Transactions on Machine Learning
and Data Mining
Vol.2, No 1 (2009) 23-43
© ISSN: 1865-6781 (Journal),
ISBN: 978-3-940501-11-0,
IBaI Publishing ISSN 1864-9734

www.ibai-publishing.org

Mining Knowledge for Differential Diagnosis of Pulmonary Nodules

Claudia Seifert[1] and Andrea Ahlemeyer-Stubbe[2]

[1]Institute of Computer Vision and Applied Computer Sciences IBaI, Germany
[2]antz21 GmbH, Gengenbach, Germany

Abstract. The development of computer-assisted radiology in PACS is associated with the use of new intelligent capabilities of computer-assisted radiology. In this paper we present our work on data mining to support medical image diagnosis. We use decision tree induction in order to learn the expert knowledge, presented in the form of image descriptions in the database. A methodology is developed to perform data mining in picture archiving systems. The tool for data mining is presented. It was applied in the task of early differential diagnosis of pulmonary nodules in lung tomograms and allowed to support image analysis and classification by the expert. Results of knowledge acquisition on the basis of data mining are presented and compared with the results of the semi-automated method of knowledge acquisition by a syndrome framework tool. The developed decision tree and developed syndrome-like decision rule are effective for early diagnosis of peripheral lung cancer, so that we are able to apply this method to other medical tasks, in particular for image analysis in mammography.

Keywords: Data mining, decision support, image interpretation, decision trees, syndrome classifier

1 Introduction

Nowadays radiology departments are in the center of radical changes in technologies of acquisition, storage, transmission, and analysis of diagnostic images. The radiographic films, which have been used for image analysis since 1895, are being replaced now by digital images acquired by new imaging techniques (CT, MRI, etc. [1][2]). Picture archiving and communication systems (PACS) were developed to provide efficient and cost-effective analysis, exchanging, archiving, and retrieving of diagnostic images presented in digital form [2]-[4]. It opens new extended possibilities for manipulation, processing, and handling medical data by computers to improve the efficacy of medical investigations. PACS gives the user the means to surpass current diagnostic capability thanks to the achievements of computer-assisted radiology such as multi-modality imaging and multimedia displaying of medical data, image processing and computer-assisted decision-making [5]-[7].

Further developments of computer-assisted radiology are associated with the use of new intelligent capabilities, such as data mining and automatic image analysis, which are being developed to support decision- making by the expert.

Medical-image diagnosis involves several task solutions: a) lesion detection; b) recognition of diagnostically important features (structures, details, objects, etc.); also c) interpretation of the image as a whole. A solution of all these tasks requires long practice and specific domain knowledge for image-feature analysis and scene understanding. In the most interesting and practical cases, such as early stage diagnosis, the physician has to detect tiny changes in the complex image background, to analyze the lesions, which have no obvious specific features, and to make a differential diagnosis on the basis of noisy, incomplete, and unbiased data, obtained as a result of image analysis.

Image processing could promote image analysis by more explicit imaging of specific features. The application of data mining can help to obtain knowledge about specific features of different classes, and to create models for decision-making. It can also provide the discovery of some inherent non-evident links between classes and their imaging in the picture.

The goal of this work was the development of knowledge acquisition methods for medical image diagnosis, based on data mining, which could help to solve some cognitive, theoretical and practical problems:

to find features, which are basic for human reasoning and classification and to discover the tasks of expert's purpose vision.

to reproduce and to display a decision model for a specific medical task solution, which can be used as a tool to support decision-making by physicians, who are not experts in the specific field of knowledge. Such a model could also be useful for teaching novices.

The second goal of the work was the development of methodology for the application of data mining in medical picture archiving systems. We applied knowledge acquisition methods to the task of early differential diagnosis of solitary pulmonary nodules in lung tomograms. In Section 2 we describe the recent approaches to knowledge acquisition in different fields of knowledge. The materials and the developed method are described in Section 3. Results are given in Section 4. We compare facilities of the developed automatic and semi-automatic knowledge acquisition methods and give the results in Section 5. Finally, we summarize our experience in Section 6.

2 Knowledge Acquisition Methods

Knowledge acquisition is the first step when developing a system to support image analysis and interpretation. The kind of method used for knowledge acquisition depends on the inference method, on which the system is based [9]-[14].

The knowledge acquisition process for a rule-based system is usually done manually by interviewing a human expert [9] or by employing interactive knowledge acquisition tools, e.g. repertory grids [10].

In model-based systems, the knowledge representation is based on semantic nets. These allow to structure the domain knowledge into concepts and their relations. The language of semantic nets determines the way in which new knowledge is elicited. Kehoe et al. [11] described a model-based system for defect classification of welding seams. The created knowledge base was maintained manually by specializing or generalizing the defect classes, their attributes, and attribute values. Schröder et al. [12] described a system, where knowledge acquisition is done automatically on the basis of 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 and availability of generalized knowledge, which is not available a priori for all applications.

When generalized knowledge is lacking, then case-based reasoning [13] seems to be a proper method. An interpretation of such a system is based on determining the case (or cases), stored in the case base, which is the closest to the actual case. The measured value of closeness and the interpretation, associated with the similar case in the case base, is presented to the user. How to interpret the closeness is left to the user. A case-based reasoning system developed for image interpretation is described in [14]. The limited explanation capability is the main lack of case-based reasoning systems.

We are developing a knowledge acquisition method for those applications, where no generalized domain knowledge is available, but where there is a large database of images, associated with expert description and interpretation. This task is very up to date, taking into account the recent trend in the wide use of picture archiving systems in medicine and in other domains.

The application of data mining could help to get some additional knowledge about specific features of different classes and variations of their manifestation on the image.

Such a system opens possibilities for an automated expert-independent image analysis based on measuring some specific features directly in the image.

3 Materials and Methods

For our experiment we used a database of tomograms with verified diagnoses. Annual screening by means of radiography identified all studied patients with pulmonary nodules. Patients with pulmonary nodules up to 3 cm in size were selected (Set 1: 218 cases =80 benign + 138 malignant (peripheral lung cancer)). Patients with a clinically verified cancer in other organs were excluded from the test to avoid cases of metastases in lung.

For our test experiment we selected 38 images (Set 2: 38 cases=20 malignant+18 benign). About a half of these images was referred to as complex cases, as they yielded ambiguous diagnostic decisions during the analysis of the unprocessed images by the expert.

3.1 Acquisition of Lung Images

Conventional (linear) coronal plane tomograms with 1-2 mm section thickness were used for a specific diagnosis. The slice was centered through the nodule. One or two contiguous sections with 3 mm advance between the slices have been done for some cases as well.

Original linear tomograms were digitized with a/the steps of 100 micron (5,0 line pairs per millimeter) to get 1024 x 1024 x 8 bits matrices with 256 levels of gray. Such a digitization promoted an acquisition of images with high spatial resolution of anatomical details that were necessary for the specific diagnosis of lung nodules.

3.2 Enhancing Diagnostically Important Details

A method of digital filtering [15] was applied to improve the specific diagnosis of small solitary pulmonary nodules in lung tomograms. The goal of the processing was to emphasize the specific nodule features and to make them more distinct. The fulfilled steps involved:

- Modeling the mixture of the background and the useful signal for the observing image set [16][17];
- Development of an optimal filter to suppress the impact of the background and to emphasize the remaining part of the image, which contained informative details [17];
- Modeling the background which fits to the considered class of images [17];
- Testing the processed images by the experts to select the background model, which allows the best imaging of diagnostic structures and the highest efficacy of image interpretation by the physician. The feedback was used to find the optimal filter, which met the expert criteria, and gave the best feature presentation for expert’s purpose vision [17][18].
- X-ray -morphological comparison was performed in order to confirm that all details, displayed in the processed tomogram, corresponded to real morphological structures found by morphologists in post surgical histotopograms [17][18].

The processing improved the manifestation of diagnostically important details on medical images and thus helped the physician to be more certain in feature reading and interpretation.

Table 1. Classes and Attribute List (1), developed for processed tomograms with small nodules

Attribute N	Attribute Type	Attribute Name	Short name	Value N	Attribute Value
1	Boolean	Class	Class	1 2	Malignant Benign
2	Categorical	Structure inside the module	StrInsNod	1 2 3 4 5 6 7 8	Inhomogeneous with disorderly structures Inhomogeneous with orderly structure Regularly decreasing film density along the periphery of the nodule Areas with calcifications Enough homogeneous structures Inhomogeneous with calcifications Inhomogeneous with orderly structures and calcifications Inhomogeneous with cavities
3	Categorical	Scar-like changes inside the nodule	ScrLikeChan	1 2 3	Irregularly shaped fragmentary dense shadow Regular dense shadow along periphery None
4	Categorical	Shape	Shape	1 2 3 4 5	Nonround Round Oval Lobular Angular
5	Categorical	Margin	Margin	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15	Nonsharp Sharp Nonsmooth Smooth Lobular Angular Spicular Nonsharp-sharp (nonsharp in some regions and sharp in others) Nonsharp and Nonsmooth Nonsharp and Angular Nonsharp and Spicular Sharp and Smooth Sharp and Lobular Sharp and Angular Nonsharp-sharp and Angular
6	Categorical	Convergence of vessels	ConvofVes	1 2	Vessels constantly exist converging to the nodule Vessels are forced away the nodule
7	Categorical	Outgoing shadows in surrounding tissues	OutgoShadinTis	1 2 3 4	Chiefly vascular Outgoing sharp tapelines (septa) None Invasion into surrounding tissues
8	Continuos	Size of Nodules	Size		Values in cm (e. g. 1,2 := 1,2 cm)
9	Categorical	Character of the Lung pleura	CharPleu	1 2 3 4	Thickening Withdrawing None Thickening with Withdrawing

However, in diagnostically complex cases good feature imaging was insufficient for correct image interpretation by the radiologist. It was caused by the absence of evident specific features at the early stage of disease, and by insufficient knowledge for making decisions on the basis of incomplete and fussy data, obtained during the image analysis. We applied data mining methods to support decision-making by the physician on the basis of the observed complex of features.

3.3 Acquisition of Expert Concepts for Image Description

First, together with the expert, we developed an attribute list, which the radiologist applied for image description. The attribute list covered all concepts about diagnostically important features and feature values, which the expert used during image analysis and interpretation.

The expert described each processed image from our set of images in the terms of selected attributes, which were then collected in symbolic form in the database of image descriptions.

We tested two variants of attribute lists, which differed in the presentation of feature values [19] (see attribute list (1), table1 and attribute list (2), table 2).

Table 2. Classes and Attribute List (2), developed for processed tomograms with small nodules

No. of Attribute	Attribute type	Attribute Name	Short Attribute Name	No. of Values	Attribute Value
...	...	...	...	...	...
8	Categorical	Sharpness of Margin	SharpMar	1 2 3	Nonsharp Mixedsharp Sharp
9	Categorical	Smoothness of Margin	SmoothMar	1 2 3	Nonsmooth Mixedsmooth Smooth
10	Boolean	Lobularity of Margin	LobMar	0 1	Nonlobular Lobular
11	Boolean	Angularity of Margin	AngMar	0 1	Nonangular Angular
12	Boolean	Spicularity of Margin	SpicMar	0 1	Nonspicular Spicular
...	...	...	...	...	...
14	Boolean	Vascular Outgoing Shadow	Vascshad	0 1	None Chiefly vascular shadows
15	Boolean	Outgoing Sharp thin tape lines	OutgoSha	0 1	None Outgoing sharp thin tape lines
...	...	...	...	...	...
18	Boolean	Thickening of lung pleura	ThLungPl	0 1	None Thickening
19	Boolean	Withdrawing of lung pleura	WithLupl	0 1	None Withdrawing
...	...	...	...	...	...

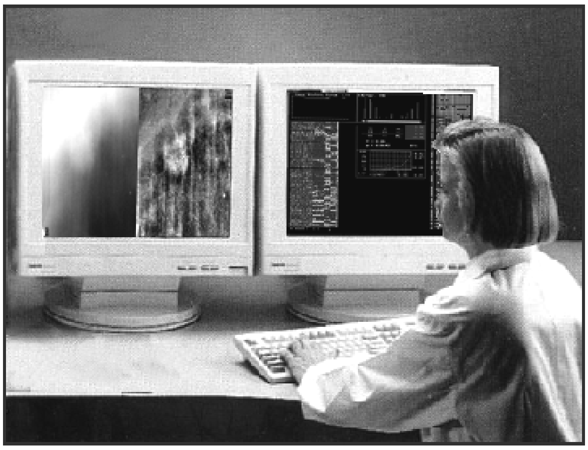

Fig. 1. Workstation for image processing and analysis

3.4 System for Image Processing and Analysis

Experiments with digital image processing and analysis were carried out, using the workstation developed in the Institute for Information Transmission Problems Russian Academy of Sciences [17]. The Workstation comprised a PC computer with hard and floppy discs, a device for image input (laser drum scanner or plane-table scanner), a high-resolution color monitor and printer.

The radiologist worked as an expert in this system. He communicated with a computer in order to determine and outline the area of interest and the area of the nodule. The parameters of the optimal filter were then calculated automatically [17]. The radiologist watched the processed image (Fig. 1, Fig. 2), displayed on-line on a TV monitor, evaluated its specific features (character of boundary and shape of the nodule, specific objects, details and structures inside and outside the nodule, etc.) according to the list of attributes (table 1 or table 2) and fed the codes of appropriate attribute values into the database program. Uncertain and missed values were available in the database as well. Hard copies of the previously processed images from the archive have also been used in this work.

Two radiologists of different qualification levels took part in our experiments. The first one was the highly skilled 'expert', who had a long experience in image reading and interpretation. The second one was a pulmonologist, who was not trained to read processed images. So he was a 'novice' in the interpretation of details in post processed tomograms. Each radiologist analyzed the processed images and gave appropriate attribute values into the database. He gave no value if he was not sure what value fitted to the case.

The collected data set was then passed as a dBase-file to the inductive machine-learning tool [20] and to the partner system [21] for a syndrome-like decision rule creation.

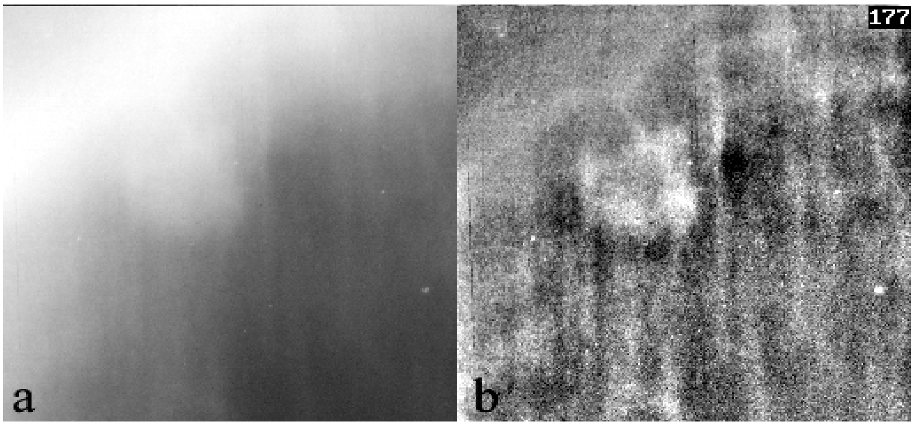

Fig. 2. Malignant nodule: a) A fragment of the original tomogram with a shadow of nodule. b) Processed tomograms with enhanced imaging of nodule features. After processing one can see non-round nodule with non-smooth irregular margin. There are irregular areas of varied densities inside the nodule. Multiple shadows of vessels are at the nodule, converging on it.

3.5 System for the Data Mining

For the data mining experiment we used the tool DECISION_MASTER, developed in the Institute for Computer Vision and applied Computer Sciences [20] (Leipzig, Germany). The tool is written in C++. It has a user-friendly interface and can be used easily by non-specialists in computer science.

The developed tool can create binary and n-ary decision trees from the data presented in the dBase-file. It has several options allowing one to specify how numerical features should be partitioned [22] and what method should be used for feature selection. An evaluation of the results can be done by test-and-train and n-fold cross-validation [23]. Missed values can be handled by different strategies [24]. The tool also provides functions for outlier detections.

For our experiment, we used a decision tree induction method, which created binary trees based on maximum entropy criteria [25]. Missing values were handled by replacing them with the mean value of the corresponding class [24]. Pruning was done based on the reduced-error pruning technique [26]. We applied a 10-fold cross-validation for the evaluation. Besides the error rate we calculated sensitivity for class_1 and specificity for class_2, which are error criteria usual for medical applications:

where S_{c1m} is a number of correctly classified samples in Class 1 and N_{C1} is a

$$E_{sens} = S_{c1m}/N_{c1} \qquad E_{Spec} = S_{C2m}/N_{C2}$$

number of all samples in Class 1 and S_{C2m} and N_{C2} in Class 2 respectively.

When the decision model (diagnosis knowledge) is instructed, the rules are provided either in TXT format for further use in the expert system, or the expert can use the learnt knowledge in the diagnosis unit of decision master for interactive work.

The architecture of the picture archiving system combined with the data-mining tool is shown in Fig. 3a [27]. The screenshot of our data mining tool decision master is shown in Fig. 3b.

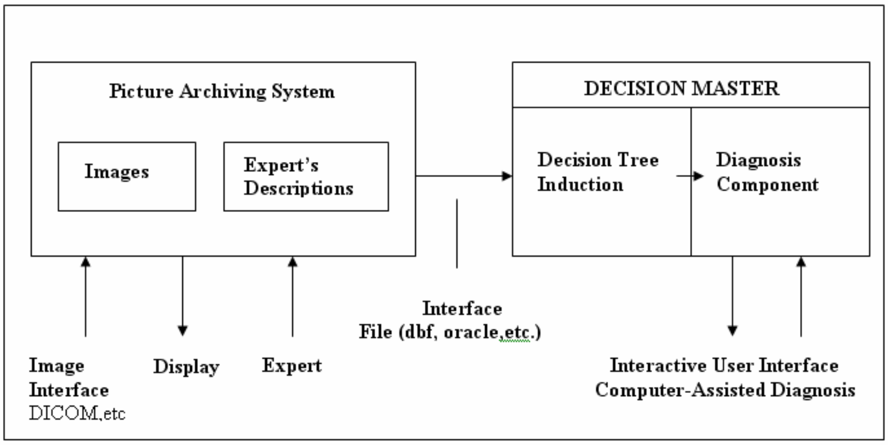

Fig. 3a. Architecture of a picture archiving system combined with the data mining tool

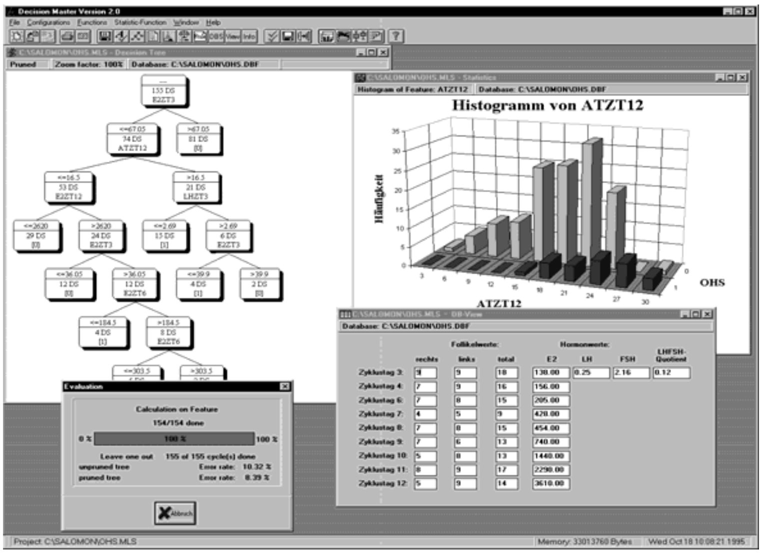

Fig. 3b. Screenshot of the data mining tool decision master

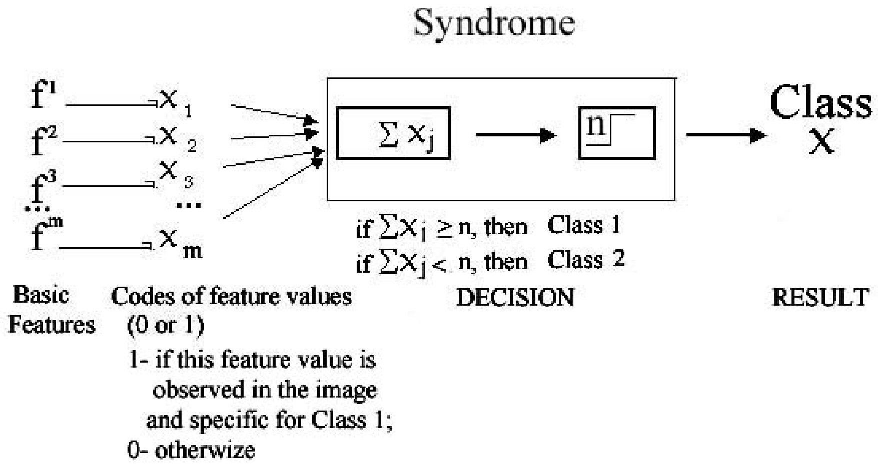

Fig. 4. Scheme of syndrome decision model

3.6. Semiautomatic knowledge aquistion Syndrome-Framework Tool

We compared the performance of the induced tree classifier to the performance of the syndrome classifier developed with the use of the syndrome–framework tool to learn diagnosis knowledge [21]. On Fig. 1 (left monitor) there is a syndrome–framework tool combined with the database and the system for image processing. This tool allows the user to develop a decision model which belongs to a specific type of classifier: It is similar to the syndrome (decision rule), conventional for medicine (see Fig. 4). The syndrome is a variant of a neuron network which operates using a multilevel threshold voting scheme.

A syndrome-framework tool supports the use of a multivariate contingency table analysis in combination with the analysis of classification results to create the effective decision model. It can be created automatically or semi-automatically (under the control of expert knowledge). In this case an knowledge-guided data analysis is organized as an interactive process with the participation of a physician. The expert controls the results of automated feature analysis: He/she checks what features and what feature values have been selected automatically as basic. He/she can correct this list according to his/her knowledge so that the new list (and discriminative values) “make sense” from his/her point of view. The obtained list has some redundancy in knowledge for improving the reliability of the decision-making model. The expert observes classification results using an analog of ROC curves displayed on a TV monitor for both classes [28]. The expert can change the threshold value according to his/her specific purpose and priorities, for example he/she selects the threshold to get a higher number of correct decisions for cancer in the score of “false alarms” for benign nodules, as it is less dangerous for the patient than missed cancer cases. (see Fig.5)

The Constructed network allows classification of the objects on the base of a fuzzy set of features, observed in the image. Being based on the voting principle, such a network is tolerant, to a large extent, to distortions, incompleteness, and uncertainty of input data. It doesn't need prior information about feature significance (or can use this information if it is available).

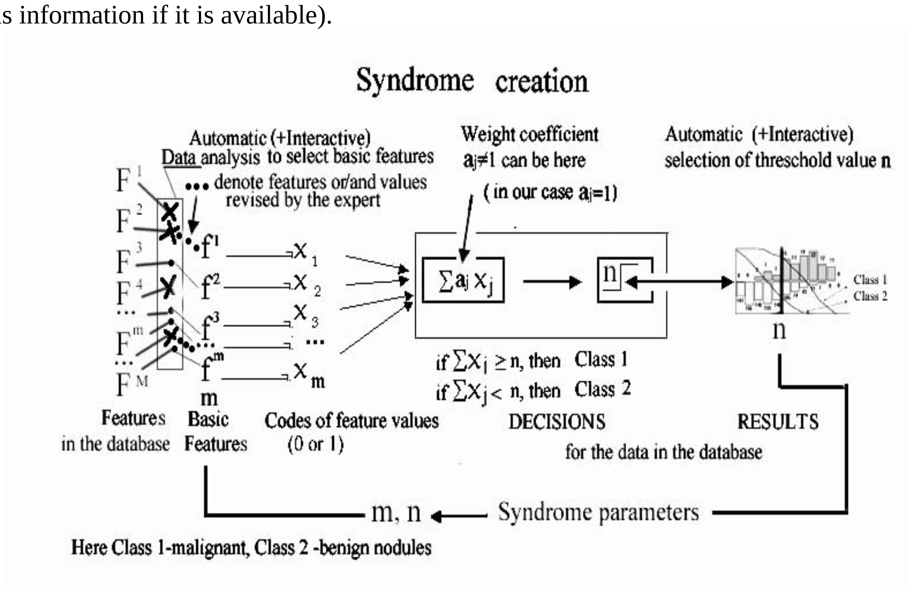

Fig. 5. Scheme of Syndrome creation. The decision rule (syndrome) is a set of feature conjunctions, based on a set of **m** binary features, selected from the initial list of features. The rule (syndrome) is fulfilled for the object if any set of **k** features (from **m** basic features) are observed (have value 1) for the object, where **k** $\in$[**n**, **m**], and **n** is a determined threshold. The tuning syndrome parameters [**n**, **m**] are fulfilled automatically or under the control of expert knowledge to obtain the most redundant and efficient decision rule.

4 Results

4.1 Efficacy of Classification for the Attribute List (1)

The induced tree for the Attribute list (1) is shown in Fig. 6 (the tool DECISION_MASTER actually shows the tree as a directed graph on the monitor).

The unpruned tree consists of 40 leaves. The pruned tree consists of 11 leaves (see Fig. 7). The 'expert' preferred the unpruned tree much more, because almost all attributes he used for decision-making appeared in this tree.

```
Scar-likeChan = 1
I       ConvofVes = 1 Class_1
I       ConvofVes = 2 Class_2
I       ConvofVes = 3
I               Margin = 1 Class_1
I               Margin = 2 Class_2
I               Margin = 3 Class_1
I               ...
I               Margin = 15 Class_1
Scar-likeChan = 2
I       Structure = 1 Class_1
I       Structure = 3 Class_2
I       ...
I       Structure = 6 Class_2
I       Structure = 7 Class_1
I       I       OutgoShadinTis = 1 Class_2
I       I       ...
I       I       OutgoShadinTis = 3 Class_2
I       I       OutgoShadinTis = 4 Class_1
Scar-likeChan = 3
I       Structure = 1 Class_1
I       Structure = 2 Class_1
I       Structure = 3 Class_2
I       ...
I       Structure = 6 Class_2
I       Structure = 7 Class_1
```

Fig. 6. Decision tree (1)

```
Scar-likeChan = 1 Class_1
Scar-likeChan = 2 Class_2
Scar-likeChan = 3
I       Structure = 1 Class_1
I       Structure = 2 Class_1
I       Structure = 3 Class_2
I       ...
I       Structure = 6 Class_2
I       Structure = 7 Class_1
```

Fig. 7. Pruned tree (2)

The 'expert' confirmed that the attribute *structure*, selected by the decision tree, is important for image analysis, and so is the attribute *scar-like changes inside the nodule*. However, he wondered why other important features, such as s*hape* and some others didn't work for classification. The 'expert' told us that he preferred to start nodule analysis from its s*tructure* and s*car-like changes inside the nodule*, then s*hape* and *margin*, and, finally*, convergence of vessels* and *outgoing* s*hadow in surrounding* *tissues*. According to his opinion, these features are also very important for his final decision about the nodule nature. So, although our decision trees represented the decision in the form convenient for the final user, it did not follow the strategy of the expert.

Therefore, we looked for the error rate and used it as the main criterion of DT fitness (table 3).

Table 3. Efficacy of classification for attribute List (1) on data set 1 and test set 2. Error rate an the set data and estimated error for cross-validation on sets.
(1) Evaluation on the data from set 1 (250 cases)
(2) Evaluation of decision tree on test data set 2 (38 cases)

Experiment	Before pruning	After pruning	
	Error	Error	Estimated error
(1)	1,0 %	1,0 %	11,2 %
(2)	11,4 %	14,4 %	11,2 %

As it is evident from table 4, we did not achieve the expert's diagnostic results. The reason might be in the choice of attribute values. During the tree-building process, the attribute appearing first in the database and satisfying the splitting criteria, is always chosen. For some categorical attributes, there were too many categorical values in the attribute list (1). That is why during the tree building process the training set was split up into very many subsets, which had a few data samples. The tree building process stopped very soon, since the remaining data set contained no sample which could be partitioned anymore.

Table 4. Comparison of the expert and decision tree (DT) classification for attribute list 1 for set 2.

Accuracy		*Sensitivity(Accuracy) for* Class_1		*Specificity(Accuracy) for* Class_2	
Human	DT	Human	DT	Human	DT
94,7 %	88,6 %	97,2 %	88,8 %	91 %	88 %

Since the attributes are nominal, we cannot find an ordering on attribute values, therefore we cannot summarize the values to obtain one more general attribute value.

For example, if one has an attribute *intensity* with attribute values "*black, dark gray, gray, light gray, white*", it is possible to generalize the values *dark gray, gray and ligh*t *gray* into gray.

Shapiro proposed the approach [29] to reduce a totally opaque, large decision tree to a hierarchy of several small decision trees, each of which "made sense" to the expert. We have chosen the way of constructing a new attribute list, which allowed us to realize Shapiro's idea.

First, we can use the generalized attribute value for the tree building processes. If we find in the induced tree that further distinction between the attribute values is necessary, we can carry out another induction experiment based on the specialized attribute values, starting with the data set, which corresponds to the leaf of the tree with a generalized attribute value.

4.2 Efficacy of Classification for the Attribute List (2)

Table 5. Efficacy of classification for attribute list (2) on data set 1 and test set2. Error rate an the set data and estimated error for cross-validation on sets

(1) Evaluation by? on data set 1 (250 cases)

(2) Evaluation of decision tree on test data set 2 (38 cases)

Result	Before Pruning	After Pruning	
	Error	Error	Estimated error
(1)	1,3%	1,3%	8,0%
(2)	4,2%	4,2%	8,0%

The initial attributes which had numerous attribute values were replaced by a set of attributes (sub-attributes), where each sub-attribute had a lower number of attribute values -attribute list (2), (see table 2).

In order to make sure that we did not develop many redundant and highly correlated attributes, we checked the reliability of features by calculating a proximity matrix based on Kruskal's tau [30], using a new data base created with the use of a new attribute list (Tab. 2). We grouped the set of features into functional groups based on an average link hierarchical clustering method [31]. We have found high correlation only between the attributes Charlung and Withlupl. For all other attributes, we were satisfied with the result. The resulting decision tree for data, collected according to the attribute list (2) is presented in Fig.8. This tree performed better than the first decision tree (compare error rates in table 3 and table 5).

However, from the expert's point of view, this decision tree was difficult to interpret, as there were not enough attributes there.

```
Angmar = 0
I    Spicmar = 1 Class_1
I    Spicmar = 0
I    I       Invsourtis = 1 Class_1
I    I       Invsourtis = =
I    I       I       Scarinsnod = 0
Class_2
I    I       I       Scarincnod = 1
Class_2
I    I       I       Scarincnod = 2
Class_1
Angmar = 1
I    Siofnod >= 0.5 Class_1
I    Siofnod <= 0.5
I    I       Regstrins = 0 Class_1
I    I       Regstrins = 1 Class_1
I    I       Regstrins = 2 Class_2
I    I       Regstrins = 3 Class_1
```

Fig. 8. Decision tree (3)

4.3 Classification Results based on the Test Data

Therefore, it was interesting to see how the tree performed on the data of set 1 and test set 2. In table 6 there are results of the DT classifier on the data, obtained by the 'expert' and by the 'novice'. Our experiments showed that the performance of the decision tree based on the attribute list 2 was better in comparison with those of the high level expert (see table 6). The resulting error rate showed that the classifier gave a reliable error rate even for very noisy and incomplete data presented by the 'novice' (see experiment (2) with the 'novice' in table 6).

Table 6. Comparison between expert decisions and classification of decision tree (DT): (1) High-level expert; (2) Middle-level expert (novice).

Users	*Accuracy*		*Sensitivity*		*Specificity*	
	Class_1	**Class_2**	**Class_1**		**Class_2**	
	Human	**DT**	**Human**	**DT**	**Human**	**DT**
(1)	94,5%	95,8%	96,2%	93,75%	90%	100%
(2)	55,2%	73%	61,1%	75%	50%	72%

Table 7. A list of features, used for decision rule creation ("***V***" marks the feature used in syndrome).

No. of Attribute	**Attribute type**	**Attribute Name**	**Short Attribute Name**	**No. of Values**	**Attribute Value**	$\mathbf{m_1}$	$\mathbf{m_2}$
CLASS	Boolean	Class	Class	1 2	Malignant Benign	**Classifying feature**	
1	Categorical	Structure inside the module	StrInsNod	1 2	Irregular Regular orderly	***V***	-
2	Categorical	Scar-like changes inside the nodule	ScrLikeChan	0 1 2	None May be existing Irregularly fragmentary dense shadow	***V***	***V***
3	Categorical	Shape	Shape	1 2 3	Nonround Round Oval	***V***	***V***
4	Categorical	Sharpness of Margin	SharpMar	1 2 3	Nonsharp Mixedsharp Sharp	***V***	-
5	Categorical	Smoothness of Margin	SmoothMar	1 2 3	Nonsmooth Mixedsmooth Smooth	***V***	***V***
6	Boolean	Angularity of Margin	AngMar	0 1	Nonangular Angular	***V***	***V***
7	Boolean	Spicularity of Margin	SpicMar	0 1	Nonspicular Spicular	***V***	***V***
8	Categorical	Convergence of vessels	ConvofVes	1 2	Vessels have forced away the nodule Vessels are converging to the nodule	***V***	***V***
9	Boolean	Vascular Outgoing Shadow	Vascshad	0 1	None Chiefly vascular shadows	***V***	***V***
10	Boolean	Outgoing shadows: Invasion into surrounding tissues	OutgoShadinTis	0 1	None Invasion into surrounding tissues		
11	Boolean	Thickening of lung pleura	ThLungPl	0 1	None Thickening	***V***	-
12	Boolean	Withdrawing of lung pleura	WithLupl	0 1	None Withdrawing	***V***	-

4.4 Classification Results based on the Syndrome Classifier

A decision rule was constructed on the basis of data obtained according to the attribute list (2).

The initial list of features contained N=25 features proposed by the expert. In table 7 there is a list of specific features selected from the initial attribute list as a result of statistical data analysis in the database and added by the expert who selected some additional features which were diagnostically important from his point of view [32]. These features were then used for decision-rule creation. Two syndromes with different numbers **m** of selected basic features and threshold values **n** were constructed and tested [28][32]**: Syndrome 1** ($\mathbf{m_1}$=12, $\mathbf{n_1}$=4), and **Syndrome 2** **(**$\mathbf{m_2}$=8, $\mathbf{n_2}$=3). **A** multivariate contingency-table analysis [33]was applied for this purpose. The syndrome referred the sample to the class 1 if any set of **k** from **m** features was observed (had value 1) for this object, where $\mathbf{n \le k \le m}$. (e.g. we must observe in the images at least **n** features which have values specific for class1). The total number of such sub-rules in the rule is equal to the sum of all possible combinations $\mathbf{C^k_m}$ of **k** features from **m** for all **k,** which satisfies the indicated condition $\mathbf{n \le k \le m}$

A set of basic features **m,** and a threshold value **n** (**m** and n -syndrome parameters) were determined, so that the resulting decision model gave the highest efficacy and stability on the data in the database. Both syndromes showed the same accuracy on data from set 1. Both constructed syndromes were applied for classification of test set 2. Our experiments showed syndrome 1 worked better with noisy and incomplete data from test set (2). So it was selected for our final experiments and comparisons with the decision tree results (see table 8).

Table 8. Comparison results of syndrome-classifier and decision tree classification on the database created with use of the attribute list (2): (1) Data readings from high-level expert, (2) Data readings of non-trained novice

Accuracy			***Sensitivity***			***Specificity***		
Radiologist's **D**iagnosis	**S**yndrome **C**lassifier	**D**ecision **T**ree	**Class_1**			**Class_2**		
RD	**SC**	**DT**	**RD**	**SC**	**DT**	**RD**	**SC**	**DT**
(1) 94,5%	95,6%	95,8%	97.2%	98,4%	93,7%	91%	91,7%	100%
(2) 55,2%	84,2%	73,0%	61,1%	78,0%	75,0%	50%	90%	72%

The developed decision rule presented effective strategies for image analysis and interpretation, orienting the user to look for specific features, and also showing how to classify the image on the basis of observed features. The rule can easily be presented in the form of a/the syndrome in the language that is natural for the expert's domain knowledge. Such a rule can easily be learned by the physician, applied in clinical practice or used for teaching novices.

5. Comparison of Data Mining with the Semiautomatic Knowledge Acquisition Method by the Syndrome–Framework Tool

The test based on image readings from a high-level expert shows that the decision tree method performs better than the syndrome (table 8). One can see that both classifiers exceed the results obtained by the 'expert'. Only in the case of the 'novice', whose data contained many missed and distorted (noisy) values, did the syndrome perform better. Our investigation showed that in the decision tree classifier a misclassification of samples is mostly based on incorrectly chosen attributes, not on the missed attribute value, since the decision tree classifier is based only on a few attributes that are the most important and always appear in the image.

Syndrome1 uses 12 features and syndrome2 uses 8 features. If we compare the set of features used by the decision tree with those used by the syndrome, we can see a large overlap among the two feature sets. However, the decision tree induction method tries to find a minimal set of features necessary for a decision. The feature 'redundancy' in the syndrome classifier might be a reason that it works so well for uncertain cases. Beyond that the decision tree was trained based on a data set created by a top-level expert. Therefore, the decision tree cannot have a strategy to deal with the reasoning behavior of a middle-level expert.

In decision tree induction we cannot easily update the tree without rebuilding the entire tree. However, there is some research going on for incremental decision tree induction [33], but that is not the subject of this paper.

It is interesting to note that the decision-tree and the syndrome–like decision rule allow one to find basic diagnostically important features. They both can support decision-making on the basis of feature reading. It means that they give the user decision models which help to decompose a complex pattern recognition task, which usually needs specific domain knowledge and long training, into a set of simpler

subtasks connected with an analysis of the object features (in our case these were: shape, margins, homogeneous area, inhomogeneous area, etc.). Most of these tasks are under consideration in the framework of the computer-vision approach. Feature computing can help to measure necessary feature values directly in the image, thus producing expert-independent results of image analysis. Our next steps will be to automate the estimation of object shapes and margins.

6. Conclusion and 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 large database with images and expert descriptions. Such databases result from the common use of picture-archiving systems in medicine.

Applying the data-mining method to the database of image descriptions, we could learn the important attributes necessary for an image analysis. Acquiring basic features and developing effective decision models, such methods could help to achieve the goals of purpose vision. It could help to decompose a complex visual task into several simpler subtasks, present effective strategies for image analysis and interpretation, show the way for development of unbiased expert-independent methods of measuring some basic features directly in the image. They can support image analysis and interpretation by the expert in the case of uncertainty.

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.

Experiments have demonstrated the attributes included automatically into the tree and represented by the expert knowledge.

We compared the decision tree and the syndrome-like classifier. The error rate for the decision tree classifier was better than the error rate of the syndrome for the data obtained from the expert. Only for handling very uncertain data was the performance of the decision tree worse than that of the syndrome. We believe that special search strategies for the decision tree method during the decision making process could improve the decision tree results. However, this is left for further research.

The accuracy of a peripheral lung cancer diagnosis with the use of the developed decision rules is close to the diagnostic results of a highly qualified expert and can improve the number of correct diagnoses for physicians of different qualifications.

The methodology, developed in this task, can be used for knowledge acquisition in other medical tasks, in particular we applied it for a/the mammogram diagnosis in our recent work [8].

Finally, we can say that picture archiving systems in combination with data mining methods open the possibility for an advanced computer-assisted medical diagnosis. However, it will not give the expected result if pictures and expert descriptions are not stored in a standard format in the archives (e.g. in PACS) for further analysis. Since standard vocabulary (8) and very good experts are available for many medical diagnostic tasks, data mining methods can be used for mining knowledge that could support image analysis and classification in the case of uncertainty. If the vocabulary is not a priori available, it can be elaborated, using for example repertory grid methodology, described in [10].

[1] In: Proceedings of 15th Symposium for Computer Applications in Radiology: Filmless radiology –reengineering the practice of Radiology for the 21st Century. J. of Digital Imaging 11 (3), Suppl. 3 (1998)

[2] Andriole, K.P.: Anatomy of Picture archiving and communication systems: Nuts and Bolts-Image Acquisition: Getting digital images from imaging modalities. J. of Digital Imaging; 12 (2), Suppl. 2, 216-217 (2000)

[3] In: Proceedings of 16th Symposium for Computer Applications in Radiology. PACS: Performance Improvement in Radiology. J. of Digital Imaging; 12 (2), Suppl. 2, (1999)

[4] In: Proceedings of 17th Symposium for Computer Applications in Radiology: The electronic Practice: Radiology and Enterprise. J. of Digital Imaging 13 (2), Suppl. 1 (2000)

[5] In: Proceedings of SPIE International Symposium Medical Imaging 1998, San-Diego, USA, 1998 , SPIE, vol. 3338 (1998)

[6] In:Proceedings of SPIE International Symposium Medical Imaging 2000, San-Diego, USA, 2000, SPIE, vol. 3981. (2000)

[7] In: Proceedings of 14th Int. Congress on Assisted Radiology and Surgery –CARS’2000, San-Francisco, USA, 2000, Excerpta Medica Intern.Congress Series, vol. 1214 (2000)

[8] Heywang-Köbrunner, S., Perner, P.: Optimized Computer-Assisted Diagnosis based on Data Mining, Expert Knowledge and Histological Verification. IBaI Report ISSN 1431-2360 (2000)
[9] Perner, P.: A knowledge-based image inspection system for automatic defect recognition, classification, and process diagnosis. Int. Journal on Machine Vision and Applications 7, 135-147 (1994)
[10] Boose, J.H., Shema, D.B., Bradshaw, J.M.: Recent progress in Aquinas: a knowledge acquisition workbench. Knowledge Acquisition 1, 185-214 (1989)
[11] Kehoe, A., Parker, G.A.: An IKB defect classification system for automated industrial radiographic inspection. IEEE Expert Systems 8, 149-157 (1991)
[12] Schröder, S., Niemann, H., Sagerer, G.: Knowledge acquisition for a knowledge based image analysis system. In: Bosse, J., Gaines, B. (Eds.) Proc. of the European Knowledge-Acquisition Workshop (EKAW 88), GMD-Studien 143, Sankt Augustin (1988)
[13] Kolodner, J.L., Simpson, R.L., Sycara, K.: A Process Model of Case–Based Reasoning in Problem Solving. In: “Proc. 9th Int. Joint conf. on Artificial Intelligence”, pp. 110-110, Los Angeles (1985)
[14] Perner, P.: Case-Based Reasoning for the Low-level and High-level Unit of an Image Interpretation System. In: Singh, S. (Ed.) Advances in Pattern Recognition, pp. 45-55, Springer Verlag, London (1998)
[15] Belikova, T.P.: Synthesis of optimal linear filters for visualization of diagnostically important objects in medical radiological tasks. In: Digital Optics in Medical Introscopy, pp. 57-72, Science, Moscow (1992) (in Russian)
[16] Barrett, H.H., Swindell, W.: Radiological Imaging: Theory of image formation, detection and processing. Academic Press, San Diego (1996)
[17] Belikova, T.P., Yashunskaya, N.I., Kogan, E.A.: Computer-Aided differential Diagnosis of Small solitary Pulmonary Nodules. Computer and Biomedical Research 29(1), 48-62 (1996)
[18] Belikova, T.P., Yashunskaya, N.I., Kogan, E.A.: Computer analysis for differential diagnosis of small pulmonary nodules. In Proc. of Int. Congress for lung cancer. Athens Greece. pp. 93-98, Monduzzi Editore (1994)
[19] Perner. P., Belikova, T.P., Yashunskaya, N.I.: Knowledge acquisition by symbolic decision tree induction for interpretation of digital images in radiology. In: Perner, P., Wang, P., Rosenfeld, A. (Eds.) Advances in Structural and Syntactical Pattern Recognition. LNAI, vol. 1121, pp. 208-219, Springer-Verlag, Heidelberg (1996).
[20] Data Mining Tool Decision Master, Description available at http://www.ibai-solutions.de
[21] Pereverzev-Orlov, V.S.: A Partner System and the Concept of Recognition Learning. Pattern Recognition and Image Analysis 2, 420-437 (1992)
[22] Perner, P., Trautzsch, S.: Multi interval Discretization for Decision Tree Learning. In: Amin, A., Dori, D., Pudil, P., Freeman, H. (Eds) Advances in Pattern Recognition. LNCS, vol. 1451, pp. 475-482, Springer-Verlag, Heidelberg (1998)
[23] Weiss, S.M, Kulikowski, C.: Computer Systems that learn. Morgan Kaufmann (1991)
[24] Perner, P.: Data Mining for Multimedia Data. LNCS, vol. 2258, Springer Verlag, Heidelberg (2002)
[25] Baird, H.S., Mallows, C.L.: Bounded-Error in Pre-classification Trees. In: Dori, D., Bruckstein, A. (Eds.) Shape, Structure and Pattern Recognition. pp. 100-110, World Scientific Publishing Co. (1995)
[26] Quinlain, J.R.: Simplifying decision tree. Intern. Journal on Man-Machine Studies **27**, 221-234 (1987)
[27] Perner, P.: Mining Knowledge in Medical Image Databases. In: Dasarathy, B.V. (Ed.) Data Mining and Knowledge Discovery: Theory, Tools, and Technology. SPIE 4057, 359-369 (2000)

[28] Belikova, T.P., Stenina, I.I., Yashunskaya, N.I.: Computer-aided methods to recover strategies for visual search and navigation. In: SPIE International Symposium Medical Imaging 2000. SPIE 3981, 240-250 (2000)
[29] Shapiro, A.: Structured Induction. John Wiley Inc. (1985)
[30] Agresti, A.: Categorical Data Analysis. John Wiley Inc. (1990)
[31] Jain, A.K., Dubes, R.: Algorithms for Clustering Data. Englewood Cliffs, Prentice Hall (1988)
[32] Belikova, T.P., Stenina, I.I., Yashunskaya, N.I.: Image interpretation based on image processing and knowledge-guided data analysis. J. Machine Graphics & Vision, 17 (1/2), 187-196 (1998)
[33] Upton, G.: The Analysis of Cross-tabulated Data. J. Wiley & Sons (1978)
[34] Utgoff, P.E.: Incremental Induction of Decision Trees. Machine Learning 4, 161-186 (1989)