

# Assessing the Classification of Liver Focal Lesions by Using Multi-phase Computer Tomography Scans

Auréline Quatrehomme<sup>1,2</sup>, Ingrid Millet<sup>3</sup>, Denis Hoa<sup>1</sup>, Gérard Subsol<sup>2</sup>,  
and William Puech<sup>2</sup>

<sup>1</sup> IMAIOS, Montpellier, France

aureline.quatrehomme@imaios.com

<sup>2</sup> LIRMM, Université Montpellier 2 / CNRS, Montpellier, France

<sup>3</sup> Department of Medical Imaging, CHU Lapeyronie, Montpellier, France

**Abstract.** In this paper, we propose a system for the automated classification of liver focal lesions of Computer Tomography (CT) images based on a multi-phase examination protocol. Many visual features are first extracted from the CT-scans and then labelled by a Support Vector Machine classifier. Our dataset contains 95 lesions from 5 types: cysts, adenomas, haemangiomas, hepatocellular carcinomas and metastasis. A Leave-One-Out cross-validation technique allows for classification evaluation. The multi-phase results are compared to the single-phase ones and show a significant improvement, in particular on hypervascular lesions.

**Keywords:** Medical Imaging, Computer Aided Diagnosis, Liver focal lesions, Multi-Phase Computer Tomography, Classification.

## 1 Introduction

Computer Aided Diagnosis (CAD) is a current dynamic field of research, with the help of recent imaging device improvements. For example, by integrating computer assistance in the diagnosis process of liver lesions, we can improve the efficiency of medical expertise and accuracy in classifying, detecting or segmenting the liver lesions. In this paper, we describe a preliminary study of a new method to classify hepatic lesions, without any detection or segmentation (as described in [1]), which is based on 4-phase CT imaging.

Section 2 introduces research on liver CT Computer Aided Diagnosis and some references dealing with multi-phase scans. Section 3 describes precisely how our dataset was built. Section 4 presents the method used and the results are analyzed in Section 5. In Section 6, we present some perspectives to improve these first results.

### 1.1 Multi-phase CT Acquisition

X-ray CT captures a large series of two-dimensional x-ray images, taken around one single rotation axis. Its usage has dramatically increased over the last two





















decades, in particular for abdominal exploration. In order to improve the contrast of the captured images, and therefore the accuracy of the diagnosis, contrast media injection is widely used. One series is first captured on the patient (pre-injection phase). The patient then receives the injection, and 3 series are taken at three different times: the first one, just after the injection, is called the arterial phase. The second, a few tens of seconds later, the portal phase. The last one, a few minutes after the injection: the late phase.

The diffusion of the media over the different phases captured will enhance the vessels and lesions. Radiologists would not imagine making a diagnosis without the essential temporal information provided from these multiphase scans. Indeed, the contrast enhancement varies from one phase to another: a lesion indistinguishable from the healthy liver in one phase will be revealed in another phase. This property is illustrated in Table 1, which visually shows these variations. Moreover, different types of lesions have different enhancement patterns and timelines. We have summarized information from a paper on strategies for hepatic CT and MRI imaging [1] in Table 2.

## 1.2 CT Liver Lesion Classification

Various papers have been published on Computer Aided Diagnosis (CAD) using liver CT scans. A team from Stanford focused on the shape of 8 types of liver nodules in [2], while they added in [3] semantic features to texture and boundary features in order to distinguish cysts, hemangiomas and metastases. These two papers apply their methods to Content-Based Image Retrieval (CBIR), which returns the images of the database which are the most similar to the query image. Mougiakakou *et al.* [4] applied multilayer perceptron neural networks, as

**Table 1.** Visual appearance of lesions by type and phase illustrating the importance of multi-phase CT scans

Phase \ Lesion	Cyst	Adenoma	Haemangioma	HCC	Metastasis
1 pre-injection					
2 arterial phase					
3 portal phase					
4 late phase					

**Table 2.** CT scan scenario and their clinical context: captured phases and context

CT scan scenario	Clinical context
<b>Single-phase</b> (portal)	No suspicion of a specific hepatic pathological condition
<b>Dual-phase</b> (arterial, portal)	Disease scenario with the primary cause outside of the liver, hypervascular hepatic metastases suspected
<b>Triple-phase</b> (before injection, arterial, portal)	Known or suspected cirrhosis, HCC, FNH or adenoma

well as a combination of primary classifiers, to the classification of liver tissue into healthy liver, cyst, hemangioma and hepatocellular carcinoma.

Surprisingly though, most image databases found in the literature contain images from one single CT phase despite its importance in the diagnosis process. We found two attempts for the study of the multi-phase CT classification of liver lesions, which will be presented below.

Duda *et al.* [5] focus on texture characteristics. Their database contains 165 lesions from 3-phase CT acquisition (no contrast, arterial and portal phase). They tested 4 sets of features (First-Order statistics, Law entropy, Run-Length matrix features and Co-occurrence matrix measures) independently at each phase, before all sets of features at each phase, then each feature set at all phases, finally all features at all phases altogether. SVM and the Dipolar Decision Tree were both used as classifiers to distinguish between healthy liver, HCC and cholangiocarcinoma.

Ye *et al.* [6] compared the results obtained from Support Vector Machines (SVM) classification on each phase with textural features: first order statistics as well as statistics computed over the image co-occurrence matrix. Furthermore, they introduced temporal tendency features over the phases. Their database consists of 131 four-phase examinations. The study is carried out on 4 classes: healthy liver, cyst, HCC and haemangioma, and the classification is always binary: normal vs. abnormal, cyst vs. other diseases, haemangioma vs. HCC. The temporal features idea seems interesting, although its application here is quite limited as the different features are computed over the mean value of the pixels (heterogeneous lesions might be hard to distinguish in this case). We also regret the lack of classification of the values obtained on the four phases, and the limitations resulting from by the binary classification scheme.

## 2 Data

### 2.1 Database Construction

With the help of 2 radiologists, we opted for five lesion diagnosis classes: cysts, metastasis and hypervascular lesions: adenomas, haemangiomas and hepatocellular carcinoma (HCC) which are presented in Table 1 and Table 3. This set

of diagnosis types cover the majority of focal hepatic lesions. Cysts are both benign and very commonly observed, but as their texture is homogeneous and their contours well defined, they have been under-represented in our database. On the other hand, adenomas, which are very rare but heterogeneous, are more present than in clinical reality. The repartition of the lesion types in our database is presented in Table 3. Our objective is not to determine whether the liver is in good condition or unhealthy, but to distinguish between nodular hepatic lesions, so no healthy tissue is present in the database.

**Table 3.** Lesion class repartition in our database

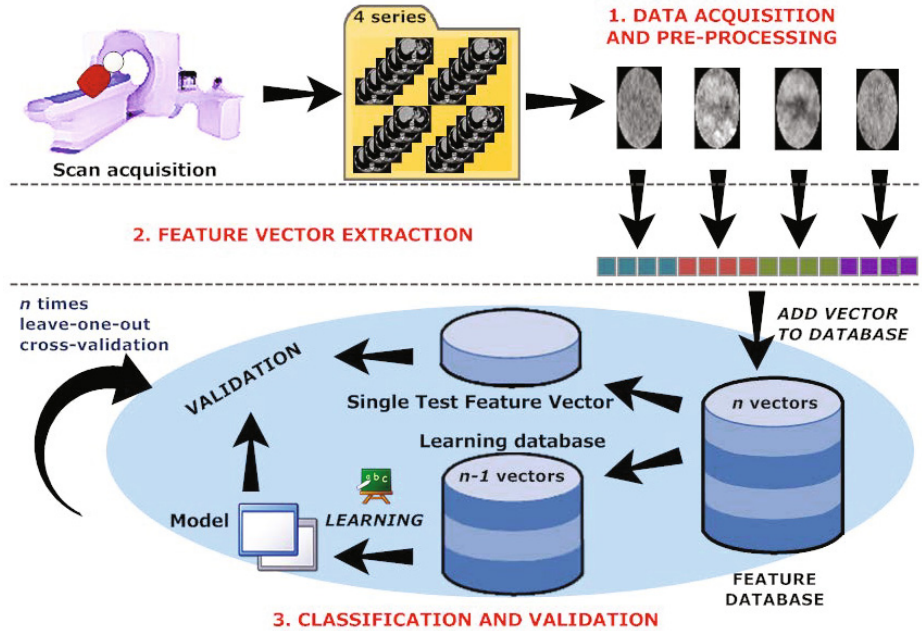
<b>CLASS</b>	<b>Cysts</b>	<b>Adenomas</b>	<b>Haemangiomas</b>	<b>HCC</b>	<b>Metastasis</b>	<b>TOTAL</b>
<b>NUMBER</b>	<b>25</b>	<b>10</b>	<b>9</b>	<b>13</b>	<b>38</b>	<b>95</b>

This is a retrospective analysis of daily CT scans conducted on two different scanners at the University Hospital of Montpellier between 2008 and 2011, so no patients were irradiated for our research, and no particular procedure other than the routine protocol was followed for the capture. An experimented radiologist looked for particular diagnosis clinical cases, and analyzed the CT images as well as the reports and complementary histological results which confirmed the diagnosis.

95 lesions of 40 different patients were selected to constitute our database. Its size is comparable to those of similar studies [5, 6]. The slice thickness and the number of phases vary, depending on what the radiologist was interested to see in the examination, which therefore determined the protocol. Slice thickness goes from 1.25 to 3 millimeters. 16 cases contain two phases images, 7 cases three phases, and 78 the four phases.

## 2.2 Data Pre-processing

We work directly with the DICOM images. As the pixel values of this format represent tissue densities, the entire range of the scale is kept and the grey levels are not normalized. The lesions are present on several CT slices, therefore a 2D rectangular bounding box was drawn around the lesions by an experimented radiologist in the middle single slice. No precise segmentation was done, in order to avoid certain problems, in particular due to the irregularity of the contours. In order to refine this rectangular box, and because we are working on *focal* lesions, the bounding ellipse in the rectangular zone defined by the radiologist will be used as region of interest (ROI), as presented in Figure 1, in the "Data acquisition and pre-processing" section. Therefore, lesion tissue will be studied instead of healthy liver. The ROI size ranges from 9\*12 to 165\*189 pixels, which is representative of the variety of hepatic lesion sizes.



**Fig. 1.** 3-step framework of proposed system: ROI then visual features extraction, before classification and evaluation

## 3 Method

### 3.1 System General Framework

Figure 1 presents an overview of the proposed system. Each lesion is a set of one to four 2D DICOM images, depending on the number of phases captured from the patient, on which a Region Of Interest (ROI) is extracted. Visual features are computed over these images and form multi-phase vectors, which are entered into a Support Vector Machine (SVM) classifier. A Leave-One-Out (LOO) cross-validation technique is finally conducted for classification evaluation.

First, the feature extraction step will be described in section 4.2, then the classification scheme in section 4.3. As in the papers by Duda *et al.* [5] or Ye *et al.* [6], our framework is broken down into 3 steps: feature extraction, training a classifier and classification (see Table 4 for comparison).

### 3.2 Feature Extraction

For segmentation, detection, retrieval or classification, the basic principle is to extract some visual features, or descriptors, from images. They describe the characteristics of the image, express its content (grey levels/colours, texture or shape). They are computed on the whole image, on each block obtained by

**Table 4.** 3 multi-phase system comparison based on multiple data, features and classification criteria

Characteristic	Ye <i>et al.</i> [6]	Duda <i>et al.</i> [5]	Our work
<b>Lesion number</b>	131	165	95
<b>Lesion size</b>	unknown	unknown	from 9*12 to 165*189 pixels
<b>Phases</b>	4	3 (late phase absent)	– 4-phase: 78 – 3-phase: 7 – 2-phase:16
<b>Diagnosis classes</b>	– HCC – cyst – haemangioma – healthy	– HCC – healthy – cholangio-carcinoma	– HCC – cyst – haemangioma – adenoma – metastasis
<b>Region Of Interest</b>	16x16 pixels square in the lesion manually delineated	manual circle of 30 to 70 pixels radii	manual rectangular bounding box around the lesion then automatically extracted inscribed ellipse
<b>Features</b>	– First Order Statistics, – Co-occurrence matrix statistics, – Temporal features	– First Order Statistics, – Co-occurrence matrix statistics, – Law measures, – Run-Length matrix features	– First Order Statistics, – Gaussian Markov Random Fields, – Law measures, – Unser histograms statistics
<b>Classifier</b>	SVM	SVM	– SVM – Dipolar Decision Tree
<b>Classification</b>	3 binomial sequential classifications: – healthy vs. pathology – if pathological: cyst vs. non cyst – if non-cyst: HCC vs haem.	Distinguish the 3 classes	Distinguish the 4 classes

dividing the image in small equally sized patches, or on Regions of Interest (ROIs), which have been delineated by a manual or automatic segmentation process. A review of the features can be found in [7] for recent CBIR systems, and in [8] for medical image classification.

We decided to begin our study with a few common features computed over the 4 phases, described below. All of them are extracted over the ellipsoid 2D ROI defined in Section 2. The first one, Unser histograms statistics, is an exception as it has never been tested to our knowledge.

**Unser Histograms:** Unser proposed in 1986 [9] an alternative method to the Grey-Level Co-occurrence Matrix (GLCM) computation, which reduces the memory requirement as well as the calculation time. GLCM, over which Haralick's well-known texture descriptors are computed, is replaced by estimates of the first order probability functions along its principal axes, which correspond to the second order probability functions over the image. These are called sum and difference histograms and they are extracted over four different directions. 9 statistical descriptors are then calculated over these two histograms in each direction, ending up with 36 attributes. Unser claims they are as accurate for classification as the GLCM statistics. We tested both Haralick and Unser measures and ended with similar and even better results with Unser, with the computation advantage already cited.

**Law Measures:** Kenneth I. Law proposed in 1980 [10] texture energy measures, which have been used for various applications. Its method to extract texture features is carried out in 3 steps. First, 25 convolution kernels are applied to the image. Secondly, a texture energy measure is computed on each convolved pixel by a windowing operation, and a new image is formed. Finally, these energy images are normalized then combined in order to obtain 14 rotation invariant final images. Mean and standard deviation are finally computed over them, ending with 28 attributes.

**Gaussian Markov Random Fields Measures:** Markov Random Fields systems model the dependency phenomena amongst image pixels using a statistical approach. The main idea is that, while neighboring pixels usually have the same intensity in an image, pixel values are independent of the pixels beyond that area. The image is therefore seen as a sample of a random process, where correlation between pixels is proportional to their geometric separation. Instead of being the real probability function computed over the image pixels, the field is a Gaussian in order to avoid high computational problems. The GMRF measures are its average, its standard deviation and 4 parameters named thetas. We keep standard deviation and thetas, while rejecting its average, which approximates very closely the image grey level average.

**Histogram Statistics:** mean, standard deviation, skewness and kurtosis computed over the grey-level histogram.

Our final set contains 303 attributes over grey levels and texture, on each phase. The feature vector for each lesion contains all the measures side to side, one

phase before another. All feature vectors are pre-computed in order to speed up the system.

### 3.3 Classification

Weka is a collection of machine learning algorithms, written in Java and developed at the University of Waikato, New Zealand (see [11] for an introduction). It can deal with missing values, which is helpful in our case where each CT scan consist of two to four series.

We tried several implemented processes before setting our choice on a classical method: Support Vector Machine (SVM). The algorithm implementation is called Sequential Minimal Optimization (SMO) and was proposed by John Platt [12]. The Support Vector Machines principle is to separate the data by a 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.

Before the classification, three pre-processing actions are conducted. First, missing values of each attribute are replaced by its mean. Our feature vectors do not all have the same length, depending on the number of phases of the CT acquisition. Then, nominal attributes are transformed into binary ones. Indeed, the SVM algorithm builds several binary models, one for each pair of classes. Finally, feature measures are normalized. The SVM kernel here is polynomial, with a 1.0 exponent.

### 3.4 Classification Validation

A Leave One Out (LOO) cross-validation technique is conducted.

Cross-validation is used to estimate how accurately our predictive model will perform in practice. One round of cross-validation consists of partitioning a sample of data into 2 complementary subsets. The analysis is performed on the first one (the training set), while the second one (testing set) is for validation. In order to reduce the effects of variability, multiple rounds as described are performed, using different partitions. The validation results are finally averaged over the rounds. Cross-validation gives more realistic results than classification and validation on the same complete database.

As its name suggests, in LOO cross-validation, a single observation of the set is designated as the validation data, and the remaining observations as the training data. The classification is conducted exhaustively  $n$  times, with  $n$  the number of observations, such that each one is used once for testing.

This classification with cross-validation is conducted in 0.19 seconds in the case of multi-phase, and 0.06 seconds in the case of mono-phase (for the complete lesion database). We are able to classify new lesions in real-time.



## 4 Results and Analysis

### 4.1 Analysis Scheme

The confusion matrix from multi-phase classification results was obtained and compared to the one from the portal phase. We extracted precision (also called true predictive value) and recall (also known as sensitivity) measures, as well as the F-measure of the test. Precision is a measure of the accuracy provided that a specific class has been predicted, whereas recall represents the ability to select instances of a certain class from a dataset. F-measure is an indicator of the global classification accuracy and it is defined by the weighted harmonic mean of precision and recall.

### 4.2 Precision, Recall and F-measure

The three measures chosen to evaluate our classification can be visualized in Figure 2. The same tendency can be observed over the three bar charts. The weighted average values show a global improvement of the three statistics by the introduction of multi-phase (+12% for precision and recall, +13% for F-score). If we have a closer look at the results obtained for each lesion type, the major phenomenon observed here is the spectacular improvement due to the multi-phase CT acquisition of the three measures for haemangioma and HCC (respectively from 56 to 63% and from 31 to 50%). Adenoma also benefits from multi-phase images, but to a lesser extent (from 5 to 8%). Regarding cysts and metastasis, portal phase evaluation seems sufficient: results are stable on cysts (8% maximum variation), and multi-phase has little positive influence on precision and F-measure (from 7 to 10%), whereas recall values goes down from 19%.

### 4.3 Confusion Matrices

Regarding the confusion matrix obtained with portal phase feature classification, cysts, adenomas and metastasis are quite well recognized (respectively 22 out of 25, 8 out of 10 and 35 out of 38), whereas heamangiomas and HCC are never recognized. One-third of the heamangiomas (3) have been labelled as adenomas and the other two (6) as metastasis. All HCC have also been classified as metastasis. This mislabelling on single phase analysis is expected as these lesions are hypervascular lesions and may be indistinguishable from a healthy liver at the portal phase. This confusion observed in portal phase has been pointed out, for example in [13], which studied the enhancement patterns of focal liver lesions during arterial time. At this phase, HCC, haemangiomas and metastasis may altogether present an homogeneous enhancement pattern, HCC and metastasis may both present abnormal internal vessels or variegated, complete ring or no enhancement pattern at all, while haemangiomas and metastasis may both present peripheral puddles or incomplete ring.

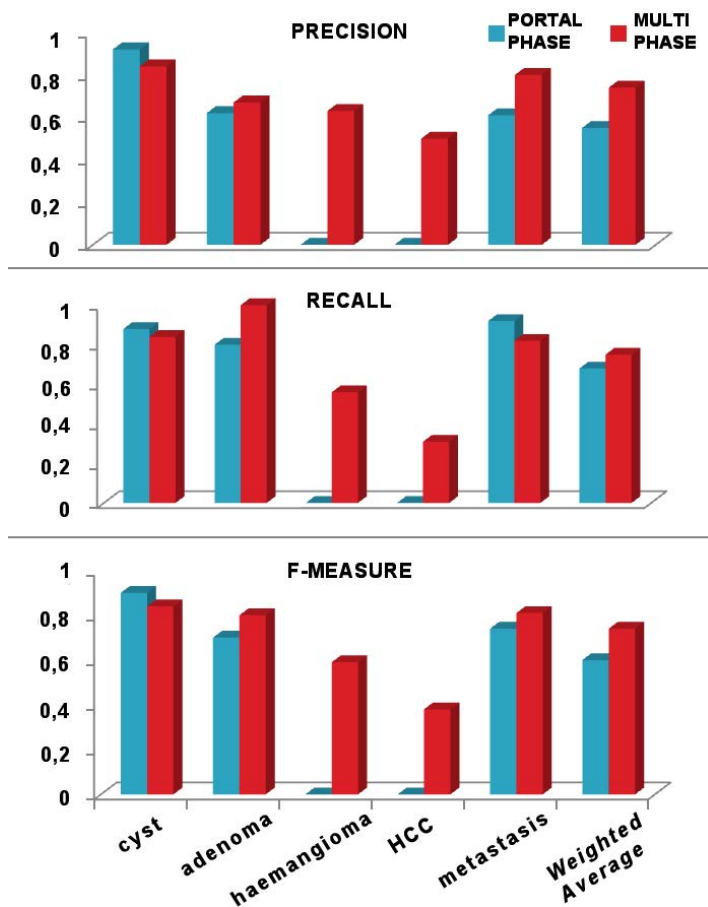


Fig. 2. Precision, Recall and F-measure values obtained on each lesion class as well as on the weighted average of all classes from portal phase and multi-phase classification

CLASS \ FOUND	PORTAL PHASE					MULTI-PHASE				
	Cy.	Ad.	He.	HCC	Me.	Cy.	Ad.	He.	HCC	Me.
Cyst	<b>22</b>	1	1	0	1	<b>21</b>	3	0	0	1
Adenoma	0	<b>8</b>	0	0	2	0	<b>10</b>	0	0	0
Haemangioma	0	3	<b>0</b>	0	6	1	0	<b>5</b>	3	0
HCC	0	0	0	<b>0</b>	13	0	0	2	<b>4</b>	7
Metastasis	2	1	0	0	<b>35</b>	3	2	1	1	<b>31</b>

Fig. 3. Confusion matrix on classification results: on the left: the real lesion type, on top: the labels determined by the classifier

As regards the confusion matrix obtained with multi-phase feature classification, and compared to portal phase results, HCC and haemangiomas recognition sharply increases, adenomas only slightly, cysts are stable while the metastasis score is falling marginally. What is significant to our earlier remark is that now 3 cysts are seen as adenomas, and the metastasis scheme has spread out over all other diagnosis classes. A sequential two-step classification could be considered: the first one, during the portal phase, to distinguish between cyst, metastasis or other nodule, and the second one, on all phases, if the first classifier labelled the instance as "other", to differentiate between adenomas, haemangiomas and metastasis. This idea coincides with the scheme detailed by Ye *et al.* in their paper [6]. For their part, haemangiomas and HCC are confused with each other in this matrix, and half of the HCC are still confused with metastasis as in the portal phase classification.

## 5 Conclusion

This paper presents a classical approach for liver lesion classification applied on multi-phase CT scans on the contrary of a majority of other studies which are based on the portal phase only. In this manner, the contrast enhancement patterns of the hepatic lesions can be taken into account.

We applied our system to a database of 95 2D CT images from 40 patients and evaluated its performances and compared them by using the portal phase only. The experimental results show a significant improvement of the classification results by using multi-phase scans, in particular for haemangiomas and HCC lesions. It is important to underline that we work on five diagnosis classes which spans most of the cases of liver lesions.

In the future, we plan to study the influence of each feature on the classification results in order to propose an automated feature selection. Temporal changes among the phases as well as a classification in sequence seem interesting leads to follow.

## References

1. Boll, D.T., Merkle, E.M.: Diffuse Liver Disease: Strategies for Hepatic CT and MR Imaging. *RadioGraphics* 29, 1591–1614 (2009)
2. Xu, J., Faruque, J., Beaulieu, C.F., Rubin, D., Napel, S.: A Comprehensive Descriptor of Shape: Method and Application to Content-Based Retrieval of Similar Appearing Lesions in Medical Images. *Journal of Digital Imaging*, 1–8 (2011)
3. Napel, S., Beaulieu, C., Rodriguez, C., Cui, J., Xu, J., Gupta, A., Korenblum, D., Greenspan, H., Ma, Y., Rubin, D.: Automated retrieval of CT images of liver lesions on the basis of image similarity: method and preliminary results. *Radiology* 256(1), 243–252 (2010)
4. Mougiakakou, S., Valavanis, I., Nikita, A., Nikita, K.: Differential diagnosis of CT focal liver lesions using texture features, feature selection and ensemble driven classifiers. *Artificial Intelligence in Medicine* 41, 25–37 (2007)

5. Duda, D., Kretowski, M., Bezy-Wendling, J.: Texture Characterization for Hepatic Tumor Recognition in Multiphase CT. *Biocybernetics and Biomedical Engineering* 26(4), 15–24 (2006)
6. Ye, J., Sun, Y., Wang, S.: Multi-Phase CT Image Based Hepatic Lesion Diagnosis by SVM. In: 2nd International Conference on Biomedical Engineering and Informatics, pp. 1–5 (2009)
7. Quatrehomme, A., Hoa, D., Subsol, G., Puech, W.: Review of Features Used in Recent Content-Based Radiology Image Retrieval Systems. In: Proceedings of the Third International Workshop on Image Analysis, pp. 105–113 (2010)
8. Deepa, S.N., Devi, B.A.: A survey on artificial intelligence approaches for medical image classification. *Journal of Science and Technology* 4(11), 1583–1595 (2011)
9. Unser, M.: Sum and Difference Histograms for Texture Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8(1), 118–125 (1986)
10. Laws, K.I.: Textured Image Segmentation. PhD thesis, University of Southern California (January 1980)
11. Witten, I.H., Frank, E., Hall, M.A.: CHAPTER 10 Introduction to Weka. In: *Data Mining*, 3rd edn., pp. 403–406. Morgan Kaufmann (2011)
12. Platt, J.: Fast Training of Support Vector Machines Using Sequential Minimal Optimization. In: *Advances in Kernel Methods - Support Vector Learning*. MIT Press (1998)
13. Nino-Murcia, M., Olcott, E., Jeffrey, R.J., Lamm, R., Beaulieu, C., Jain, K.: Focal liver lesions: pattern-based classification scheme for enhancement at arterial phase CT. *Radiology* 215, 746–751 (2000)