI and investigation of breast MRI.<sup>4</sup> It also has shown the advantage of data mining over other techniques for improving the quality of image diagnosis in medical applications and it provides in the long run the opportunity for the development of fully automatically image diagnosis systems.

In Section 2, we describe the problems by the development of computer-assisted medical diagnosis systems and the application itself. The method used for data mining are described in Section 3 and results are given in Section 4. In another experiment, we use feature subset selection before applying decision tree induction. The method and the results are described in Section 5. Finally, we summarize our experience in Section 6.

## 2. Application

Knowledge acquisition is the first step in developing an image interpretation system. The kind of method used for knowledge acquisition depends on the inference method the image interpretation system is based on.

The knowledge acquisition process for rule-based system is usually manually done by interviewing a human expert<sup>5</sup> or by employing interactive knowledge acquisition tools such as e.g. repertory grid.<sup>6</sup>

In model-based systems, the knowledge about the objects is represented based on semantic nets that structure the knowledge into concepts and their relations. The language of the semantic net determines the way new knowledge is elicited. Kehoe et al.<sup>7</sup> describe a model based system for defect classification of welding seams. The knowledge base is manually maintained by specializing or generalizing the defect classes, their attributes, and attribute values. Schröder et al.<sup>8</sup> described a system where knowledge acquisition is done automatically based on the language of the semantic net. Although semantic nets seem to be the most convenient way of representing and eliciting knowledge, this method requires a deep understanding of the domain, which is not given a-priori for all applications.

When generalized knowledge is lacking, then case based reasoning<sup>9</sup> seems to be a proper method. This system is based on a case base consisting of a set of cases. An interpretation is made by determining the closest case or cases in the case base to the actual case and by displaying the value of the closeness measure and the interpretation associated with the similar case of the case base. How the closeness measure should be interpreted is left to the user. The limited explanation capabilities are the main drawback of case based reasoning systems.

We want to develop a knowledge acquisition method for such applications where no generalized knowledge about the domain is available but a large data base of images associated with expert description and interpretation. If we think of the recent trend to picture archiving systems in medicine and other domains, tasks such as these become quite important.

Therefore, the aim of our project is development of knowledge acquisition methods for medical image diagnosis, which can help to solve some cognitive, theoretical and practical problems:

1. Decision model of an expert for specific tasks solution will be reproduced and displayed.
2. It will show the pathway of human reasoning and classification. Image features which are basic for correct decision by expert will be discovered.
3. Developed model will be used as a tool to support decision-making of physician, who is not an expert in a specific field of knowledge. It can be used for teaching of decision-making.

The application of data mining will help to get some additional knowledge about specific features of different classes and the way in which they are expressed in the image (can help to find some inherent non-evident links between classes and their imaging in the picture).

It can help to get some nontrivial conclusions and predictions can be made on the base of image analysis.

For our experiment, we used a database of tomograms of 250 patients with verified diagnoses (80 cases with benign disease and 138 cases with cancer of lung). Patients with small pulmonary nodules (up to 5 cm) were selected for this test. Conventional (linear) coronal plane tomograms with 1 mm thickness of section were used for specific diagnosis.

Original linear tomograms were digitized with step of 100 micron (5,0 line pairs per millimeter) to get 1024 x 1024 x 8 bits matrices with 256 levels of gray.

The use of linear tomograms and such a digitization enabled an acquisition of high spatial resolution of anatomical details that were necessary for the specific diagnosis of lung nodules.

To improve results of specific diagnosis of small solitary pulmonary nodules we used optimal digital filtering and analysis of post-processed images. The processing emphasized diagnostically important details of the nodule and thus helped to improve the reliability of image analysis: the physician was more certain in feature reading and interpretation. The radiologist worked as an expert on this system.

### 3. Methods and Approach Taken

First, an attribute list was set up together with the expert, which covered all possible attributes used for diagnosis by the expert as well as the corresponding attribute values, see Table 1. We learned our lesson from another experiment and created an attribute list having no more than three attribute values. Otherwise, the resulting decision tree is hard to interpret and the tree building process stops very soon because of the splitting of the data set into subsets according to the number of attribute values.

Then, the expert collected the database and communicated with a computer answering to its requests. He determined whether the whole tomogram or its part had to be processed and outlined the area of interest with overlay lines and he also outlined the nodule margins. The parameters of optimal filter were then calculated automatically. A radiologist watched the processed image (see Fig. 1) displayed on-line on a TV monitor, evaluated its specific features (character of boundary, shape of the nodule, specific objects, details and structures inside and outside the nodule, etc.), interpreted these features according to the list of attributes and inputted the codes of appropriate attribute values into the database program. Hard copies of the previously processed images from the archive have been used in this work as well.

The collected data set was given as a dBase-file to the inductive machine learning tool.

For the data mining experiment we used our tool DECISION\_MASTER.<sup>10</sup> It can create binary and n-ary decision trees from the data. It has several options which makes it possible to specify how numerical features should be partitioned<sup>11</sup> and what method should be used for feature selection. Evaluation of the results can be done by test-and-train and n-fold crossvalidation. Missing values can be handled by different strategies. The tool also provides functions for outlier detections. Once the diagnosis knowledge has been learnt, the rules are provided weather in txt-format for further use in an expert system or the expert can use the diagnosis component of DECISION\_MASTER for interactive work. The tool is written in C++ and runs under Windows95 and Windows NT. It has a user-friendly interface and is set up in such a way that it can be handled very easily by non-computer specialists.

Figure 1 shows our structure for a Picture Archiving System combined with the data mining tool.

We used a decision tree induction method for our experiment which creates binary-trees based on maximum entropy criteria.<sup>12</sup> Pruning is done based on reduced-error pruning technique.<sup>13</sup> Evaluation was done by 10-fold crossvalidation. Besides the error rate we calculate Sensitivity for Class\_1 and Specificity for Class\_2, which are error criteria normal required for medical applications:  $E_{sens} = S_{c1m} / N_{c1}$   $E_{Spec} = S_{C2m} / N_{C2}$  with  $S_{c1m}$  the number of misclassified samples of class 1 and  $N_{C1}$  the number of all samples of class 1 and  $S_{C2m}$  and  $N_{C2}$  respectively.

<b>Attribute</b>	<b>Short Name</b>	<b>Attribute Values</b>
Class	<i>CLASS</i>	1 malignant 2 benign
Structure inside the nodule	<i>STRINSNOD</i>	1 Homogeneous 2 Inhomogeneous
Regularity of Structure inside the nodule	<i>REGSTRINS</i>	1 Irregular Structures 2 Regular orderly
Cavitation	<i>CAVITATIO</i>	0 None 1 Cavities
Areas with calcifications inside the nodule	<i>ARWCAL</i>	0 None 1 Areas with calcifications
Scar-like changes inside the nodule	<i>SCARINSNOD</i>	0 None 1 Possibly exists 2 Irregular fragmentary dense shadow
Shape	<i>SHAPE</i>	1 Nonround 2 Round 3 Oval
Sharpness of margins	<i>SHARPMAR</i>	1 NonSharp 2 MixedSharp 3 Sharp
Smoothness of margins	<i>SMOMAR</i>	1 NonSmooth 2 MixedSmooth 3 Smooth
Lobularity of margins	<i>LOBMAR</i>	0 NonLobular 1 Lobular
Angularity of margins	<i>ANGMAR</i>	0 Nonangular 1 Angular
Convergence of vessels	<i>CONVVESS</i>	1 Vessels constantly 2 Vessels are forced away the nodule 3 None
Vascular Outgoing Shadows	<i>VASCSHAD</i>	0 None 1 Chiefly vascular
Outgoing sharp thin tape-lines	<i>OUTSHTHIN</i>	0 None 1 Outgoing sharp thin tape-lines
Invasion into surrounding tissues	<i>INVSOURTIS</i>	0 None 1 Invasion into surrounding tissues
Character of the lung pleura	<i>CHARLUNG</i>	0 No Pleura 1 Pleura is visible
Thickening of lung pleura	<i>THLUNGPL</i>	0 None 1 Thickening
Withdrawing of lung pleura	<i>WITHLUPL</i>	0 None 1 Withdrawing
Size of Nodule	<i>SIOFNOD</i>	Numbers (eg, 1.2)in cm

Table 1 Attribute List

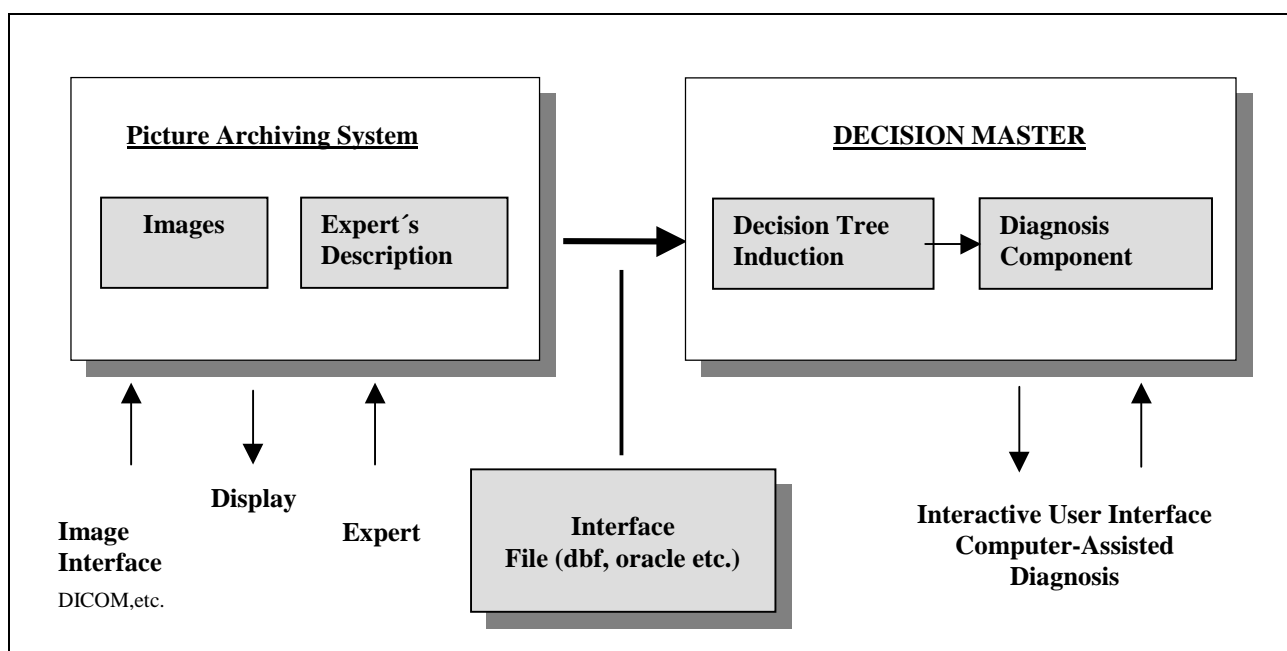


Figure 1 Architecture

#### 4 Results

The unpruned tree consists of 20 leaves, see Figure 2. The pruned tree consists of 6 leaves, see Figure 3. Our expert liked the unpruned tree much more since nearly all attributes he is using for decision making appeared in the tree. The expert told us that the attribute *Structure* is very important, also the attribute *Scar-like changes inside the nodule*.

However the expert wonders why other features such as *Structure* and some others didn't work for classification. The expert told us that he usually analyzes a nodule starting with its *Structure*, then tests *Scar-like changes inside the nodule*, then *Shape* and *Margin*, then *Convergence of Vessels* and *Outgoing Shadow in Surrounding tissues*. Although decision trees represent the decision in a comprehensible format to human, the decision tree might not represent the strategy used by an expert since it is always the attribute appearing first in database and satisfying the splitting criteria that is chosen.

Therefore, we investigated the error rate as main criterion, see Tab. 2 and Tab. 3.

We did not come close to the expert's performance. One reason might be the choice of attribute values. For some categorical attributes, there are too many categorical values. That causes that during the tree building process the training set is split up into too many subsets with few data samples. As a result the tree building process will stop very soon since no discrimination power is left in the remaining data samples.

Although decision trees represent the decision in a human understandable format, the decision tree might not represent the strategy used by an expert since always the attribute appearing first in database and satisfying the splitting criteria, is chosen.

Therefore, we looked for the error rate as main criterion, see Tab. 2 and Tab. 3. The decision tree could not outperform the expert but showed good results.

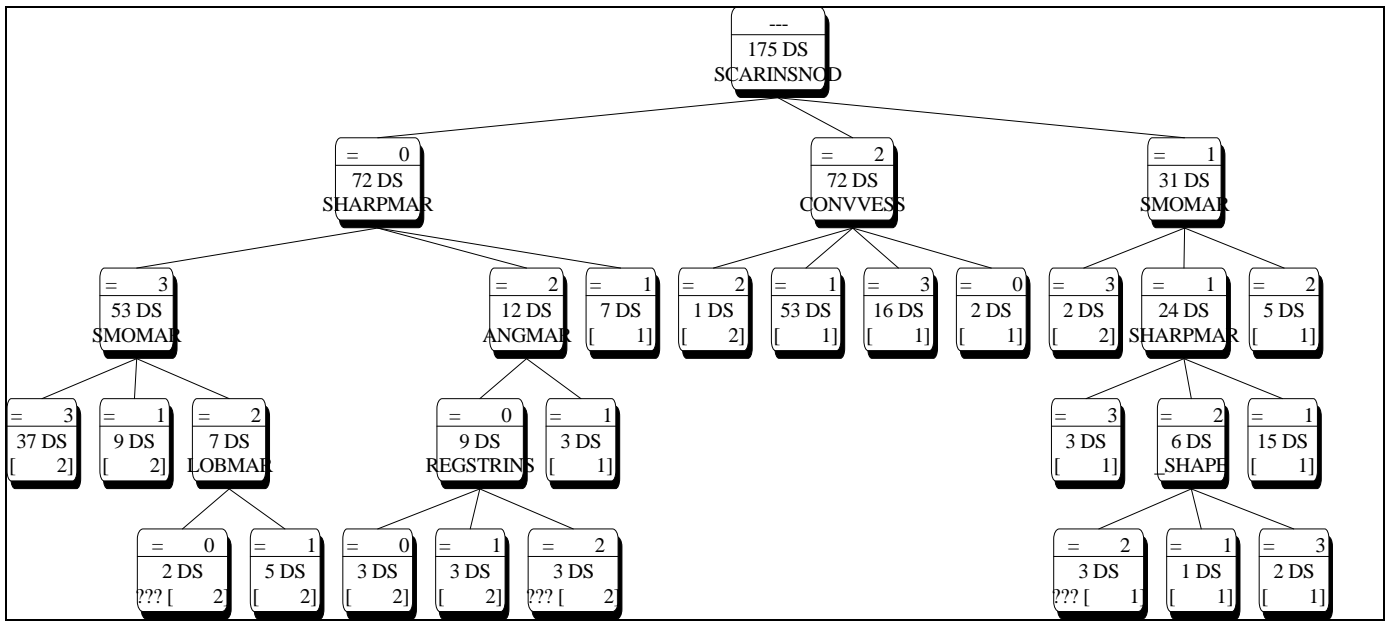


Figure 2 Decision Tree (unpruned)

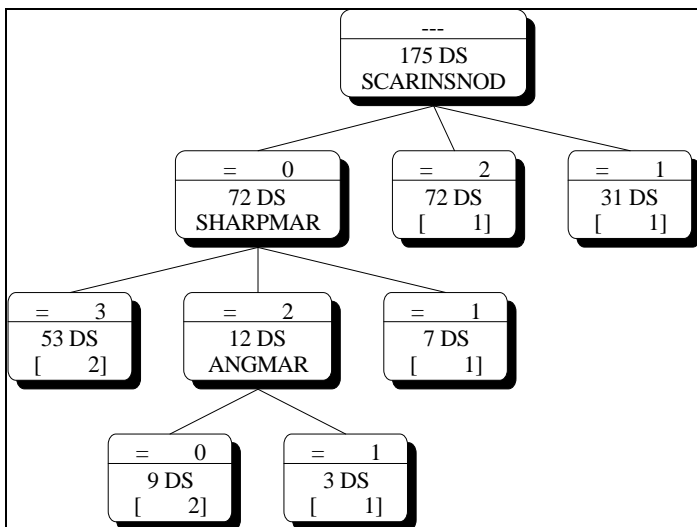


Figure 3 Decision Tree (pruned)

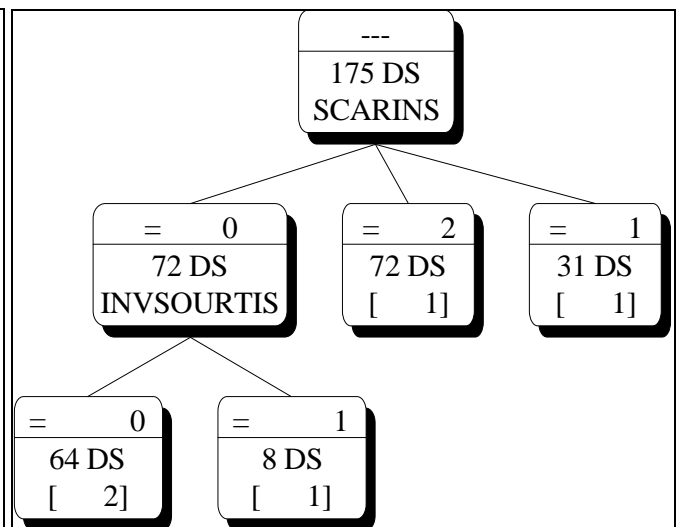


Figure 4 Decision Tree for Feature Subset

	<i>Before pruning</i>	<i>After pruning</i>
	<b>Error Rate</b>	<b>Error Rate</b>
(1)	6,857 %	7,428 %
(2)	6,30%	7,3 %

Table 2 Result (1) and Evaluation of Decision tree on Test Data (2)

<i>Accuracy</i>		<i>Sensitivity/Specificity</i>			
<b>Human</b>	<b>DT</b>	<b>Class_1</b>		<b>Class_2</b>	
		<b>Human</b>	<b>DT</b>	<b>Human</b>	<b>DT</b>
94,4 %	93,2 %	97,5 %	93 %	91,4 %	90 %

**Table 3 Comparison between Human Expert and Decision Tree Classification**

In another test, we used test data set obtained by image reading from a middle level expert on difficult cases of images. This expert had no special courses of processed image reading and interpretation. It was our special intention to obtain as much uncertainty and noise in the data as possible. As the expert did not know how to read a new roentgenological picture that appeared after digital image processing, in many cases he could not choose which attribute value was suitable for this case (marked as missing attribute value in data base).

<i>Accuracy</i>		<i>Sensitivity/Specificity</i>			
<b>Human</b>	<b>DT</b>	<b>Class_1</b>		<b>Class_2</b>	
		<b>Human</b>	<b>DT</b>	<b>Human</b>	<b>DT</b>
94,5%	95,7%	96,2%	93,65%	90%	99%
55,2%	73%	61,1%	74%	50%	72,5%

**Table 4 Comparison between Human Expert and Decision Tree Classification (1) High-level Expert (2) Middle-level Expert**

These readings were given to the decision tree for classification. The resulting error rate showed that classifier based on decision tree gives reliable error rate even by bad image readings, see Table 4.

## 5. Feature Subset Selection

Ideally, decision tree induction should use only the subset of features that leads to the best performance. Induction algorithm usually conduct a heuristic search in the space of possible hypotheses. This heuristic search may lead to induced concepts which depend on irrelevant features, redundant, or correlated features. The problem of feature subset selection involves finding a "good" set of features under some objective function. We consider our feature subset selection problems as a problem of finding the set of features which are most dissimilar to each other. Two Features having high similarity value are used by the experts in the same manner for the target concept. They are redundant and can be removed from the feature set. It can be shown in practice that this assumption holds for most of the applications.<sup>15</sup> For our experiment, we used Kruskal's tau<sup>14</sup> as similarity function and single linkage method to visualize the similarity relation, see Figure 5. We can see that *Character of Lung Pleura* and *Within Lung Pleura* are more or less used in the same manner. The expert confirmed this observation. However, on his opinion it is necessary to have both features since sometimes one feature does exist and the other does not exist. From the dendrogram shown in Figure 5, we created different subsets of features with 10, 13, 15, and 17 features by selecting the most dissimilar features. The first subset included the following 10 features: REGSTRINS, ARWCAL, LOBMAR, CONVVESS, VASC SHAD, OUTSHTHIN, INVSOURTIS, SIOFNOD, SPICMAR, and SHAPE. The next subset of features included three more features with high dissimilarity value and so on. From these subsets Decision Master induced decision trees and calculated the error rate based on cross validation. We observed that a better error rate can be reached if the decision tree is only induced from a subset of features, see Figure 6. The method used in this paper does not tell us what is the right number of features. This, we can only find out by running the experiment. Another side effect is that the resulting decision tree is more compact.

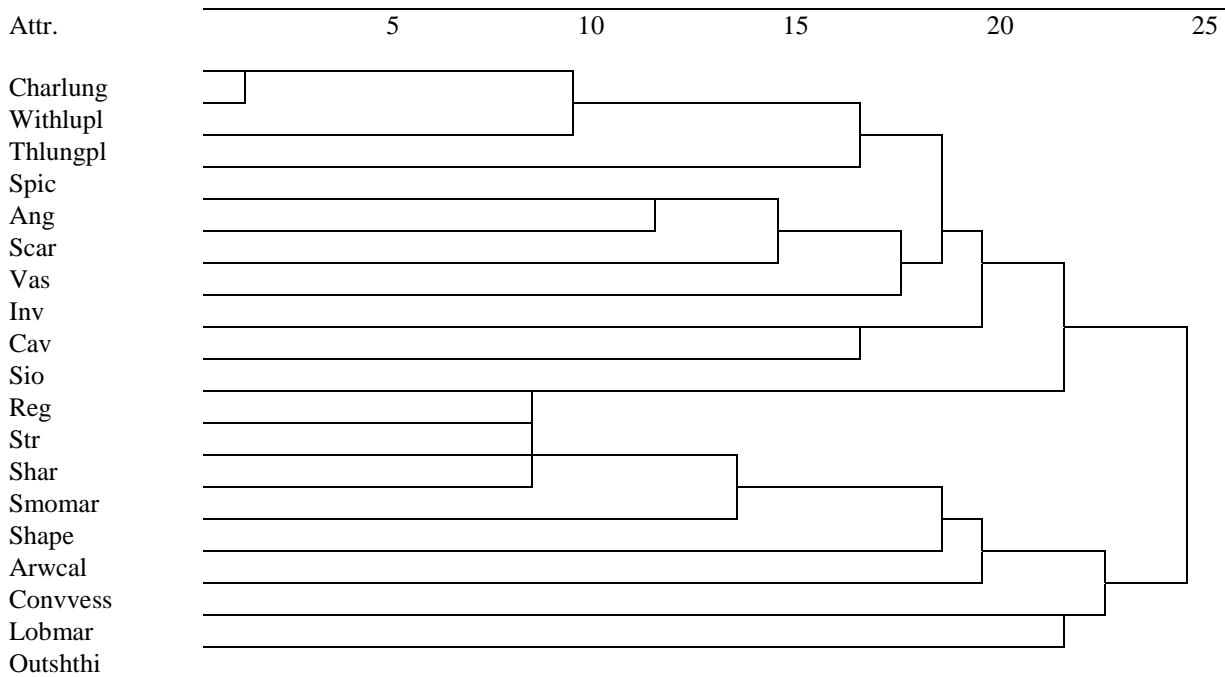


Figure 5 Dendrogram Similarity between the Features

Feature Number	Unpruned Decision Tree Error Rate	pruned Decision Tree Error Rate
19	6,8571	7,428
10	10,85	14,85
13	7,4286	4,5714
15	7,429	4,5714
17	10,28	7,42

Table 6 Error Rate for different Feature Subsets

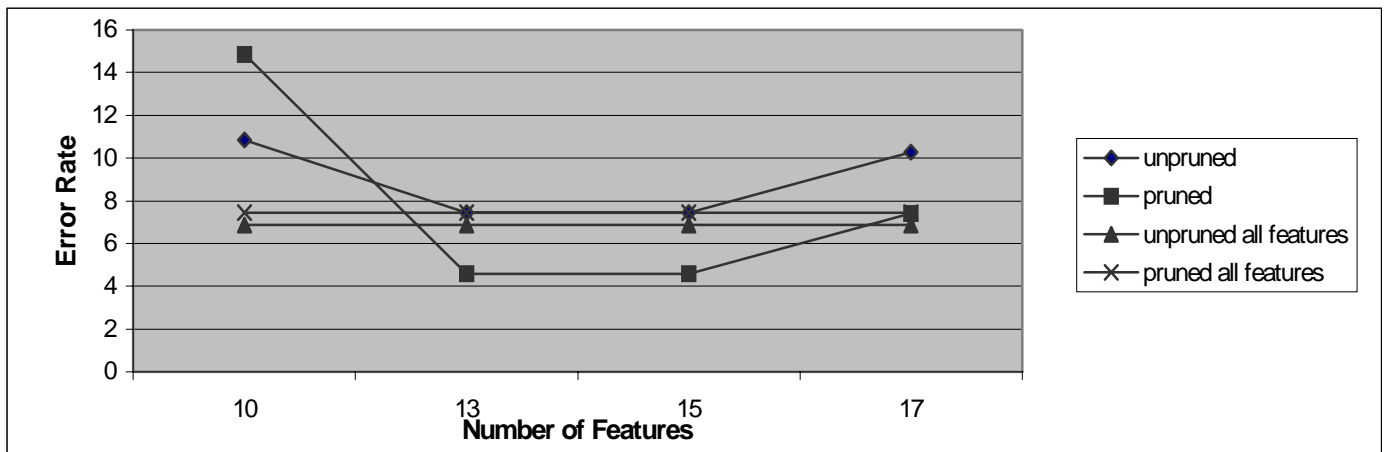


Figure 6 Diagramm Error Rate for different Feature Subsets



## 6 Conclusion and Further Work

In this paper, we presented our methodology for data mining in picture archiving systems. The basis for our study is a sufficiently large database with images and expert descriptions. Such databases result from the broad use of picture archiving systems in medical domains.

We were able to learn the important attributes needed for interpretation and the way in which they were used for decision making from this database by applying data mining methods. We showed how the domain vocabulary should be set up in order to get good results and which techniques could be used in order to check reliability of the chosen features.

The explanation capability of the induced tree was reasonable. The attributes included in the tree represented the expert knowledge.

Finally, we can say that picture archiving systems in combination with data mining methods open the possibility of advanced computer-assisted medical diagnosis systems. However, it will not give the expected result if the PACS have not been set up in the right way. Pictures and experts descriptions have to be store in a standard format in the system for further analysis. Since standard vocabulary<sup>8</sup> and very good experts are available for many medical diagnosis tasks this should be possible. If the vocabulary is not a priori available, then vocabulary can be determined by a methodology based repertory grid.<sup>15</sup> What is left is to introduce this method to the medical community, which we are going on to do recently for mammography image analysis and lymph nodule diagnosis. Unfortunately, it is not possible to provide an image analysis systems, which can extract features for all kind of images. Often it is the case that it is not clear how to describe a particular feature by automatic image feature extraction procedures. The expert's description will still be necessary for a long time. However, once the most discriminating features have been found the result can lead in the long run to fully automatic image diagnosis system which is set up for specific type of image diagnosis.

## References

1. H.U. Lemke, Medical imaging and computer assisted radiology, Tutorial Notes of CAR'91 (The 5-th International Symposium on Computer Assisted Radiology) Berlin, July 2-3 1991.
2. D.M. Chaney, Coordinate your plans for EMR and PACS investments, Health Management Technology (Aug. 199) vol. 20, no.7, p.24, 26-7.
3. K. Adelhard, S. Nissen-Meyer, C. Pistitsch, U. Fink, M. Reiser, Funtional requirements for a HIS-RIS-PACS-interface design, including integration of "old" modalities, Methods of Information in Medicine (March 1999) vol. 38, no. 1, p. 1-8.
4. S. Heywang-Köbrunner, P. Perner, Optimized Computer-Assistet Diagnosis based on Data Mining, Expert Knowledge and Histological Verification, IBaI Report August 1998 ISSN 1431-2360.
5. P. Perner, „A knowledge-based image inspection system for automatic defect recognition, classification, and process diagnosis.“ Int. Journal on Machine Vision and Applications, 7 (1994): 135-147.
6. J.H. Boose, D.B. Shema, and J.M. Bradshaw, „Recent progress in Aquinas: a knowledge acquisition workbench,“ Knowledge Acquisition 1 (1989): 185-214.
7. A. Kehoe and G.A. Parker, „An IKB defect classification system for automated industrial radiographic inspection,“ IEEE Expert Systems (1991) 8, pp. 149-157.
8. S. Schröder, H. Niemann, G. Sagerer, „Knowledge acquisition for a knowledge based image analysis system,“ In: Proc. of the European Knowledge-Acquisition Workshop (EKAW 88), Bosse, J. and Gaines, B. (ed.), GMD-Studien Nr. 143, Sankt Augustin, 1988.

P. Perner, Mining Knowledge in Medical Image Databases, In Data Mining and Knowledge Discovery: Theory, Tools, and Technology, Belur V. Dasarathy (eds.), Proceedings of SPIE Vol. 4057 (2000), 359-369

9. P. Perner, Case-Based Reasoning for the Low-level and High-level Unit of an Image Interpretation System, Sameer Singh (Eds.), Advances in Patter Recognition, Springer Verlag 1998, p. 45-54.S.
10. DECISION\_MASTER <http://www.ibai-solutions.de>
11. P. Perner and S. Trautzsch, Multinterval Discretization for Decision Tree Learning, In: Advances in Pattern Recognition, A. Amin, D. Dori, P. Pudil, and H. Freeman (Eds.), LNCS 1451, Springer Verlag 1998, S. 475-482
12. H.S. Baird and C.L. Mallows, „Bounded-Error Preclassification Trees,“ In: Shape, Structure and Pattern Recognition, Eds. D. Dori and A. Bruckstein, World Scientific Publishing Co., 1995, pp.100-110.
13. J.R. Quinlain, „Simplifying decision tree,“ Intern. Journal on Man-Machine Studies, 27 (1987): 221-234.
14. A. Agresti, „Categorical Data Analysis,“ John Wiley Inc. 1990
15. P. Perner, “ How to use Repertory Grid for Knowledge Acquisition in Image Interpretation,“ HTWK Report, January 1994