Combining image content and annotated text for medical image categorization and retrieval

Filip Florea, Olivier François, Eugen Barbu, Alexandrina Rogozan, and Abdelaziz Bensrhair

LITIS Laboratory, 1060 Avenue de l'Université
F-76801 St. Etienne du Rouvray Cedex, France

Abstract. The richness of health-information available on-line requires the development of efficient information retrieval methods. The CISMeF heath-catalogue provides indexing and searching capabilities for health-resources. Medical images are representing a significant part of on-line medical knowledge and a valuable component of diagnosis and teaching. In this context, a combined text and image extraction approach is desired for the indexing and retrieval of on-line medical information. This paper presents the capabilities of the MedIC system of automatically extract image information from the image visual content or the image annotated text, as well as the current architecture for combining the decisions issued at each of these levels, using Bayesian Networks. The results we are presenting are showing that combining these decisions can create a better result. Extracting further image-related information from the documents where the images are placed, will further test the this information fusion approach, and ultimately enrich the MedIC image categorization module, of the CISMeF catalogue.

1 Introduction

With the development of the Internet and its increasing importance as a major source of information, the creation of tools and methods for accurately and easily find medical information on the web becomes indispensable. In this context, several tools for indexing and retrieval of health information were proposed and are available on the Internet, like MedHUNT[1], MedWebPlus[2], CliniWeb[3] and CISMeF[4].

CISMeF (French acronym for Catalog and Index of French-language health resources) is a quality-controlled subject gateway [1] initiated by the Rouen University Hospital in 1995 [2]. Its role is to provide online searching capabilities for health resources (i.e. documents), by describing and indexing the most important documents of institutional health information in French (language). From the beginning CISMeF is indexed manually by a team of experienced health-librarians.

[1] http://www.hon.ch/
[2] http://www.medwebplus.com/
[3] http://www.ohsu.edu/cliniweb/
[4] http://www.cismef.org/

Considerable efforts were undertake by the CISMeF team to develop automatic textual indexation architectures, and recently significant advancements were reported [3]. Automated indexing has drawbacks thought, and one of the biggest is the difficulty of indexing non-text media, like the images.

Medical imaging has grown over the last decade to become an essential component of diagnosis, medical teaching and civic education. The development of Internet technologies has made medical images available in large numbers in online repositories, collections, atlases, and other heath-related resources. These images are representing a valuable source of knowledge and are of significant importance for medical information retrieval. Unfortunately, the shear amount of medical visual data available online makes it very difficult for users to find exactly the images that they are searching for. The annotation of images with relevant image-related medical keywords could provide the users with the possibility of using textual queries to search medical images. However, the cost of manually annotating images is prohibitively high as it is time-consuming and requires medical knowledge. To provide efficient and fast access to medical visual-data, automatic systems for extracting relevant medical information from images are needed.

Developed by the CISMeF team, the MedIC (Medical Image Categorization) system has the goal of adding a ”search-by-image” capability to the Doc'CISMeF search engine. Thus, the aim is to allow the users (i.e. health professionals, students or general public) to specify additional image-related terms (in addition to the text-related terms currently used), when performing queries.

The current architecture of the MedIC system is designed to extract five types of image-related medical information:

1/. medical modality
2/. anatomical region, biological system and/or organ under observation
3/. acquisition view-angle
4/. diagnostic information: pathology
5/. other acquisition parameters: resolution, field of view, side, slice number, contrast substances used

Once extracted, this information is to be used when the CISMeF resource containing the image(s) is indexed. For example, if an image which MedIC recognized as 1/. = ”RX”, 2/. = ”upper-leg”, 3/. = ”coronal/frontal plane”, 4/. = ”pertrochanteric fracture”, 5/. = ”left side”, is part of a medical document that CISMeF is going to index, then the indexer will add that the document is containing additional resources (e.g. additional RT Resource Type = JPG) with the description provided by MedIC. Adding this information to the index of CISMeF resources would be very useful to allow image-oriented queries (e.g. <Find me all the resources containing RX images>) or to add image-related criteria to the initial query (e.g. <Find me the resources related to ORTOPEDICS> and <containing RX images of LEG FRACTURES>”).

These types of information are representing different concepts and are occurring in different places in images or the documents that are containing the images. The MedIC architecture allows the extraction of information at several levels:

a/. from the image visual content
b/. from the textual annotations marked directly on the image
c/. from the image-related text-regions of the document (e.g. image caption, image name, the paragraph/sentence that points to the caption number)

At each of these levels, different information could be expected to be present. For example, from the image itself (a/.), using image representation spaces and machine learning, information like (1/.) (2/.) and (3/.) can be extracted while (4/.) and especially (5/.) are very difficult to obtain.

To combine these five possibly very rich sources of complex information can represent a challenge especially when contradictory information are extracted at different levels (a/.-c/.) with different probabilities. In this paper we are presenting the current approaches for combining the information extracted by MedIC for an accurate medical image categorization. For these tests we considered the extraction of the most accurate modality decision (1/.) taking into consideration the image's visual content and the textual annotations presented on the images.

In this paper we are presenting the current approaches of medical information fusion for image categorization. For these tests we considered the extraction of the most accurate modality decision (1/.) taking into consideration the image's visual content and the textual annotations presented on the images.

The rest of this paper is organized as follows. Section 2 will describe some of the related work. In Section 3 the materials and methods used for these experiments are presented, with the presentation of image database, the description of the two information extraction approaches and the fusion architecture we are proposing. The obtained results are presented in section and finally the discussions, conclusions and perspectives in Section 5.

2 Related work

There are a number of systems that are treating the problem of ”automatic extraction of image-related information” in the form of medical image categorization and retrieval applications. These systems are usually designed for retrieval inside a given modality: KMeD [4] is treating MRI head images, ASSERT-system deals with lung CT images [5], I-Browse [6] operates on histological slices and [7] with X-Rays. Given the fact that the principles used by each of these systems are highly-dependent of the particular conditions of each medical modality, they are not directly and/or entirely applicable to other cases. However systems better adapted to cope with various image modalities, anatomical regions and pathologies were proposed more recently: MedGIFT [8] and IRMA [9].

The systems specifically designed for automatic medical image analysis (indexing, categorization and/or retrieval) are coping well with medical images but usually in off-line, image-only contexts. To our knowledge, none of these systems are adapted to deal and efficiently use medical images placed in rich on-line resources or in other environments where text descriptions were available along with the images. Not having to deal with several sources (e.g. image content, image annotation, image-related text paragraph) of (possibly) the same information, the existing image categorization or indexing systems did not need to consider decision fusion in their architectures.

In the context of CISMeF catalogue, the MedIC system uses several sources (a/.) - (b/.) and several criteria (i.e. categories) (1/.)-(5/.) for categorizing medical images. Each of the sources is providing information on some of the 5 categories. This will produce very uneven (and possibly contradictory) decision sets. Extensive tests were done on the first two of the image-related information sources (a/. and b/.) for extracting as much as possible of the categories (1/.-5/.). Promising results were noted for some of the categories when considering the sources independently (see results at 3.2 and 3.3). To efficiently combine these decisions, we are presenting in this paper the current information fusion approaches we are considering for medical image categorization.

3 Material and Methods

3.1 Medical image database

The image database used for our experiments consists of 10317 anonymous images extracted part from the Rouen University Hospital clinical file and part from web-resources indexed in CISMeF. For the tests presented in this paper, we considered the main 6 categories of medical-imaging modalities: angiography, ultra-sonography (US), magnetic resonance imaging (MRI), standard radiography (RX), computer tomography (CT) and scintigraphy, each containing a number of anatomical (e.g. head, thorax, lower-leg), sub-anatomical regions (e.g. knee, tibia, ankle) and acquisition views-angles (coronal, axial, sagittal).

Medical		no. of images	
modality		absolute	relative
Angio	(C1)	280	2.71%
US	(C2)	1046	10.14%
MRI	(C3)	3806	36.89%
RX	(C4)	2588	25.08%
CT	(C5)	2387	23,14%
Scinti	(C6)	210	2,04%

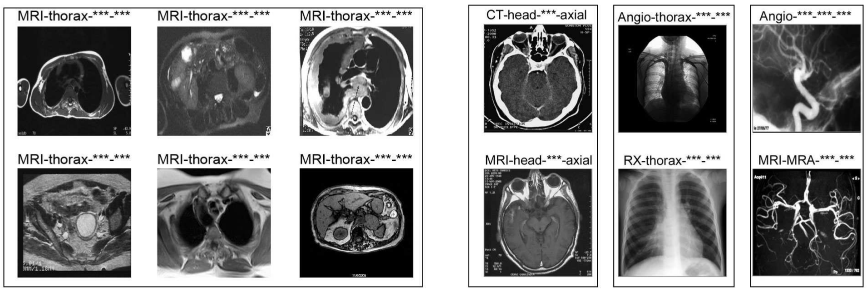
Table 1. The distribution of the 6 modalities

Table 1 shows the distribution of the 6 modalities in the database. We used this image database for extracting the most reliable medical modality (1/.) from both the image visual content (a/.) and the textual annotations marked directly on images (b/.), as we shall see in the next two sections.

3.2 The image visual content (a/.)

The image itself contains a significant amount of information. This information is usually very difficult to interpret (e.g. diagnostic) and entirely capture in words (manually annotate), as it is very complex and requires domain knowledge and experience. Even though pathological decision (4/.) or acquisition parameters (5/.) are more difficult to extract directly from the visual content of medical images, experiments show that information like the modality (1/.), the anatomical region (2/.) and the acquisition view-angle (3/.) can be accurately extracted using proper image representations and machine learning.

The main difficulty is represented by the high image variability caused by the various image sources and the additional processing necessary before publishing the images on Internet. We noted strong variability in variability in size, compression ratio, contrast, background, addition of superposed didactical annotations and drawings. Not being able to control the parameters and quality of the images posted on web-resources, and thus to reduce this supplemental variability, makes dealing with on-line published images even more difficult. We are, thus faced with an increased intra-class variability that should negatively influence the overall classification result (see Fig. 1(a)). The difficulty is further increased by the strong inter-class similarity between some classes, representing different modalities and/or anatomical regions (see Fig. 1(b))



(a) Intra-class variability.

(b) Inter-class similarity.

Fig. 1. Database difficulty.

To automatically extract image categories we first defined category classes: modalities, anatomical regions and view angles (as presented in 3.1). The images were represented using statistical and texture features from local representations of the 256x256 downscaled images [10]. Then we applied dimensionality reduction and classification algorithms (e.g. *k-NN*, *SVM*) to project the test images in the previously defined classes.

In spite of the significant image variability, using only image representation and machine learning techniques, an accuracy of 89.59% was obtained for modality categorization. This result was obtained training an *SVM* classifier (C=100, second degree polynomial kernel) on 8818 images and, then classifying the remaining 1499.

We aiming at improving this result using this approach combined with the decision issued at the next section.

3.3 Image annotations (b/.)

Nowadays images produced by modern medical acquisition equipments, are stored in DICOM, a standard for medical image production. Along with the image itself, DICOM stores a series of information on the production modality, anatomic regions, view angles, acquisition parameters, and others. Some of the acquisition parameters are presented directly superposed on the images, when these are consulted on diagnostic stations inside PACS (Picture archiving and Communication System - hospital specific networks dedicated to the storage, distribution and presentation of images).

To be posted on web pages, the images have to be exported in Internet specific formats, such as JPEG and GIF. Unfortunately, when this type of conversion is done, the images are loosing the text layer of the DICOM format, and with it, all the image annotation "marked" directly on the images. In the rare cases when these image annotations/markings are still present available they are containing valuable information (about acquisition parameters (5/.), but also about the modality (1/.) and anatomical region (2/.).

We conducted test on the same 10317 image database used at previous section. First the text marked on the images is extracted using morphology-based filtering (i.e. a multi-level *TopHat* combined with additional morphological operations to take into consideration the horizontal disposition of text in lines). A state-of-the-art Optical Character Recognition application was then used for the optical recognition of the textual annotations [13]. Given that the images are usually, of good quality, and thus the text is correctly extracted, the OCR recognizes the annotations properly.

These annotations can be used directly for image indexing but image retrieval using directly this kind of information is unlikely, given that the majority of these are technical acquisition parameters. However, by filtering and interpreting these annotations more useful image concepts could be derived, such as the medical modality. We thus proceed at extracting the most accurate modality decision using only the textual annotations (markings) present on images (if they exist).

For the interpretation of the recognized textual annotations, a set of medical modality production rules (the most pertinent and discriminative modality annotations) was defined by a medical specialist. A sample of the full set of 96 rules is presented in Table 2 (e.g. `TR`, `TE` and `TA` are typical annotations for MRIs, and stand for `Repetition Time`, `Exposure Time` and respectively `Acquisition Time` in French).

The extracted annotations for each image were classified using a *C4.5 Decision Tree (DT)* [11] trained on the 8818 learn images, and a performance of +99% modality recognition precision was obtained. The structure of the *DT* is presented in figure 2. We can observe the meaningful annotations that de tree

Textual	Meaning	Modality					
Markings		C1	C2	C3	C4	C5	C6
`TR, TE, TA`	repetition, exposure and acq. time			✓			
`NEX`	number of averages			✓			
`GADOLINIUM`	contrast agent			✓			
`Post CM`	injection			✓		✓	
`Tilt`	tipping angle					✓	
`FLTR`	filtre	✓					
`LNDMK`	-	✓					
`dB`	decibel		✓				
`GAIN`	-		✓				
`SonoCT`	-		✓				
`kV`	kilovolt					✓	
`mA`	miliampere					✓	
`MIBG`	injected isotope						✓
`DMSA`	”						✓
`THALLIUM`	”						✓

Table 2. An extract of the production rules

keept, and the number of error for each class. Also we observe the there are no decisions for the classes C4 (RX) and C6 (Scintigraphy). This is explained by the absence of any annotations for practically all the RX and Scintigraphy images presented in our database.

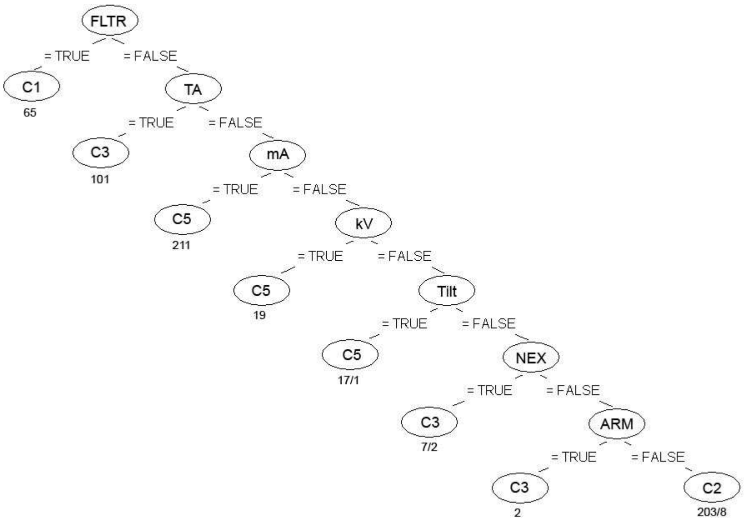


Fig. 2. Decision Tree

The good performances obtained by this approach shows that the annotations representing acquisition parameters are meaningful and should be used for modality decision when these annotations are present on images. Unfortunately, in reality, only 625 images from the total of 10317 had textual annotations. This is reflected by the small overall recall of this method (less than 10% for modality recognition). This makes this approach less interesting when used independently.

However the precision of this approach being very high (+98%) we can expect that combining the two approaches better results could be obtained.

3.4 Combining the decisions/Fusion architecture

As we could observe that the design of MedIC allows the extraction of image-related medical information from several sources (image, image annotation, text-regions). As a consequence of the system's multi-source approach, we can come across a situation where contradictory or incomplete decisions are issued from different sources. In these cases an additional fusion step is needed to combine the decisions in function of their reliability.

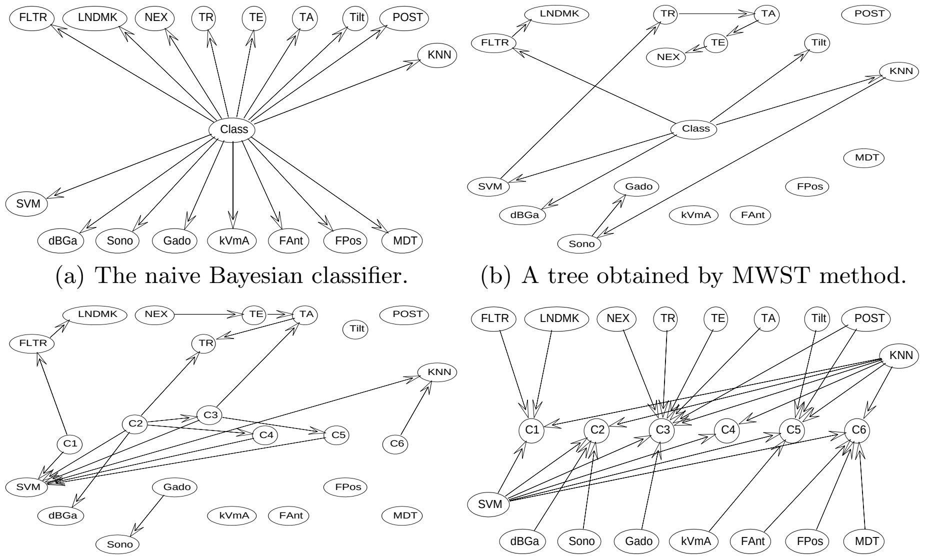
In the case of these experiments the aim was to combine the two obtained decisions, the modality decision obtained by *SVM* classifier (applied to the statistical and texture feature space) (a/.) and the decision issued from the decision tree annotation classification (b/.), to improve the performance of the modality categorization.

Giving that the second decision (b/.) is very accurate ($\sim$98% of modality recognition precision) we imagined a first approach in which we combined (a/.) and (b/.), by validating each time the prediction of (b/.) when this existed (when textual annotations are present). This fairly easy approach is already improving significantly the modality overall decision, up to 94.46% of accuracy being noted.

However, this simple approach is only viable when the textual annotations of the method (b/.) are influencing the resulting classes in a determinist manner and there are no confusions among different annotations. In reality there are a number of annotations that are specific to more than one modality or that even though specific to a certain modality, can be find (due to abbreviations, OCR errors ...) on images of different modalities. This makes some annotations more reliable than other, and groups of annotations stronger indicators of a certain category (modality in the experiments presented here). To model and learn all these probabilistic relations, we have considered to use the *Bayesian Network* formalism [12, 13].

Bayesian Networks (*BN*) aim at modelling systems by taking graphically into account conditional independences between variables (by means of a directed acyclic graph) and by giving a compact representation of the joint probability distribution as the product of local conditional probability distributions (one for each node in the graph).

We first we used a naive Bayesian classifier as this model has proved its efficiency in many fields [14]. The naive *BN* applied to our problem is illustrated



(a) The naive Bayesian classifier.

(b) A tree obtained by MWST method.

(c) A *BN* obtained by GES (close to ones found by K2) for 6 classe nodes.

(d) The model we have built.

Fig. 3. Some of the Bayesian Networks used for the classification task.

on figure 3(a). Additionally, we have used the Minimum Weighted Spanning Tree (MWST) [15] as this approach produces a *BN* structure that is a tree. Therefore the result could be viewed as a probabilistic extension of a Decision Tree. An instance of a MWST *BN* model is shown on figure 3(b). Then, we have tried others standard methods of structure learning as, for instance, K2 [16] or the Greedy Equivalent Search (GES) [17]. Finally, to designed a *BN* model specific to our problem which could be seen on figure 3(d). Once the model was created, we have learned its conditional probability tables. The model is created using the medical knowledge depicted form the production rules presented in 3.3.

All these experiments were done using the Structure Learning Package [18] for the Bayes Net Toolbox [19].

4 Results

The results obtained by the two individual approaches (a/.) and (b/.), as well as the simple decision are represented in Table 3

As we can observe, the results are already consistent for method (a/.), and rather high but not on the entire dataset for the method (b/.). Using the simple fusion approach allows to correct some of the errors obtained by the (a/.) method.

Method	Modality (1/.) categorization	% classified
(a/.)	89,59%	100%
(b/.)	97,28%	6,05%
simple (a/.)and(b/.)	94,46%	100%

Table 3. Individual decision and simple fusion

We then passed to test the different designs of *BN* we presented at 3.4. As the test dataset numbers only 1499 samples and as we want to evaluate the generalization error, we have chosen to use a k-fold cross-validation. Moreover to evaluate the robustness of learning technics, we have tested with $k = 3$ and with $k = 15$, which gave us learning datasets of different sizes.

Given that some features are specific to a given modality class, we have chosen to consider a new variable representing the logical OR of these features. For instance, we join together the gain variables (dB and GAIN → dBGA) of US modality, the electrical measures (kV and mA → kVmA) of CT modality and the type of isotope used in scintigraphy (MBIG, DMSA and THALLIUM → MDT). Thus, the remaining number of features is 18.

Moreover, to graphically evaluate the influence of features on classes, and to simplify the conditional probability table of the class node (as we have few samples to evaluate it), we decided to also test *BN* structure learning using different nodes for each of the 6 classes. Thus, the remaining number of features is 23. We choose these approach to model the *BN* structure designed specifically for our application (Fig. 3(d)).

Upon inspection of the models created by various *BN* approaches we had considered, we noted that in many cases not all the nodes were connected. In the classification phase only the variables that are connected to the class node(s) are used for the prediction. For instance, we can observe in figure 3(b)(c) that the nodes POST, kVmA, FAnt, FPos, MDT are almost always disconnected from the network structures. If the dataset were representative, we could have deduced that these features were not relevant. In our case, the most probable conclusion is that our dataset is not relevant enough to learn the influence of these features.

Cross-validation classification rate means are shown in Table 4.

		naive	MWST	K2	ges	model
C1	cv3	96.20	96.40	96.40	96.40	96.00
	cv15	96.73	96.73	96.60	96.65	96.97
C6	cv3	/	96.46	95.94	95.30	96.13
	cv15	/	96.46	95.90	95.58	96.97

Table 4. Percentages of good classification evaluated by cross-validation in three and fifteen folds.

We could see that all approaches obtained similar performances (around 96.3%). Nevertheless, we think that the model we have built could be more efficient on a more representative training dataset.

5 Conclusions

With the continuous development of Internet technologies, more and more medical information is becoming available on-line. Efficient and easy to use on-line tools capable of indexing and searching for medical resources using criteria related to both the text of the resource and the images that may appear are still lacking. The MedIC system was designed to function in such a bi-modal (text-image) context. When dealing with various sources of information, proper fusion architectures are needed to combine the decisions extracted at each of these levels.

We presented, in this paper, our current approach of medical information fusion used by MedIC, pointing out how we can extract image-related medical information from different sources, and insisting on the the way we can design and adapt Bayesian Networks, to extract the most accurate medical modality. This approach is relatively easy to extend to treat other extracted information (anatomical region, acquisition view-angle, pathology) by updating the *BN* design to take into consideration the additional knowledge.

On going experiments are taking into consideration the disposition of images in complex documents (as they are on health resources indexed by CISMeF), and thus, the extraction of additional image-related medical knowledge from the document text-regions. Once the most accurate and complete image information is extracted by MedIC, it can be used for automatic indexing of CISMeF resources (containing the images). This will provide additional search capability to the catalogue, to better assist users on their searches for quality heath-information on the Internet.

6 References

References

1. Koch, T.: Quality-controlled subject gateways: definitions, typologies, empirical overview. Online Information Review **24**(1) (2000) 2434
2. Darmoni, S., Leroy, J., Thirion, B., Baudic, F., Douyére, M., Piot, J.: Cismef: a structured health resource guide. Meth Inf Med **39**(1) (2000) 30–35
3. Neveol, A., Rogozan, A., Darmoni, S.: Automatic indexing of health resources in french with a controlled vocabulary for the cismef catalogue: a preliminary study. Medinfo (2004)
4. Chu, W.W., Hsu, C.C., Cardenas, C., , Taira, R.K.: Knowledge-based image retrieval with spatial and temporal constructs. IEEE Transactions on Knowledge and Data Engineering **10**(6) (1998) 872–888

5. Shyu, C.R., Brodley, C.E., Kak, A.C., Kosaka, A., Aisen, A.M., Broderick, L.S.: Assert: A physician-in-the-loop content based retrieval system for hrct image databases. Comp.Vision and Image Understending **75**(1/2) (1999) 111–132
6. Tang, H.L., Hanka, R., Ip, H.H., Cheung, K.K., Lam, R.: Semantic query processing and annotation generation for content-based retrieval of histological images. In: International Symposium on Medical Imaging. Volume 3976 of SPIE Proceedings., San Diego, CA, USA (2000)
7. Marée, R., Geurts, P., Piater, J., Wehenkel, L.: Biomedical image classification with random subwindows and decision trees. In: Proc. ICCV workshop on Computer Vision for Biomedical Image Applications (CVIBA 2005). Volume 3765. (2005) 220–229
8. Müller, H., Rosset, A., Vallée, J.P., Geissbuhler, A.: Integrating content-based visual access methods into a medical case database. In: Proceedings of the Medical Informatics Europe Conference (MIE 2003), St. Malo, France (2003)
9. Lehmann, T.M., Güld, M.O., Thies, C., Fischer, B., Keysers, M., Kohnen, D., Schubert, H., Wein, B.B.: Content-based image retrieval in medical applications for picture archiving and communication systems. In Proceedings, S., ed.: Medical Imaging. 5033, San Diego, California (2003) 440–451
10. Florea, F., Rogozan, A., Bensrhair, A., Darmoni, S.: Medical image retrieval by content and keyword in an on-line health-catalogue context. In: Proceedings of Mirage 2005 (Computer Vision / Computer Graphics Collaboration Techniques and Applications), INRIA Rocquencourt, France (2005) 229–236
11. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann (1992)
12. Jensen, F.V.: An introduction to Bayesian Networks. Taylor and Francis, London, United Kingdom (1996)
13. Pearl, J.: Graphical models for probabilistic and causal reasoning. In Gabbay, D.M., Smets, P., eds.: Handbook of Defeasible Reasoning and Uncertainty Management Systems, Volume 1: Quantified Representation of Uncertainty and Imprecision. Kluwer Academic Publishers, Dordrecht (1998) 367–389
14. Sebe, N., Lew, M., Cohen, I., Garg, A., T.S., H.: Emotion recognition using a cauchy naive bayes classifier. In: Proceedings of the International Conference on Pattern Recognition. (2002)
15. Heckerman, D., Geiger, D., Chickering, M.: Learning Bayesian networks: The combination of knowledge and statistical data. In de Mantaras, R.L., Poole, D., eds.: Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence, San Francisco, CA, USA, Morgan Kaufmann Publishers (1994) 293–301
16. Cooper, G., Hersovits, E.: A bayesian method for the induction of probabilistic networks from data. Maching Learning **9** (1992) 309–347
17. Chickering, D.M.: Learning equivalence classes of bayesian-network structures. Journal of machine learning research **2** (2002) 445–498
18. Leray, P., Franois, O.: Bnt structure learning package: Documentation and experiments. Technical Report 2004/PhLOF, Laboratoire PSI, INSA de Rouen, FRE CNRS 2645 (2004) `http://bnt.insa-rouen.fr/`.
19. Murphy, K.: The BayesNet Toolbox for Matlab, *Computing Science and Statistics: Proceedings of Interface*, 33 (2001) `http://www.ai.mit.edu/~murphyk/Software/`.