Rapport Technique PSI 2004 - no.01/FIF

Comparison of Feature-Selection and Classification Techniques for Medical Images Modality Categorization

Filip I.Florea[1,2] Alexandrina Rogozan[1,2] Abdelaziz Bensrhair[1] Stèfan J.Darmoni[2]
[1] PSI Laboratory (FRE CNRS 2645), INSA de Rouen
[2] CISMeF Team, Rouen University Hospital

{Filip.Florea,Alexandrina.Rogozan,Abdelaziz.Bensrhair}@insa-rouen.fr
Stefan.Darmoni@chu-rouen.fr

Keywords :
Cataloguing, Abstracting and Indexing, Classification, Internet, Content-Based Image Retrieval (CBIR), Medical Modality Categorization.

1 Introduction

This research work is concerned with the development of the French health catalogue CISMeF to assist healthcare professionals, students and general public in their search for quality health resources. The CISMeF project (French acronym of Catalogue and Index of Health On-Line Resources) [5] was initiated in 1995 in order to meet the users' need to find precisely what they are looking for among the numerous health documents available online [1]. CISMeF describes and indexes the most important resources of institutional health information in French. Indexing is a decisive step for the efficiency of information retrieval within the CISMeF catalogue, and it is also one of the most time consuming tasks for the librarians, demanding high-level documentary skills. Indeed, the textual content of resources is manually annotated with a metadata set and a structured terminology similar to a documentary ontology of the medical field [6].

Being aware of the medical image importance in healthcare, we currently aim to enrich the health catalogue CISMeF with an image retrieval engine allowing query by keyword and/or by the visual content. Therefore, the cataloguing of medical images present in the CISMeF resources requires not only visual-feature extraction (i.e. color, form, texture), but also metadata extraction (i.e. imaging modality, body region) to be added to the visual content. It is a challenging idea since existing medical retrieval systems are mostly based on textual queries and annotations of the medical images [10], whereas those by visual content are still on a prototype state, dedicated to a very specific medical context (e.g. pathology of the lungs) [24] and not always accessible via Internet. This makes impossible their validation and integration as effective tools to train or to assist medical students and healthcare professionals in the diagnosing stage, as well as the acquisition of iconographic knowledge on various pathologies.

2 Objective

The goal of this paper is the automatic categorization of medical images according to their corresponding modalities. This is a process containing 3 stages: a) the extraction of representative vector features, to describe image content, b) the selection of the best features from each subset (minimizing the number of features and maximizing the discriminative information carried by them)(they can be used individually or combined), and c) the training and classification of the resulting vectors in the desired classes (i.e. the desired modalities).

Being applied to an on-line medical resource health catalogue open to all medical resources, like CISMeF, we could not focus ourselves on a single modality problematics.

The paper aims at the extraction of most accurate medical image modality categorization algorithm, using texture and statistical global features (with the perspective of improving the decision by the extraction and interpretation of textual annotations found around the images, in complex medical documents).

[1] http://www.cismef.org

Table 1: Database Content

Modality	no. of images	repartition
Angiography	348	23.5%
Ultrasonography	246	16.6%
MRI	373	25.2%
X-ray	126	8.5%
CT scan	295	20%
Scintigraphy	87	5.8%
Total	1475	100%

3 Materiel and methods

3.1 Constitution of a medical image database

The first step in the implementation of a medical image categorization algorithm is the constitution of a representative medical image database.

A list of medical modalities used in daily practice was constituted by a medical expert from the Rouen University Hospital (RUH) , and implemented in *CISMeF* terminology as ressources type [6]. The initials developments and tests were carried out on a six-modality image database containing: angiography, ultra-sonography, magnetic resonance imaging (MRI), standard radiography, CT scan, and scintigraphy. This six were considered as the most frequent information-bearer modalities used in medical imaging.

This initial image database contains 1475 images extracted from Radiology, Radio Pediatry and Nuclear Imaging departaments of RUH. The images are imported from the hospital DICOM internal format, or secondary digitized, to JPEG format, as they are mostly found in any medical resource database, on the Internet. The images don't have the same dimension and quality, being acquired with different digital or analogical equipments, in different hospital services, with different parameters, in a time-period of several years. This results in an increased intra-modality variability, which means an increased difficulty of modeling discriminative measures but also means that the resulting algorithm is more adapted to real-live practice challenges.

3.2 Categorization algorithm

Region Of Interest Selection. Due to the fact that each modality-acquisition is applied to a number of different services inside a health-center, there is important intra-modality variability. Our purpose is to extract the criteria that maximize the differences between the modalities, and minimize (ignore) those between the examples of the same modality. The differences between modalities, are diminished by the presence of the background (which, due to the digitization process and jpeg compression, is not of uniform gray-level - can be approximated with a uniform texture) and the textual annotations (the text layer in DICOM format). These textual annotations are not always present, due to legal constraints, but when they are, the regions with this text have a relatively resembling texture.

That's why the extraction of the features that will differentiate between the classes (modalities) must be made outside the background/text regions. The background on the six modalities treated in this paper is either black or white, and it was relatively easy to extract by tresholding of the gray-level histogram. The dimension and font of the text annotations in the images treated were sufficiently close to be approximated and extracted with a TopHat filter set on the character's thickness [Fig. 1].

Once obtained the image without the background and text a rectangular region must be extracted, that will maximize the treated information [Fig. 2]. However, such a process being complex and difficult to employ, by observation from our experiments, a centered 512x512 window, of the background/text filtered image was considered, and proved sufficient.

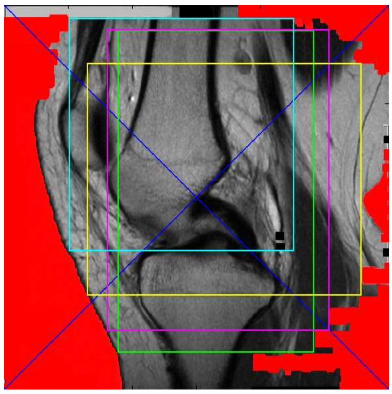

Figure 2: ROI selection

Image features. There always was a debate on whether global or local features should be used in describing images. Due to the technical difficulties that arise when trying to extract features at the object level from unknown images, and the good, usually sufficient, performance, of global extracted features, the general-purpose CBIR systems (like Photobook [22], QBIC [8]) use them.

This idea is not entirely applicable in medicine, due to the vastly superior complexity of medical images, and especially due to the importance of local char-

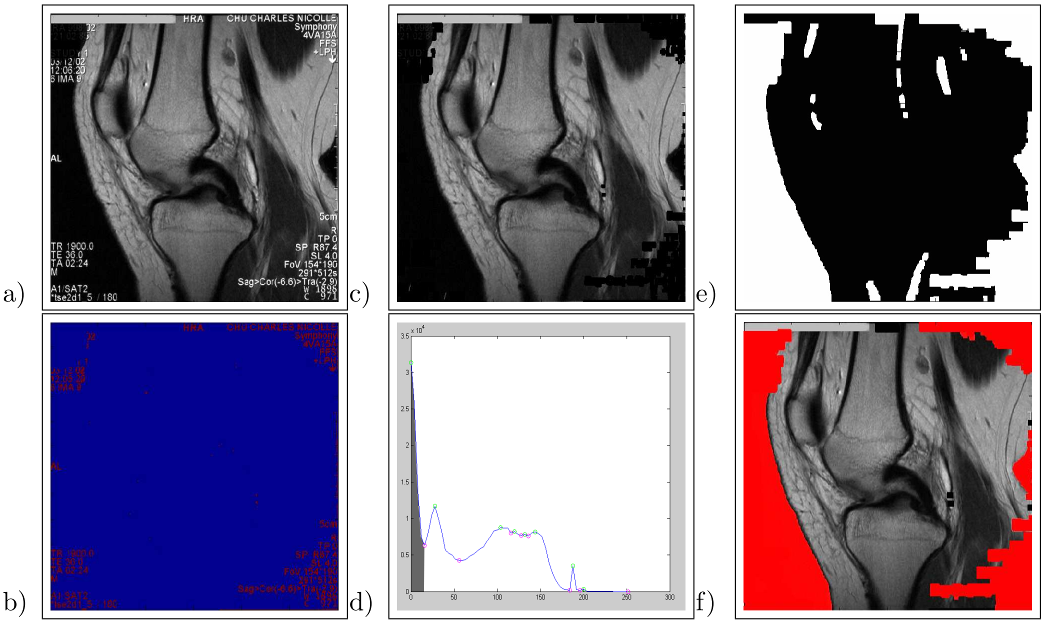

Figure 1: Discarding of the text/background information. a).initial image; b).TopHat filtering; c).the image without the text; d).the image gray-level histogram (the local min are shown in magenta and the local max in green)(the background approximation for the considered image is highlighted in grey); e).the background approximation (in white; there are some errors, but they are corrected with some supplemental conditions); f).the image without the text/background regions

acteristics in deciding the images content (modality, the anatomical region/biological system, pathology) (ex: the angiography, being a radiography-based sub-modality, could be so alike to a standard-radiography, that only some local characteristics, like contrasted vessels, could make the difference). Although, the global features, appear to be sufficiently discriminative for the categorization of medical images by modality or anatomical-region [12].

To capture the semantics of images, three main types of features can be extracted: color, texture and shape. Historically, color was the most successfully employed feature in general purpose CBIR. In medical imaging, the color is very seldom present and when it is, it describes only a fraction of the image information, rendering the color as a feature, unusable. To characterize the shapes inside medical images requires a certain amount of apriori knowledge, but can be proved useful with the extracting of organs or tumor edges (contours) once the acquiring modality and the body-region/biological-system are known [2]. Even with this high-level of apriori knowledge provided, the extraction of shapes can be an illusive one on some cases (i.e. ultrasonography, even radiography), so the shape-based features aren't appropriate for automatic modality categorization either. The texture remains the best suited descriptor and combined with statistical grey-level measures to account for the different acquisition biases (like contrast and brightness), proved to be a well suited global descriptor for medical images.

From the large amount of methods developed for describing texture [14], three were employed based on performance criteria: Harlick's grey-level co-occurrence matrix, the fractal dimension and the Gabor wavelets.

[13] proposed the cooccurrence matrix as a representation of texture. This approach explore the gray level spatial dependence of texture. Four cooccurrence matrixes were computed, one on each direction (horizontal, vertical and the two diagonals), and on each one, four features were extracted: Energy, Entropy, Contrast and Homogeneity, producing a 16 feature vector.

[21] made the assumption that textures are fractals for a certain range of magnifications. He used statistics of differences of gray levels between pairs of pixels at varying distances as indicators of the fractal properties of the texture. Fractal dimension is not an integer in contrast to the dimension in Euclidean geometry, but a number between 2 and 3; the more the texture is smooth (respectively rough), the more the fractal dimension is close to 2 (respectively 3). We used a modified box-counting texture analysis technique based on the probability density function described by [15]. The computing of the fractal dimension generates a single feature.

[9] suggested the use of so-called Gabor wavelets for feature extraction. The Gabor filters, computed at a certain wavelength and orientation, produce decomposition by ϕ orientations and λ scales. We obtain an $\phi^*\lambda = 24$ level decomposition on witch we compute the mean and standard deviation, resulting a 48 feature vector. Coarse textures will have spectral energy concentrated at low spatial frequency, while fine textures will have larger concentrations at high spatial frequency.

In addition we used features derived from gray-level statistical measures. The second moment (variance) is of particular importance because it measures gray-level contrast and can therefore be used to calculate descriptors of relative smoothness. We computed first moment mean, median and mode, second moment variance, l2norm, third and forth order skewness and kurtosis.

Overall we obtain a vector composed of 72 features.

Feature selection. When applying machine learning in practical settings the first difficulty is raised by the feature selection phase for the data at hand. The basic idea of feature selection algorithms is searching through all possible combinations of features in the data to find which subset of features works best for prediction. The selection is done by reducing the number of features of the feature vectors, keeping the most meaningful (features which together convey sufficient information to make learning tractable), discriminating ones, and removing the irrelevant or redundant ones. That way we can dispose of the false regularities, and increase the interpretability and efficiency. For the feature selection phase, two objects must be set up: a feature evaluator and a search method. The evaluator determines what method is used to assign a worth to each subset of features. The search method determines what style of search is performed.

The feature selection can be done two ways: 1) using full training set (the worth of the feature subset is determined using the full set of training data), or 2) by cross-validation (the worth of the feature subset is determined by a process of cross-validation).

In addition, the classifying time grow dramatically with the number of features, rendering the algorithm impractical for problems with a large number of features. In most cases, in our experiments, with the feature selection phase activated, minor accuracy decreases were noted (1-2%) but the classifying time was importantly reduced.

In practice, the choice of a learning scheme (the next phase) is usually far less important than coming up with a suitable set of features.

We experimented with several evaluators and search methods:

Evaluators:

- CfsSubsetEval - Evaluates the worth of a subset of features by considering the individual predictive ability of each feature along with the degree of redundancy between them; subsets of features that are highly correlated with the class while having low inter-correlation are preferred.
- ConsistencySubsetEval - Evaluates the worth of a subset of features by the level of consistency in the class values when the training instances are projected onto the subset of features.
- InfoGainFeatureEval - Evaluates the worth of an feature by measuring the information gain with respect to the class. InfoGain(Class,Feature) = H(Class) - H(Class — Feature).
- GainRatioFeatureEval - Evaluates the worth of an feature by measuring the gain ratio with respect to the class. GainR(Class, Feature) = (H(Class) - H(Class — Feature)) / H(Feature).
- SymmetricalUncertFeatureEval- Evaluates the worth of an feature by measuring the symmetrical uncertainty with respect to the class. SymmU(Class, Feature) = 2 * (H(Class) - H(Class — Feature)) / H(Class) + H(Feature).
- ChiSquaredFeatureEval - Evaluates the worth of an feature by computing the value of the chi-squared statistic with respect to the class.
- PCA - Performs a principal components analysis and transformation of the data.
- SVMFeatureEval - Evaluates the worth of an feature by using an SVM classifier.

Search methods:

- BestFirst - Searches the space of feature subsets by greedy hill-climbing augmented with a backtracking facility.
- GeneticSearch - Performs a search using the simple genetic algorithm described in [11]
- Ranker - Ranks features by their individual evaluations. Use in conjunction with feature evaluators (ReliefF, GainRatio, Entropy etc).

Table 2: Feature selection

	BestFirst	GeneticSearch	Ranker
CfsSubsetEval	FS1	FS2	×
ConsistencySubsetEval	FS3	FS4	×
InfoGainFeatureEval	×	×	FS5
GainRatioFeatureEval	×	×	FS6
SymmetricalUncertFeatureEval	×	×	FS7
ChiSquaredFeatureEval	×	×	FS8
PCA	×	×	FS9
SVMFeatureEval	×	×	FS10

3.3 Selected features classification

For the classification phase we experimented with 5 classifiers:

- **Multilayer Perceptron** - Multilayer perceptrons (MLPs) are feedforward neural networks trained with the standard backpropagation algorithm.
- **Support Vector Machines** - Implements John Platt's sequential minimal optimization algorithm for training a support vector classifier [23].
- **Random Forest** - Class for constructing a forest of random trees. For more information see [3]
- **Logistic Model Trees** - Classifier for building 'logistic model trees', which are classification trees with logistic regression functions at the leaves [17].
- **K Nearest Neighbors** - K-nearest neighbors classifier [1]

A standard PC architecture was used, with implementation in MatLAB, for the feature extraction methods, and Java, for the feature selection and classification (the Weka Data Mining Machine Learning Software[2]).

4 Evaluation

As previously mentioned, five features types were extracted [Table 3]. The features extracted from the grey-level histogram, were considered only for comparison. The 7 statistical measures and the texture-based features are forming uneven feature vectors, ranging from a single feature for the fractal dimension, to 48 for the Gabor wavelets. These vectors were evaluated separately to show the relevant information contribution of each one, or combined to combine the performance.

Eight feature selection methods were applied to this combined feature vector. [Table 2] shows all the possible combinations between the Evaluators and SearchMethods. The first two evaluators, CfsSubsetEval and ConsistencyEval can be combined with either BestFirst or GeneticSearch methods. The other 7 Evaluators are used with the Ranker search method, that provides an ordered feature vector (i.e. and not a selected-reduced one), a threshold being needed for the actual selection. From the initial combined vector of 72 features, we experimented with the most meaningful 30, 50 and 70 for each of the 7 algorithms that use the Ranker search method.

For the classification stage, the features extracted from the entire image data-base were used, and a 10-fold cross-validation scheme was employed. Basically cross-validation means that you use one part of the data to build a model, which you then apply to the other part of the data to assess how well the model fits the data. k-fold validation means that you divide the data randomly into k equal sized parts, use k-1 parts to build the model, and the remaining part to validate. You do this k times, each time using a different part for the validation. At the extreme k=N and you have a leave one out approach.

5 Results

[Table 4] shows the precision and time obtained by the 5 classifiers with each of the features individually or combined. The features extracted from the gray-level histogram were considered only for comparison, and, indeed, they does not exceed 75.21%. The Gabor-features obtained the highest precision from all the individual features, with rates between 74.71% and 84.13%, and the performances of the statistical measures an the cooccurence matrix are close. Of course, the statistical measures were intended to be used in combination with textural measures so the

[2]www.cs.waikato.ac.nz/ml/weka/

Table 3: Feature Vectors

name	dimension	content
histogram	256	[histo1 . . . histo256]
statistical measures	7	[stat1 . . . stat7]
cooccurence	16	[co1 . . . co16]
fractal dimension	1	[fd]
gabor wavelets	48	[gb1 . . . gb48]
combined	72	[stat1 . . . stat7]+[co1 . . . co16]+[fd]+[gb1 . . . gb48]
feature selection on the combined vector	?	([stat1 . . . stat7]+[co1 . . . co16]+[fd]+[gb1 . . . gb48])FS
FS1	9	[mean, median, mode, l2norm, skewness, gb4, gb12, gb19, gb41]
FS2	20	[mean, median, mode, l2norm, skewness, kurtosis, co4, co16, fd, gb1, gb2, gb5, gb6, gb8, gb11, gb15, gb21, gb24, gb30, gb38]
FS3	10	[median, mode, l2norm, kurtosis, co10, co11, df, gb1, gb4, gb25]
FS4	23	[mean, mode, co7, co8, co9, co11, df, gb6, gb10, gb11, gb12, gb13, gb15, gb16, gb19, gb24, gb25, gb30, gb32, gb38, gb41, gb43, gb47]
FS5	72	[l2norm, mean, median, mode, skewness, kurtosis, gb4, gb12, gb11, gb3, gb5, gb6, gb10, gb9, std, gb24, gb1, gb19, gb16, gb43, gb2, gb23, gb7 ...]
FS6	72	[mean, median, l2norm, skewness, kurtosis, mode, gb2, gb4, gb9, gb12, gb30, gb34, gb3, gb11, std, gb22, gb1, gb6, gb5, gb10, gb15, gb18, gb7, ...]
FS7	72	[mean, median, l2norm, skewness, mode, kurtosis, gb4, gb12, gb11, gb3, gb9, gb5, gb6, std, gb10, gb2, gb1, gb24, gb16, gb19, gb15, gb23, gb7, ...]
FS8	72	[l2norm, mean, median, mode, skewness, kurtosis, gb5, gb11, gb4, gb19, gb12, gb3, gb9, gb10, gb6, std, gb1, gb17, gb21, gb7, gb43, gb23, gb15, ...]
FS9	12	12*[0.104mean+0.106median-0.12std+0.073mode+0.092l2norm +0.027skewness+0.022kurtosis+0.02 co1+0.126co2+0.242co3 -0.231co4+0.047co5+0.12 co6+0.241co7-0.234co8+ ...]
FS10	72	[gb41, mean, gb39, gb24, std, co10, gb4, gb23, co4, co1, median, gb44, gb19, df, mode, gb38, l2norm, gb33, gb5, gb40, gb25, gb12, gb11, co12 ...]

poorer results for them when used alone, are expliquable.

For all the classifiers, considering a combination of the textual and statistical features, improve the recognition rate with couple of percents.

The best classification results are obtained with the MultiLayer Perceptron, resulting nearly 90% of precision, but at a high time-cost. The SVM classifier obtained the poorest results, with all the considered vectors.

From the computing-time point of view, the classifiers employed can be divided in two categories: MLP and LMT that take even several hours of computation time (but obtained better classifing results), and SVM, RF and K-NN that can classify the data in only a couple of minutes.

The performance of the 10 feature selection methods and the precision and computing-time obtained by the 5 classifiers, on the resulting selected feature vectors, are presented in [Table 5].

The variation of performance between the feature selection methods is rather small, the precisions varying of approximatively 10% for a classifier. The FS1 obtained the highest precision then classified with a MLP, and FS2, FS3, FS4 and FS8 obtained similar performances. FS6 performed the poorest.

For those feature selection algorithms, that used the "Ranker" search-method, reducing the number of features always produced a drop in precision. The variations between the results obtained with 70, 50 or 30 features are not very important, proving that the most important features are, indeed kept. Of course the classifying time are significantly reduced.

6 Discussion

[20] raised the modality categorization problem, presenting a frame that decides the modality by analysis of Required and Frequently Occurring features. The IRMA project propose a general structure for semantic medical image analysis [18], and recently, body-region classification results are presented, taking in consideration multiple modalities but focusing on radiographs [12]. Even that the modality categorization step is important in a medical image retrieval context, it was not intensely researched because the medical image retrieval approaches to this day, where mostly concerned with the retrieval inside a certain modality. There are a series of implementations for different modalities like: KMeD and COBRA [4] [7] that are treating MRI head images, ASSERT-system that deals with lung CT images [24], I-Browse that operates on histological slices [25], the system presented in [26] that investigate bone X-rays in ophthalmology, and another X-ray based system is presented in [[16] that describes the retrieval of tumor shapes in mammogram X-rays. A more recent approach in [19] employs shape-analysis on spine X-rays to automatically recognize a series of pre-defined pathological patterns. Given the fact that the principles used by each of these systems are dependent of the particular conditions of diagnostics context, including image modality, they are not applicable to other cases.

For the experiments presented in this paper, only 6 modalities were considered. To treat real live problems, like the CISMeF catalogue, the number of modalities will grow to over 60 (e.g. doppler ultasonogaphy, echocardiography, ...). Increasing the number of modalities will complicate the problem, and decrease results of the classification.

The best results, obtained with the MLP, are situated around 90%. This limitation is due to:

- the images that exist in our database, being acquired from a real-live health-care environment, are covering a large scale of body regions, pathologies or acquisition parameters. This results in a increased intra-modality variability, making very difficult to generalize ... [Figure3]
- Furthermore, there are modalities that are visually, very close, like MRI an CT-scan or Angiography and X-Ray [Figure4].
- The database is not equilibrated, meaning that the repartition of the images in modalities is not in concordance with the internal complexity of each one. The Scintigraphy is clearly the most under-represented, but given the significant visual difference, with regards to the other modalities, the recognition rates are rather high [Table matrice de confusion]. The X-ray class has also a smaller number of images, and thus, the poorer results of X-ray recognition are somehow expliquable.
- The presence of Angiographys images saved in negative and MRI-MRA(Magnetic Resonance Angiography) perturbed even more the classifying results. The first is used for better localize some local characteristics (blood vessels) (fig). The second is an MRI study of the blood vessels, and thus has the appearance of an Angiograhy [Figure5].

Since the beginning of this study, the CISMeF team has already developed an exhaustive list of medical

Table 4: Individual or combined feature vectors

Feature \ ClassifyMethod		**MLP** Multi-layer Perceptron		**SVM** Support Vector Machines		**RF** Random Forest		**LMT** Logistic Model Tree		**K-NN** K-Nearest Neighbours	
method	nf	p(%)	t(s)	p(%)	t(s)	p(%)	t(s)	p(%)	t(s)	p(%)	t(s)
Histogram (Histo)	256	69.21	19282	67.86	289	73.08	243	75.21	10183	71.59	69
Statistical measures (Stat)	7	64	198	62.77	40	64.40	30	66.10	1071	60.88	4
Co-occurrence (CO)	16	63.59	448	50.44	59	62.44	45	70.16	4037	58.77	9
Gabor (GB)	48	84.13	2024	74.71	58	81.83	55	83.81	13792	82.03	25
Stat+CO+DF+GB	72	89.55	5383	84.61	69	85.83	72	87.11	18889	86.84	41

Table 5: Feature selection

Feature \ ClassifyMethod		**MLP** Multi-layer Perceptron		**SVM** Support Vector Machines		**RF** Random Forest		**LMT** Logistic Model Tree		**K-NN** K-Nearest Neighbours	
method	nf	p(%)	t(s)	p(%)	t(s)	p(%)	t(s)	p(%)	t(s)	p(%)	t(s)
(Stat+CO+DF+GB)FS1	9	80.27	236	75.45	32	82.44	16	81.62	1161	80.54	5
(Stat+CO+DF+GB)FS2	20	83.45	655	78.91	55	82.71	27	83.38	3903	82.50	12
(Stat+CO+DF+GB)FS3	10	77.55	296	73.49	49	82.16	19	80.74	1032	79.25	6
(Stat+CO+DF+GB)FS4	23	84.13	696	79.25	59	84.88	30	84.13	6145	85.55	13
(Stat+CO+DF+GB)FS5	70	**89.15**	5422	**84.40**	89	85.42	62	**86.77**	9032	**87.05**	36
(Stat+CO+DF+GB)FS5	50	87.32	3010	83.05	54	**85.43**	41	86.37	3942	85.69	30
(Stat+CO+DF+GB)FS5	30	86.71	1482	82.23	43	84.80	30	85.49	2023	84.88	16
(Stat+CO+DF+GB)FS6	70	**88.81**	4980	**84.27**	52	**86.77**	57	**87.05**	8522	**86.54**	36
(Stat+CO+DF+GB)FS6	50	88.67	3172	83.38	46	86.50	44	86.57	3920	86.30	30
(Stat+CO+DF+GB)FS6	30	84.47	1592	80.61	35	83.72	29	85.93	2002	84.54	16
(Stat+CO+DF+GB)FS7	70	89.35	5043	**84.40**	52	85.22	45	**86.84**	8720	**87.05**	37
(Stat+CO+DF+GB)FS7	50	**89.42**	3210	82.71	39	**86.10**	44	86.37	4013	86.03	28
(Stat+CO+DF+GB)FS7	30	86.84	1602	82.23	36	84.74	30	85.69	2112	84.88	16
(Stat+CO+DF+GB)FS8	70	**90.50**	5361	**84.40**	86	85.08	63	**87.18**	8647	**87.05**	38
(Stat+CO+DF+GB)FS8	50	88.06	2891	84.20	67	**85.62**	45	87.05	5292	86.30	25
(Stat+CO+DF+GB)FS8	30	86.57	1302	80.47	44	85.15	30	83.11	2963	84.67	17
(Stat+CO+DF+GB)FS9	72	82.23	408	78.91	41	82.91	21	82.57	672	85.83	6
(Stat+CO+DF+GB)FS10	70	**88.99**	5112	**84.67**	80	86.30	62	**87.11**	8920	86.64	37
(Stat+CO+DF+GB)FS10	50	88.33	3282	84.33	50	**87.05**	50	87.05	4120	**87.11**	27
(Stat+CO+DF+GB)FS10	30	87.79	1708	83.38	38	86.64	29	86.50	2103	86.23	16

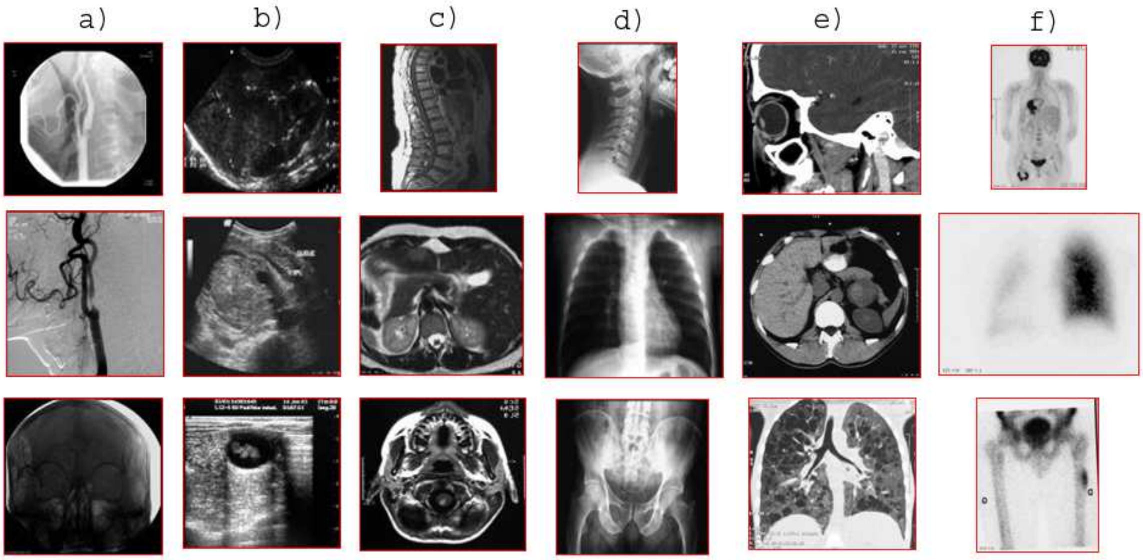

Figure 3: Intra-classe variability. a).Angiography; b).Ultrasonography; c).MRI; d).X-Ray; e).CT Scan; f).Scintigraphy

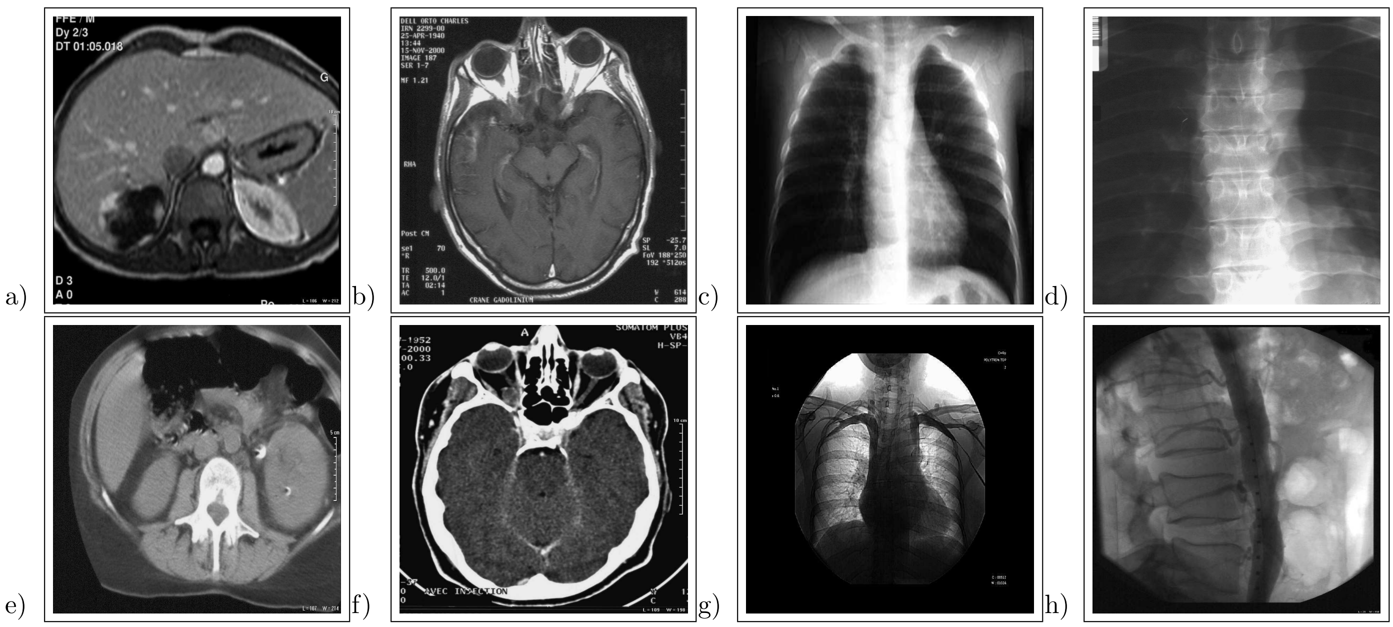

Figure 4: Confusion a) and b).MRI; c) and d).X-Ray; e) and f).CT-scan; g) and h).Angiography; The confusion is between a-e, b-f, c-g and d-h

a) b) c) d) e) f)

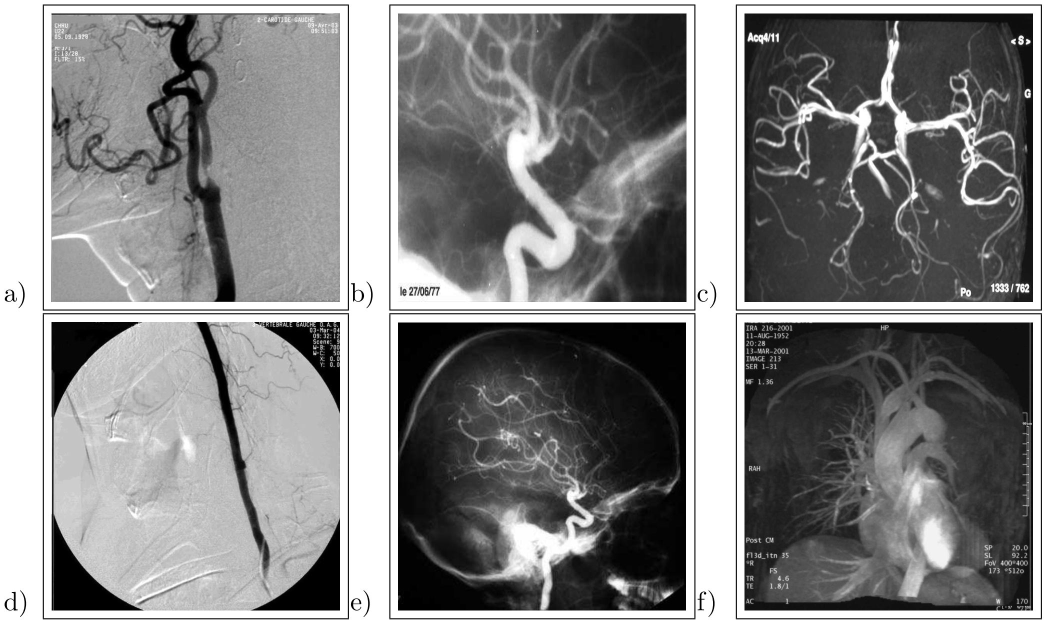

Figure 5: a). and d). Standard Angiography; b). and e). Negative Angiography; c). and f). Magnetic Resonance Angiography

image (sub)types (N=65) derived from the MeSH tree of diagnostic imaging. This image type list will improve the resource type already developed in the CISMeF terminology. The ultimate goal of the image categorization algorithm will be to automatically classify the N image types defined and not only the main six modalities evaluated in this first phase.

References

[1] D. Aha and D. Kibler. Instance-based learning algorithms. *Machine Learning*, 6:37–66, 1991.

[2] S. Antani, L. R. Long, and G. R. Thoma. A biomedical information system for combined content-based retrieval of spine x-ray images and associated text information. In *Proceedings of the 3rd Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP 2002)*, Ahamdabad, India, 2002.

[3] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, October 2001.

[4] W. Chu, C. Hsu, A. Cardenas, and R. Tiara. Knowledge-based image retrieval with spatial and temporal constructs. *IEEE Trans KDE*, 10(6):872–888, 1998.

[5] S. Darmoni, J. Leroy, B. Thirion, F. Baudic, M. Douyère, and J. Piot. Cismef: a structured health resource guide. *Meth Inf Med*, 39(1):30–35, 2000.

[6] M. Douyère, L. Soualmia, A. Névéol, A. Rogozan, B. Dahamna, J. Leroy, B. Thirion, and S. Darmoni. Enhancing the mesh thesaurus to retrieve french online health resources in a quality-controlled gateway. *Health Information and Libraries Journal 2004 (in press)*, 2004.

[7] E. El-Kwae, H. Xu, and M. R. Kabuka. Content-based retrieval in picture archiving and communication systems. *Journal of Digital Imaging*, 13(2):70–81, 2000.

[8] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D.Steele, and P. Yanker. Query by image and video content: The qbic system. *IEEE Computer*, 28(9):23–32, 1995.

[9] I. Fogel and D. Sagi. Gabor lters as texture discriminator. *Biological Cybernetics*, 61:103–113, 1989.

[10] T. Frankewitsch and U. Prokosch. Navigation in medical internet image databases. *Medical Informatics*, 26(1):1–15, 2001.

[11] D. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Publishing Company, Inc., Reading, MA, 1989.

[12] M. Güld, D. Keysers, T. Deselaers, M. Leisten, H. Schubert, N. Ney, and T. Lehmann. Comparison of global features for categorization of medical images. In *Proceedings SPIE 2004*, volume 5371, 2004.

[13] R. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *TransSMC*, 3(6):610–621, November 1973.

[14] B. Jähne, H. Haußecker, and P. Geißler. *Handbook of computer vision and applications vol.2*. Academic Press, 1999. Interdisciplinary Center for Scientific Computing University of Heidelberg, Heidelberg, Germany.

[15] J. Keller, S. Chen, and R. Crownover. Texture description and segmentation through fractal geometry. *CVGIP*, 45:150–166, 1989.

[16] P. Korn, N. Sidiropoulos, C. Faloutsos, E. Siegel, and Z. Protopapas. Fast and effective retrieval of medical tumor shapes. *IEEE Trans. on Knowledge and Data Engineering*, 10(6):889–904, 1998.

[17] N. Landwehr, M. Hall, and E. Frank. Logistic model trees. *ECML 2003*, 2003.

[18] T. Lehmann, M. Güld, C. Thies, B. Fischer, K. Spitzer, D. Keysers, H. Ney, M. Kohnen, H. Schubert, and B. Wein. The irma project - a state of the art report on content-based image retrieval in medical applications. In *Proceedings 7th Korea-Germany Joint Workshop on AdvancedMedical Image Processing*, pages 161–171, EWHA Womans University, Seoul, Korea, 14.-15.October 2003. Kim, M.H. and Meinzer, H.P.

[19] L. Long, S. Antania, D. Leeb, D. Krainakc, and G. Thomaa. Biomedical information from a national collection of spine x-rays. film to content-based retrieval. In *Proceedings SPIE 2003*, volume 5033, 2003.

[20] A. Mojsilovic and J. Gomes. Semantic based categorization, browsing and retrieval in medical image databases. *Proc. Int. Conf. Image Processing ICIP2002*, Sept 2002.

[21] A. Pentland. Fractal-based descriptors of natural scenes. *IEEE Trans on PAMI*, 6(6):661–674, 1984.

[22] A. Pentland, R. W. Picard, and S. Sclaro. Photo book: Tools for content-based manipulation of image databases. *International Journal of Computer Vision*, 18(3):233–254, 1996.

[23] J. Platt, B. Schölkopf, C. Burges, and A. Smola. Fast training of support vector machines using sequential minimal optimization. In *Advances in Kernel Methods - Support Vector Learning.* MIT Press, 1998.

[24] C. R. Shyu, C. E. Brodley, A. C. Kak, A. Kosaka, A. M. Aisen, and L. S. Broderick. Assert: A physician-in-the-loop content based retrieval system for hrct image databases. *Comp. Vision and Image Understending*, 75(1/2):111–132, 1999.

[25] H. Y. Tang, Lilian, R. Hanka, H. H. S. Ip, K. K. T. Cheung, and R. Lam. 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.

[26] W. Zhang, S. Dickinson, S. Sclaroff, J. Feldman, and S. Dunn. Shape-based indexing in a medical image database. In *Procs IEEE Workshop on Biomedical Image Analysis*, pages 221–230, 1998.