EUSIPCO 2013 1569743515

ASSESSMENT OF AN AUTOMATIC SYSTEM CLASSIFYING HEPATIC LESIONS ON MULTI-PHASE COMPUTER TOMOGRAPHY IMAGES

Auréline Quatrehomme [1,2], *Ingrid Millet* [3], *Denis Hoa* [1], *Gérard Subsol* [2], *William Puech* [2]

[1] IMAIOS, Montpellier, France
[2] LIRMM, University Montpellier 2 CNRS, Montpellier, France
[3] Department of Medical Imaging, Lapeyronie University Hospital, Montpellier, France

ABSTRACT

In this paper, we propose to assess the results obtained by our automatic system for classifying hepatic lesions. Our database contains 107 multi-phase liver nodules from 7 different diagnosis types. As multi-phase scans are not commonly found in Computer Aided Diagnosis systems, an analysis of the potential improvement introduced by working on multi-phase versus single portal phase CT acquisitions is made, as well as the comparison of our tool classification results to the ones obtained by two radiologists. Experimental results led to the validation of the method.

***Index Terms*—** Multi-Phase Computer Tomography, Computer Aided Diagnosis, Liver focal lesions, Classification, Medical Imaging, Expert validation

1. INTRODUCTION

This study is part of a larger project of retrieval and classification of medical images led at IMAIOS company which offers healthcare e-learning products *(www.imaios.com)*.

The database is built from Computer Tomography (CT) images, commonly known as CT scans and widespread in abdominal medical imaging. Most existing classification or retrieval systems work on mono-phase CT (on the one called the portal phase), and on a two or three-class database. The particularities of the current work are the 7 diagnosis class set and the fact that the tool processes on multi-phase CT: a multiple acquisition using a contrast media to enhance the liver over time and improve the visibility of its structures. From a clinical point of view, it seems a very hard task to correctly identify the various types of lesions only on a single phase. Some recent efforts as [1] [2] [3] are focused on multi-phase examinations.

A specific multi-phase classification system was built [4]. The aim of this work was first of all the validation of this multi-phase approach, and then the comparison between the classification of the hepatic lesions achieved by our computerized tool and by the eye of the radiologists.

2. PRELIMINARIES

A few papers as [3], [5] have been published about Computer Aided Diagnosis (CAD) using liver CT scans. Surprisingly, most databases found in the litterature contain images from one single CT phase. In order to improve the contrast of the captured images, and therefore the accuracy of the diagnosis, contrast media injection is widely used in clinical practice. One series of images is first captured on the patient (pre-injection phase). Then, the patient receives the injection, and three different series are acquired at three different times: this is multiphase CT examination. The gradual diffusion of the media will cause vessels and lesions enhancement. Radiologists would not make a diagnosis without the essential temporal information arised from these multi-phase scans. Indeed, a lesion indistinguishable from the healthy liver in one phase will be revealed in another phase. Moreover, different types of lesions have different enhancement patterns and timelines.

We found two attempts of classifying liver lesions on multiphase CT, which will be presented briefly below (see [4] for a detailed comparison of our approach to these methods). Both use classic visual features and classification algorithms. Their main drawback is their small diagnosis set (3, 4 classes).

Ye *et al.* [1] compared the results obtained from Support Vector Machines (SVM) classification on each phase separately with textural features: first order statistics as well as statistics computed over the image co-occurrence matrix. Besides, they introducted temporal descriptors over the phases (but on the mean value of the pixels only). Their database was made of 131 four-phase examinations from 4 classes: healthy liver, cyst, hepatocellular carcinoma (HCC) and haemangioma. The classification was always binary: normal vs. abnormal, cyst vs. other diseases, haemangioma vs. HCC.

Duda *et al.* [2] focuse on texture characteristics. Their database includes 165 lesions from 3-phase CT acquisition. They tested 4 sets of features (first order statistics, Law entropy, Run-Length and Co-occurrence matrices measures) independantly at each phase, then all sets of features at each phase, each feature set at all phases, and finally all features at all phases altogether. Both SVM and Dipolar Decision Tree were used as classifiers to distinguish between healthy liver, HCC and cholangiocarcinoma.

3. PROPOSED METHOD

3.1. Data

A skilled radiologist conducted a retrospective analysis of daily CT scans captured on two different scanners at the University Hospital of Montpellier between 2008 and 2011 and selected 107 lesions from 40 different patients. Thus, no patients were irradiated for this research, and no particular procedure other than routine protocol was followed. This dataset size is comparable to the ones of similar studies ([1], [2]). Our set of diagnosis types covers the majority of focal hepatic lesions: cysts, adenomas, haemangiomas, HepatoCellular Carcinoma (HCC), Focal Nodular Hyperplasia (FNH), and metastasis. Their visual aspect among the different CT phases is illustrated in Table 1 as well as their repartition in our database. Each nodule is a set of two to four 2D DICOM images, depending on the number of phases captured from the patient.

Our system works directly on the DICOM images format, which is the standard for medical images (on grey-level pixels). A 2D rectangular bounding box has been drawn manually by the same radiologist very closely around the lesion on its middle slice, featuring the region of interest (ROI) throughout this work.

3.2. Automatic classification tool

Figure 1 presents an overview of the proposed system. Visual features are computed over the ROIs and form multiphase vectors, which are entered into a SVM classifier. A cross-validation technique is finally conducted for classification evaluation. Given the speed of our software, we would be able to classify new lesions in real-time. Our method will now be detailed: first the feature extraction step before the classification and evaluation scheme.

Feature Extraction

Visual features describe the characteristics of the image, express its content (colors, texture or shape). We implemented 4 sets of features computed over all the CT phases, described below.

- First order statistics over grey-level histogram: 4 values
- Law texture measures: 28 attributes
- Gaussian Markov Random Fields: 5 statistical values
- Unser histograms: 36 descriptors (9 over 4 directions)

The last three of them characterize the texture of the lesion. Indeed, discussions with clinicians suggest that texture is of paramount importance for lesion recognition, more than grey levels which can vary a great deal according to the patient and / or examination conditions.

To our knowledge statistics computed over Unser histograms) have never been tested out of its reference article [6]. Sum and Difference histograms are a high-speed and low memory alternative to Grey-Level Co-occurrence Matrix computation. We ensured that its classification performance is similar or even superior than that from the classic method.

The final set contains 73 attributes on grey levels and texture over the four phases, which leads to a total of 292 descriptors. The feature vector for each lesion will contain all the measures side to side, one phase before another. All feature vectors are pre-computed in order to fasten the system.

Classification

Weka is a collection of machine learning algorithms, written in Java and developed at the University of Waikato, New Zealand *(www.cs.waikato.ac.nz/ml/weka)*. It can deal with missing values, which is helpful because routine CT scans may be made of one to four series. We tried several implemented methods before setting our choice on a SVM algorithm called Sequential Minimal Optimization (SMO) [7]. This classifier separates the data by an hyperplane (or a set of hyperplanes) in a high or infinite-dimensional space. In this new space, separations in the data that could not be seen in the initial one may be revealed. At first, feature values are normalized, missing ones replaced and nominal attributes transformed into binary ones. Indeed, the SVM algorithm builds several binary models, one for each pair of classes, over a polynomial kernel.

Classification Validation

A Leave One Out (LOO) cross-validation technique was conducted. Cross-validation is used to estimate how accurately the predictive model will perform in practice. One round of cross-validation consists of partitioning the dataset into 2 complementary subsets. The training is performed on the first one, while the second one is used for validation purpose. Multiple rounds are achieved over different partitions. As suggested by its name, in LOO cross-validation, a single observation of the set is designated as the validation data, and the remaining ones as the training set. The classification is conducted exhaustively n times, with n the number of observations, such that each one is used once for testing.

3.3. Expert analysis

Two radiologists have determined a diagnosis class for each lesion. They both have a 10-years of experience in reading liver CT images. Only the ROI was visible, on the available phases of each examination, the rest of the image being completely black. Our experts both reported the difficulties inducted by this lack of visual environment around the nodules. We merged their analysis as followed: a lesion is considered as well recognized if both clinicians labelled it correctly.

The DICOM files were anonymised, so no information was given regarding the patient, its age, conditions, or the other analysis that were conducted. Of course, this context is not realistic, but the purpose was to evaluate how clinicians would perform when being put in the ”same” conditions as the computer.

Table 1. *Database class repartition and visual appearance of lesions by type and phase*

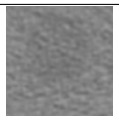
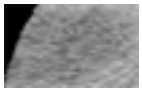
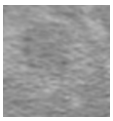
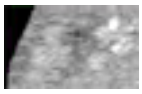
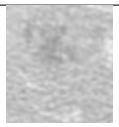
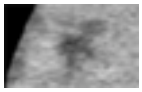
PHASE - LESION	Abcess	Adenoma	Cyst	FNH	Haemangioma	HCC	Metastasis
Number of lesions (TOTAL: 107)	6	10	25	8	8	13	38
1 pre-injection							
2 arterial phase							
3 portal phase							
4 late phase							

Fig. 1. *System framework overview*

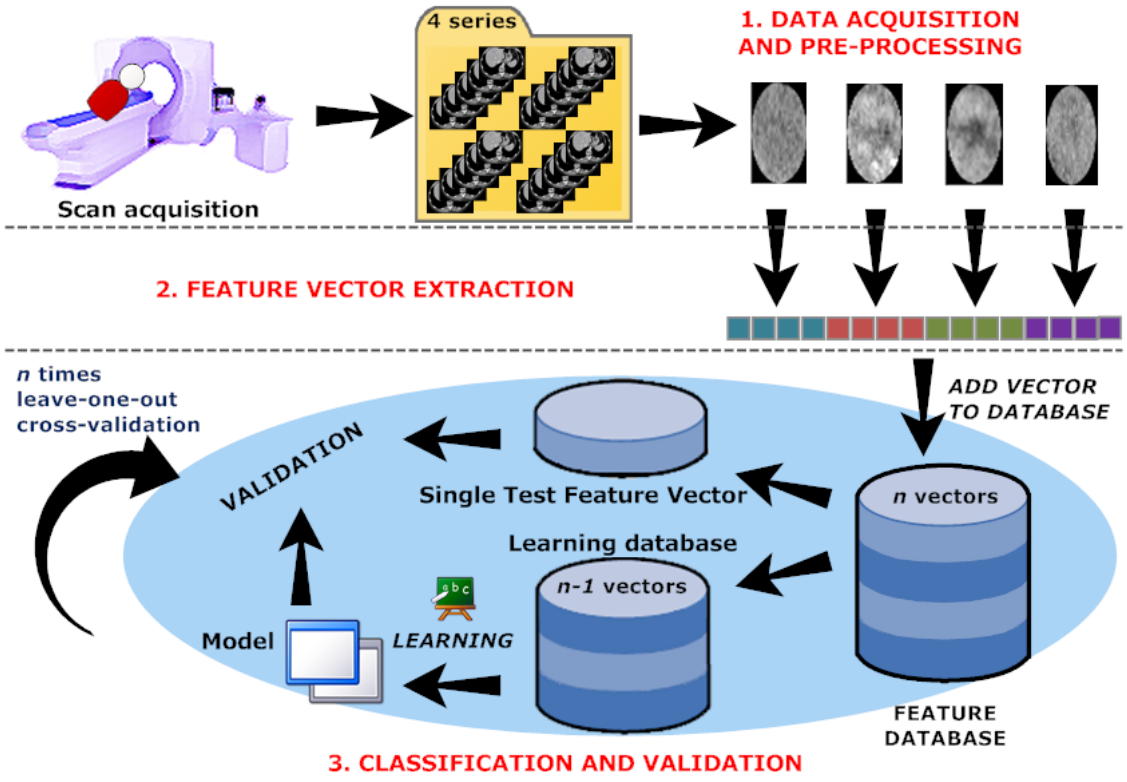


Table 2. *Classification results confusion matrix from our tool on portal and multi-phase*

	PORTAL PHASE							MULTI PHASE						
CLASS \FOUND	**Ab**	**Ad**	**Cy**	**FN**	**Ha**	**HC**	**Me**	**Ab**	**Ad**	**Cy**	**FN**	**Ha**	**HC**	**Me**
Abcess	**2**	0	1	1	1	1	0	**1**	0	0	1	0	0	3
Adenoma	0	**10**	0	0	0	0	0	0	**9**	0	0	0	0	1
Cyst	0	1	**23**	0	0	0	1	0	0	**24**	0	0	0	1
FNH	0	0	0	**0**	2	0	4	0	0	0	**1**	0	1	4
Haemangioma	3	1	0	2	**2**	0	1	0	1	1	0	**7**	0	0
HCC	1	2	0	0	0	**3**	7	0	0	1	0	1	**6**	5
Metastasis	4	2	3	1	2	2	**24**	3	1	1	3	0	5	**25**

4. RESULTS

4.1. Portal vs multi-phase classification

The confusion matrices of the classification results on single portal phase and multi-phase, obtained by our system are presented in Table 2. Results show that multi-phase introduction improves haemangioma and HCC recognition, respectively from 2 to 7 (out of 9) and from 3 to 6 (out of 13) lesions well labelled. On other nodule types, results are globally similar between portal and multi-phase classification. This can be easily explained as haemangioma and HCC nodules are hypervascular, and thus present a strong enhancement pattern variation over the phases. An overall improvement of 8.4% is measured on true positive score between portal and multi-phase classification.

4.2. Expert vs automatic classification

Table 3 presents the number of lesions from a diagnosis type correctly labelled over the overall number of lesions of this class. For expert analysis, a lesion is considered properly recognized if both radiologists designed its true class during the test. The similarity measurement known as Dice coefficient is also presented in order to evaluate if the lesions well identified by the radiologist and our tool are the same or not. This function ranges from 0 to 1, being 1 when the sets are identical. The recognition rate on multi-phase classification is similar between clinicians and our system results on four diagnosis classes out of the existing seven: abcesses, cysts, FNH and haemangiomas. Our system outperforms the radiologists on adenomas, HCC and metastasis.

Table 3. *Multi-phase lesions correctly classified by experts and our tool, and Dice coefficient over the two result sets.*

CLASS	**#**	**Success**		**Dice coeff.**
		Experts	Tool	
Abcess	6	1	1	0
Adenoma	10	0	**9**	0
Cyst	25	25	24	0.958
FNH	6	1	1	0
Haemangioma	9	7	7	0.667
HCC	13	0	**6**	0
Metastasis	38	17	**25**	0.468
TOTAL	**107**	**51**	**73**	**0.618**

5. CLASS ASSESSMENT

The results will now be discussed for each diagnosis class.

Abcesses: they are hard to identify out of clinical context. This is the reason why only one out of six is identified correctly, by human and by our tool, and the nodule recognized is not the same in both cases (null Dice coefficient). These lesions may display a wide range of visual aspects, as seen in Figure 2.

Fig. 2. *Illustration of abcess variability: three different abcesses captured on the same portal phase*

The lesion is darker than healthy liver but heterogeneous.

The nodule looks like a dark bean, with a strong separation with the grey healthy hepatic tissue, and some white zones on the contours of the lesion (which are the inflammation).

The lesion is heterogenous, with areas that darker and whiter, in comparison with healthy liver, but white zones being spread out and not on the contours.

Adenomas: this nodule type was poorly identified by the radiologist (10 % in both portal and multi-phase examinations), but well labelled by our tool (respectively 100% and 90%). Adenomas are supposed to be easy to identify, but it seems that our database contains not enough typical adenomas. This diagnosis class needs to be expanded, however adenomas are rare among the whole population.

Cysts: these very common lesions, of which visual aspect does not vary among phases are well recognized, both by our system (92 to 96% recognition) and by the radiologist (100%), in mono-phase as well as in multi-phase process.

FNH: results for this class are poor, both for radiologists (1 success out of 6 cases), and our tool (0 and 1 out of 6), with a zero Dice coefficient (the lesion identified by the clinicians is not the same than the one recognized

by our software). FNH typical visual information is a central scar which occurs in 60-70% of patients, but this may be a subtle sign.

Haemangiomas: there is a spectacular rising of the good classification rate with multi-phase scans by our software (from 22% to 78%). This is justified by the fact that its enhancement pattern among the different phases is important clue for diagnosis. With multi-phase CT scan, radiologists and software obtain an identical good score (7 out of 9). The Dice similarity measure of 0.67 tells us that 4 out of the 7 recognized lesions are commonly labelled by both the experts and the software.

HCC: results here are poor. These nodules are usually recognized in a clinical suspicion context (cirrhosis, overweight), and with the help of other analyses. The multiphase slightly improves the results using our tool (from 23 to 46% recognition). Our multi-phase classification outperforms experts, which obtain a null score.

Metastasis: The confusion vector is spread out: metastasis aspect may vary and look similar to any other kind of lesion, which makes this class recognition a hard task. Surprinsingly though, our tool obtains a good score on metastasis recognition (79% versus 49 % for experts). This may be due to this visual aspect variability combined to the important number of lesions in this class: our software might tend to put in the metastasis class all the lesions it cannot label in another class, leading to a bias in results. FNH label is the strongest illustration of this phenomena in portal vs multi-phase experiment. FNH confusion vector supports this hypothesis. The Dice coefficient tells us that only half of the lesions correctly labelled were common to the radiologists and our tool.

The first conclusion is that our system seems to hold the confrontation with radiologists.

The results obtained on haemangiomas and HCC support the idea of working on multi-phase examinations instead of single-phase. The distinction between cysts and other lesions could be made on a single portal phase basis, before discriminating the other diagnosis types on multi-phase captures. There is work to do in order to differentiate metastasis from other classes, and other classes from metastasis. Specific features on lesion variations over the phases may be extracted to improve some classes (abcess, FNH, HCC) recognition.

Eventually, the results obtained by our radiologists on HCC and FNH suggest that visual aspect alone may not be enough to classify some types of hepatic lesions, and that clinical context might be introduced in some way in CAD systems. An attempt is thus being made in our team to build a simple semantic ontology which will be introduced in our classification scheme.

6. CONCLUSION

We proposed the validation of a hepatic lesions classification system, with comparison to human expertise, of which major assets are a 7-class database and multi-phase CT scans images. Results states that introducing multi-phase examinations improves significantly the recognition of two hypervascular lesion classes, and that our software proceeds as well or even better than human experts put in the same conditions to recognize the class of hepatic nodules. As a future work, we plan to extend our database and explore temporal changes over the CT phases features as well as hybrid semantic and image processing techniques. We would like a more flexible model to take into account the variability inherent to some type of lesions.

7. REFERENCES

[1] J. Ye, Y. Sun, and S. Wang, "Multi-Phase CT Image Based Hepatic Lesion Diagnosis by SVM," in *2nd International Conference on Biomedical Engineering and Informatics*, pp. 1–5, 2009.

[2] D. Duda, M. Kretowski, and J. Bezy-Wendling, "Texture Characterization for Hepatic Tumor Recognition in Multiphase CT," *Biocybernetics and Biomedical Engineering* **26(4)**, pp. 15–24, 2006.

[3] S. Mougiakakou, I. Valavanis, A. Nikita, and K. Nikita, "Differential diagnosis of CT focal liver lesions using texture features, feature selection and ensemble driven classifiers," *Artificial Intelligence in Medicine* **41**, pp. 25–37, 2007.

[4] A. Quatrehomme, I. Millet, D. Hoa, G. Subsol, and W. Puech, "Assessing the Classification of Liver Focal Lesions by Using Multi-phase Computer Tomography scans," *Third MICCAI International Workshop MCBR* **7723**, pp. 80 – 91, 2012.

[5] J. Xu, J. Faruque, C. F. Beaulieu, D. Rubin, and S. Napel, "A Comprehensive Descriptor of Shape: Method and Application to Content-Based Retrieval of Similar Appearing Lesions in Medical Images," *Journal of Digital Imaging* , pp. 1–8, 2011.

[6] M. Unser, "Sum and Difference Histograms for Texture Classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8**(1), pp. 118 – 125, 1986.

[7] J. Platt, "Fast Training of Support Vector Machines Using Sequential Minimal Optimization," in *Advances in Kernel Methods - Support Vector Learning*, MIT Press, 1998.