

## DETECTION OF RENAL CYSTS BY A METHOD OF SVM CLASSIFICATION

Mekhaldi Nadia, Benyettou Mohamed

Laboratory Modeling and Optimization of Industrial Systems: LAMOSI,  
Faculty of Computer and Mathematical, Computer Science Department, USTO,  
B.P. 1505 El M'Naouer 31000 Oran – Algeria

**ABSTRACT:** During the last years the medical imagery became a vital part for detection, diagnostic and treatment of cancer. In the majority of the cases the medical imagery is the first stage in the prevention of certain diseases of cancer. The medical imagery allows a fast treatment, an effective storage and a rapid transmission of the images for diagnostic on remote sites. In the current work, we present a system for processing CT images and MRI kidney, which is based on the Split and Merge method for segmentation step, and application of SVM (Support Vector Machine) for detecting abnormalities in these images.

**KEYWORDS:** Images scanning's, Split & Merge, segmentation, SVM, lesions.

### 1) Introduction

Computer analysis of medical images can not only help doctors in the diagnosis of diseases, but it is considered as a powerful tool for therapy as well. Medical imaging has long evolved from exploratory discipline centered on diagnosis, discipline making in situations of growing, of supported comprehensive patient including interventional and therapeutic activities.

In recent years, new methods of learning are developed on the basis of Statistical Learning Theory (Statistical Learning Theory) of Vapnik and Chervonenkis. One of these methods is called Support Vector Machine. SVM were quickly adopted for their ability to work with large data, the small number of parameters they have, the fact that they are well founded theoretically, and for the good results they generate in practice.

### 2) Medical imaging and renal cysts

The importance of medical imaging is first the fact that an image is focused much efficacious more than text information or a verbal explanation.

The interpretation of medical images is one of the most promising areas of research, as it provides facilities for diagnosis and treatment decisions of a large number of diseases such as renal cyst.

### 3) Segmentation

Image segmentation is the heart of any vision system, which is an important step in the process of image analysis. Segmentation aims at extracting visual cues in an image. It allows to partition an image into sets of points called homogeneous regions for one or more characteristics (intensity, color, texture ...) and are different for at least one of the characteristics of the neighboring regions

Formally, the segmentation of a digital image  $I$  to homogeneity using a predicate  $P$  is commonly defined as a sheet.pamela f

$S = R_1, R_2, R_3, \dots, R_n$  as:

1.  $I = \cup R_i, i \in [1 \dots n]$
2.  $R_i$  is related,  $\forall i \in [1 \dots n]$
3.  $P(R_i) = \text{real}, \forall i \in [1 \dots n]$
4.  $P(R_i \cup R_j) = \text{false} \forall i \neq j$

Thus, a variety of segmentation techniques is presented in the literature; can be grouped into three main groups: methods of contour segmentation, the region segmentation methods and classification methods.

We chose a method of segmentation by region approach "Split and Merge" which is also known as "merger & division" for the performance of this method in image processing. The split & merge algorithm, was first introduced in 1974 by Horowitz and Pavlidis. This algorithm consists of two steps. In the first step, we will use the operation "Split" (division) using the quadtree structure to partition the image. Once the image is partitioned, we can proceed to the second step of the algorithm: the operation "Merge" (fusion) [HP77]

#### 3.1) Split & Merge

##### Algorithm Split and Merge

Let  $X_i, i = 1, \dots, N$  the set of regions forming a partition of the image  $I$  (we can have  $N = 1, X_1 = I$ ). Let  $H$  be a predicate of homogeneity [Sam80].

**Step 1 :** SPLIT (recursive algorithm)

For any region  $X_i$   
if  $H[X_i]$  So wrong Split ( $X_i$ )

**Step 2 : MERGE** For any couple ( $X_i, X_j$ )  
If Voisin ( $X_i, X_j$ ) and  $H[\text{Union}(X_i, X_j)]$   
Then Fusion ( $X_i, X_j$ ).

**3.1.1 Phase of Split (Division)**

The algorithm can be summarized in the following steps:

1. Divide the original image into 4 regions (see figure 1) where at  $H$  (pictures) = False.
2. If for any region  $R_i$ , the predicate of homogeneity  $H$  is not checked. ( $H(R_i) = \text{false}$ ), then divide by the area 4.
3. Go to 2).

The algorithm stops when all regions satisfy the predicate  $H$ .

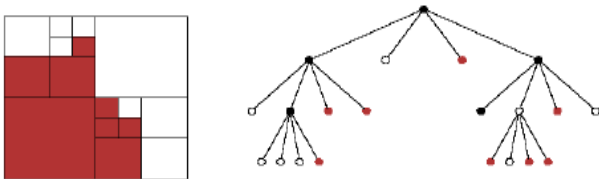


Figure 1. Structure quadtree

**3.1.2 Phase of Merge (fusion)**

The algorithm starts with an image-segmented, then:

1. Merge any two adjacent regions that satisfy a homogeneity criterion.
2. Define a predicate  $\text{Merges}(R_i, R_j)$  where  $R_i$  and  $R_j$  are two adjacent regions.
3. This method is based on the labeling. It allows you to assign the same value to all the pixels belonging to the same object, while it applies to binary images resulting from the division quadtree.
4. Labeling is characterized for its connection i.e. 4 or 8 (see figure):

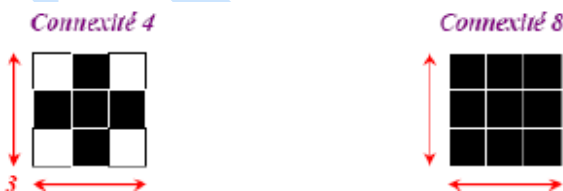


Figure 2. Principe of the labeling

**4) Running principle of SVM**

**4.1) Basics: Hyperplane, margin and Support vector:** For two classes of examples, the goal is to find a SVM classifier that will separate the data and maximize the distance between these two classes. With SVM, this classifier is a linear classifier called hyperplane (see figure 3).

In the diagram below, a hyperplane that separates the two sets of points is determined.

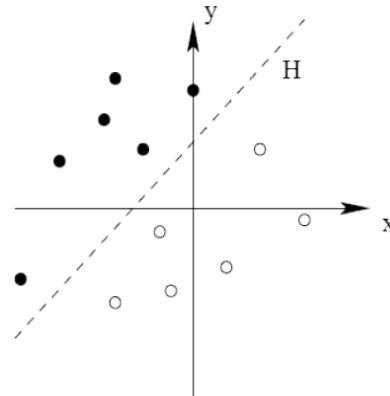


Figure 3. Hyperplane

Closest points, which are used only for determination of the hyperplane are called support vectors.

Obviously there are a multitude of hyperplane valid but the remarkable property of SVM is that the hyperplane must be optimal. We'll search through more valid hyperplanes one that passes "middle" points of the two examples classes. Intuitively, this amounts to finding the hyperplane the «safest». Indeed, suppose an example was not perfectly describes, a small change will not affect its classification if its distance to the hyperplane is big. Formally, this amounts to finding a hyperplane with the minimum distance to the training examples is maximized.

This distance is called "margin" between the hyperplane and the examples. The optimal separating hyperplane is the one that maximizes the margin. As we seek to maximize the margin, we speak about separators wide margin. [MF06]

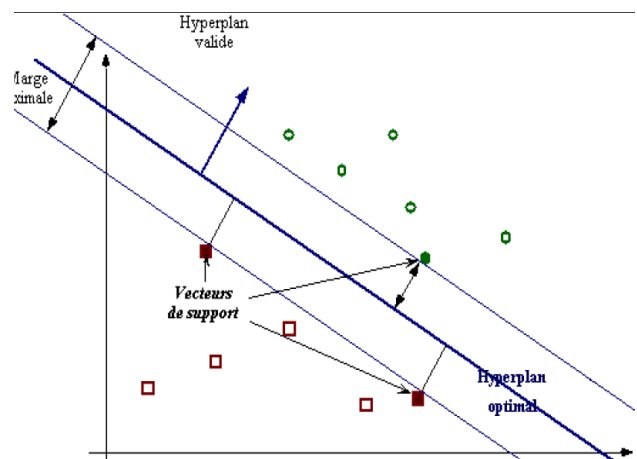


Figure 4. Representation of the basic concepts of SVM

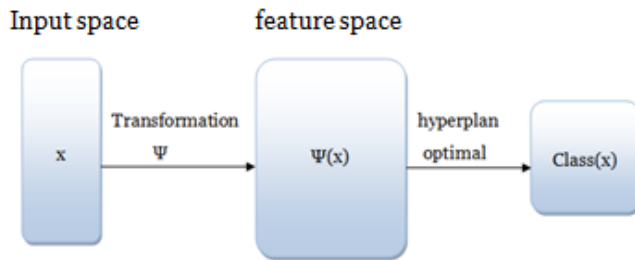


Figure 5. Principal technique SVM

### 5) The interpretation system of medical images

The system that we have developed contains five essential steps are:

- Preprocessing the images
- Segmentation
- Feature extraction regions
- Activating the SVM
- The classification (the image determining class)

#### 5.1 Database

We worked on images procured from two sources, images downloaded from the internet: <http://radiopedia.org> and digitized photographs of medical imaging clinic patients with various types of anomalies.

#### 5.2 Preprocessing Images

After the acquisition of brain medical images, we pass preprocessing is done in two steps:

**Image resizing:** The purpose of this step is to make all the images to be processed in the same size 256 \* 256 so that it is easy to handle by the segmentation method that uses square footage.

**Median filter:** The median filters are a type of low-pass filters whose principle is to do the median values of neighboring pixels.

We conclude that overall, the median filter (see figure 6) has better fidelity to the original image. This filter gives excellent results on noise "salt and pepper". In addition, it preserves the contours.

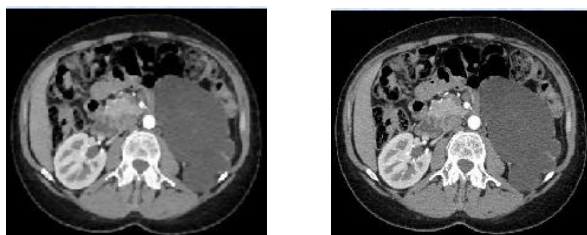


Figure 6. Median filter

#### 5.3 Segmentation

Medical images are segmented by split & merge method detailed in section (3.1) The predicate of homogeneity for a region R is based on two criteria:

1. The average gray level of the region R is lower than a threshold.
2. The variance of the gray levels in the region R is greater than a threshold.

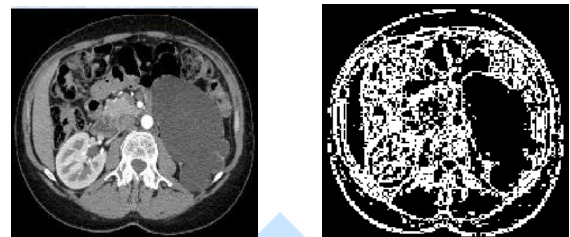


Figure 7. Segmentation

#### 5.4 Extracting characteristics of the segmented image

The choice of attributes to characterize a region is a difficult problem which requires the whole experience of catering images

This choice depends on the image processing and problem solving. From the results of the segmentation 'split & merge' applied to the images, we get a map of homogeneous regions where the pixels of each region are a single value characterizing the region.

To calculate the attributes of each region of the image, we look for the values of the gray levels of each pixel of the region obtained by the segmentation in the original image. These attributes are defined as following:

- **Surface (sur):** is the number of pixels of the region.
- **Variance (var):** is the homogeneity criterion calculated as following:

$$var = \frac{\sum_{i=1}^{Nbp} (Ngpi - Ngm)^2}{sur}$$

with Ngpi: Grayscale pixel i of the region R.

Ngm: The average grayscale R.

- **Perimeter (Per):** the length of the boundary of the region or the number of pixels forming the border.
- **Mean gray level (Ngm):** This is the average of the gray levels of pixels in the same region.

$$Ngm = \frac{\sum \text{niveaux de gris de la région}}{sur}$$

- **Elongation (Elg):** determining the rectangle excircle regions. The elongation of the regions is likened to that of the rectangle (the length to

$$\text{width ratio } L I). \quad Elg = \frac{L}{I}$$

- **Compactness (Com):** also called circularity

$$Com = \frac{4\pi \text{sur}}{\text{per}^2}$$

factor defined by:

With these attributes, each image is represented as a matrix of size N \* M.

N: the total number of resulting regions of the segmented image and M is the number of attributes that characterize each region, with the following sequence (On, Var, Ngm, Elg, Com). [MF11]

Here is a sample extraction characteristic of a region:

**Table 1. Sample Table**

Num region	Surface	Variance	Mean gray level	Elongation	Compactness
1	4	1.5840e+04	188.75	1	12.5664
2	11888	4.1588e+03	115.9671	1	2.38745e+07
3	8	48.4107	95.125	0.3333	0.1963
4	4	37.5833	88.25	1	0.7854
5	4	32.25	91.75	1	0.7854
6	4	27	97.50	1	0.7854
7	4	55.3333	62	1	0.7854
8	20	119.7368	62.5	1.6667	0.0559
9	4	35.6667	61.5	1	0.7854

**5.5 The activation of SVM**

This step allows you to enter the following parameters:

- Files learning.
- Test files.
- The type of kernel to use.
- Kernel parameters.
- The upper bound of the Lagrangian coefficients C.

The results are:

- Number of subregion in the base.
- Learning rate.
- Learning time.
- Rate test.
- Test time.
- The number of misclassified region.

**Some kernels:**

A kernel characterizes the degree of similarity or correlation between the two elements. It has a high value for two vectors that are similar and low for two different vectors. Most conventional cores are:

- **Linear kernel** :  $K(x, x') = x \cdot x'$ ,
- **Polynomial kernel** :  $K(x, x') = (x \cdot x' + 1)^q$  more the degree q is high , more the shape of the hyperplane is complex
- **RBF kernel (Radial Basis Function)**

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{\sigma}\right)$$

$$\frac{\|x - x'\|^2}{2\sigma^2}$$

- **Gaussian kernel** :  $K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right)$ . [Zam08].

**6) The practical results**

The processed images are sized 256 \* 256 and segmented by Split and Merge method. Here are some results with different kernels used:

**a) Gaussian kernel**

$$K(x,y)=\exp(-\|x-y\|^2/2 \sigma^2)$$

The following table shows the results obtained with the Gaussian kernel:

**Table 2. Results of gaussian kernel**

C	P1	Learning Rate	Learning time	Rate test	Test time	Error rate	Nbr misclassified regions
1000	2	81.52	2.54	72.94	0.8	27.05	23
1000	1	34.89	0.76	27.05	0.6	72.94	62
1000	3	88.56	0.78	99.41	0.9	10.58	9
1000	4	89.14	0.9	90.58	1	9.41	8
100	5	89.14	0.4	90.58	0.9	9.41	8

**b) Linear kernel**

The following table (give the number of table, in case it move away) shows the results obtained with the linear kernel:

**Table 3. Results of linear kernel**

C	P1	Learning Rate	Learning Time	Rate test	Test time	Error rate	Nbr of misclassified regions
1000	1	87.09	0.5	89.41	0.5	10.58	9
1000	2	87.09	0.7	89.41	0.5	10.58	9
1000	3	87.09	0.4	89.41	0.9	10.58	9
100	1	87.09	0.6	89.41	0.5	10.58	4

**c) Polynomial kernel**

The results obtained by the polynomial kernel are less good compared to cues obtained by Gaussian and linear kernel. (say more, for example why they are not good, etc.)

**Table 4. Results of polynomial kernel**

C	P1	Learning Rate	Learning time	Rate test	Test time	Error rate	Nbr of misclassified regions
1000	1	80.09	0.5	89.41	0.4	10.58	9
10	1	80.09	0.7	89.41	0.1	10.58	9

**d) RBF**

**Table 5. Results of RBF**

C	P1	Learning time	Learning test	Test rate	Test rate	Error rate	Nbr misclassified regions
1000	1	99.41	2	74.11	1	25.8	22
1000	2	99.12	1.8	74.11	1	25.8	22
1000	3	98.53	2	72.94	2	27.05	23
1000	4	93.54	2	72.94	2	27.05	23
100	1	99.41	2	74.11	2	25	22

**e) Sigmoid**

**Table 6. Results of sigmoid**

C	P1	Learning time	Learning test	Test rate	Test rate	Error rate	Nbr misclassified regions
1000	1	89.14	0.6	90.58	0.5	9.41	8
1000	2	60.41	0.7	4.7	0.7	95.29	81
100	1	89.14	1	90.58	0.7	9.91	8
10	1	89.14	1	90.58	0.7	9.91	8

Results noticed in the experimental part, are obtained by balancing between multiple parameters for support vector machines (SVM).

On supports vector machines, the value of the regularization parameter C is a parameter that governs the performance of SVM. This parameter is used to set the regularization between the minimization of the learning error and maximizing the margin.

**Conclusion**

In this work we presented the different phases used to lead to the creation of a medical image analysis system, and the different approaches used in order to do our work, and the practical results obtained. After various pretreatments on our database, we decided to apply a segmentation and classification by SVM. According to our results with different kernels, we can conclude that the best results are with the sigmoid and Gaussian kernels in learning rate, and the test time too.

We showed examples of successful SVM, and suggest some ways to go further:

- A large data base
- A comparison with other methods.
- Hybridization with other optimization methods
- A passage of minimizing the overall error in the error minimization local SVM individual

**References**

[BB07] **Berrichi Fatima Zohra, Benyettou Mohamed** - *L'Interprétation des Images Médicales*

*Cérébrales segmentation par L'approche Split et Merge*, CEUR-WS / Vol-547/23 (2007).

[HP77] **Horowitz S.L., Pavlidis T.** - *Picture segmentation by a directed split and merge procedure*, Proceedings of the 2<sup>nd</sup> International Joint Conference on Pattern Recognition, Copenhagen, Denmark (1977).

[MF06] **Mohamadally H., Fomani B.** - *SVM: Machines a Vecteurs de Support ou Separateurs a Vastes marges*, Versailles St Quentin, France janvier (2006).

[MF11] **Mekhaldi Nadia, Frida Yasmina** - *Application du classifieur pour la détection des problèmes rénaux*, Projet de Fin d'Etude, University of science and technology Oran (2011).

[Sam80] **Samet H.** - *Region representation: quadtree from binary arrays*, Computer Graphics and Image Processing, (1980).

[Zam08] **Zammit Oliver** - *Détection de zones brûlées après un feu de forêt à partir d'une seule image satellitaire spot 5 par techniques SVM*, These of doctorat. Page 38, (2008).