A new scheme for automatic classification of pathologic lung sounds

Fatma Ayari[1], Mekki Ksouri [2] and Ali Alouani [3]

[1,2] University of Tunis Elmanar, Electrical Engineering Department, National School of Engineers Tunis (ENIT), B.P.342, Le Belvedere, 1002 Tunis, Tunisia

[3] Tennessee Technological University Electrical Engineering Department, Po Box 5004, Cookeville, TN 38 505, USA

Abstract

In this paper, a classification scheme has been proposed to classify crackles based on waveform features and frequency domain features. This purpose is very important in the analysis of respiratory disorders. In fact, morphological characters of crackles can be well represented by time amplitude distribution. Thus, they convey significant diagnostic information, for their precise timing in the respiratory cycle, their repeatability, and shape all mightily correlate with pulmonary diseases. The ability to analyze the acoustic patterns of these breathing-induced phenomena will enhance the expertise of the physiology and pathophysiology of respiratory disorders that can be very useful in clinical considerations.

Keywords: *Lung sound, Classification, Crackles Fuzzy Logic, Fine, Coarse, Medium.*

1. Introduction

Crackles have been the most adventitious lung sound treated by numerous researchers. In fact, due to their transitory character and difficulty to separate them from noise in original lung sound, different theories and approaches were investigated along the last decades. In fact, computerized analysis of lung sounds have been a longstanding technique that has included a large verity of methodologies and procedures presented by experts in this field. However, the most important earliest work is due to Gavriely et al. [1]; they grasp the versatility of lung sound analysis by revealing additional useful information needed to diagnose a patient with a respiratory disease. Murphy et al [2]. Holford et al [3] and Piiril et al [4] have proposed empirical rules defining crackles as remarkable characteristic of multiple lung diseases.

A normalisation of a large number of features that can parameterize the lung crackles have been established via CORSA. Nevertheless, we need still to investigate in this field because of great correlation between many lung diseases and crackles and also, till nowadays, statistics about their identification and estimation during a respiratory signal needs additional homogenisation and standardisation. Also it is worthy to note that crackles have moving character during time. The pitch, spectrum, and timing of the crackles are different in each species of disease. In addition, the precise number of crackles can indicate the severity of a disease [4]. Crackles also have a short duration and low intensity that is inaccessible to the human ear without computerized analysis that can provide precious information in the early detection of lung pathologies.

Yeginer and Kahya [5] have used a wavelet network to characterize crackles within a lung sound signal by optimizing two weight factors, scaling, time-shifting, and frequency parameters. A fine/coarse crackles identification based on probabilistic based rules has been established. The most recent work of M Bahoura [6] has been focused on an automatic system for crackles extraction and classification. A high separation rate has been found; and this filter was justified to be very advantageous with less signal distortion. In this paper we will detail our methodology for crackles detection and classification in the case of two pulmonary diseases. Crackles detection and extraction have been achieved using a Hilbert-multiscale product algorithm that is well described by Ayari et al [7]. We will be focused in this paper on a new crackles classification scheme to enhance the morphology of crackles detected in some pulmonary diseases and to ascertain their real characteristics. This scheme is based on two complementary approaches: the first one is based on a statistic method and the second one is based on fuzzy logic non linear classifiers. It is shown that those two approaches have lead to an automatic decision scheme for the classification of two pulmonary diseases with a high accuracy.

IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 1, July 2012
ISSN (Online): 1694-0814
www.IJCSI.org

2. Crackles parameters

According to the study of Murphy et al. [8], different parameters have been used to characterize and classify crackles, mainly the initial deflection width (IDW) which is the duration of the first deflection of the crackle and the two-cycle duration (2CD) which is the duration of first two cycles of the crackle. In the study of Hoevers and Loudon [9], four parameters have been utilized to classify crackles: mainly, largest deflection width (LDW_1) which is the largest deflection of the crackle and the widths of its first three rights and left neighbours denoted by (LDW_2, LDW_3 and LDW_4). Those parameters are denoted by (Fig. 1). Homma and Matsuzakill [10], and Mitsuru Munakat & al [11] have introduced also time expanded waveforms: 1/4 cycle duration of crackles waveform, and 9/4 cycle duration of crackle waveforms, denoted by T_1 and T_2 in Fig. 2. Thus, it was demonstrated in their studies that such features can be the best basics to distinguish between fine and coarse crackles.

To bring more precision to the crackles distinction and then classification, we have choice T_1, T_2, IDW, 2CD, LDW_1 to LDW_3 which are described previously and also the maximum amplitude in the crackle waveform denoted by Amp. The maximum peak frequency extracted from the spectral frequency analysis of the crackle waveform denoted by PF is also used as a frequency domain crackle feature. A part of those features was used by standard definitions to classify crackles into fine or coarse. In fact, according to the American Thoracic Society (ATS) [12], the mean durations of initial deflection width (IDW) and two-cycle duration (2CD) of coarse crackles are specified to be 1.5 and 10 ms, and those of fine crackles are 0.7 and 5 ms, respectively. Thus, according to Computerized Respiratory Sound Analysis (CORSA) definitions, a coarse crackle is defined with 2CD >10 ms [8, 9], and a fine crackle has 2CD <10 ms. In the review of pulmonary analysis [13] it is assumed that a fine crackle is defined with deflection width (IDW) and two cycle duration (2CD) of 0.92 ms, and 6.05 ms respectively. But coarse crackles were considered to be designated with IDW and 2CD > 1.25 ms and > 9.23 ms respectively.

It is aimed in the present study to develop an algorithm that allows calculating a set of crackles features that we have chosen to detect and classify crackles and also to count their number in a breathing phase (inspiration or expiration). The algorithm build in this study involves two purposes which are developed simultaneously.

The first purpose is based on time-amplitude response analysis of the lung signal and the second one involves fast Fourier transform analysis of the lung signals. This issue was successfully obtained based on the following detailed description and also illustrated with the flowchart presented by Fig. 3.

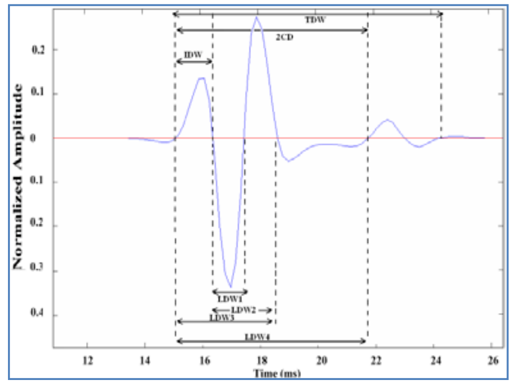

Fig. 1 A typical crackle waveform: four parameters, largest deflection widths ($LDW_{1\text{-}4}$)

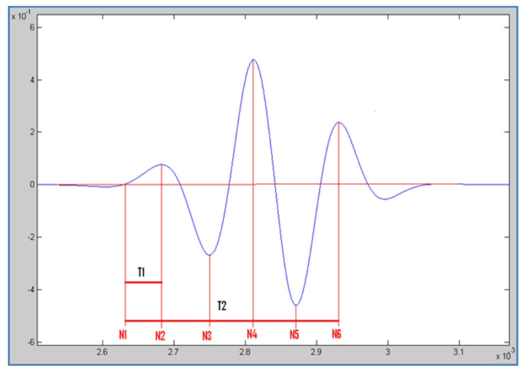

Fig. 2 A typical crackle waveform: two parameters T_1 and T_2 defined by Holford [3]

The first purpose is based on time-amplitude response analysis of the lung signal and the second one involves fast Fourier transform analysis of the lung signals. This issue was successfully obtained based on the following detailed description and also illustrated with the flowchart presented by Fig. 3.
Thus, the precise starting points of crackles are indicated by locations of peaks of LDW_1 in time-amplitude distribution which is used as the reference point for selection of attracts zones AZ, and the wave that follows the starting point is manifested in time-amplitude distribution which is illustrated by Fig. 1. Thus, we obtain relative narrow band waveform representing the oscillating pattern of crackles. This reference point renders the automatic processing of the crackle possible once it is detected Fig. 4. The second extremum before and the third extremum after the peak of LDW_1 are appointed as the start–end points of a region that includes 2CD and are exemplified by Fig. 1. The crossing zeros and the first derivative of the crackles are calculated via a Matlab script to fix the 1/4 cycle duration and 9/4 cycle duration parameters T_1 and T_2 Fig. 5. Then, a feature subroutine is called to calculate wave form features (T_1, T_2, IDW, 2CD, $LDW_{1\text{-}3}$, Amp) and also to count the number of crackles in a respiratory phase. Crackle wave forms can be therefore plotted Fig. 6

and saved to be used for frequency analysis Fig. 7. It is outstandingly noted that an intermediate step of identification is used to ascertain that the captured depicted crackle wave does not contain a vesicular sound. This step was achieved using a subroutine that computes the maximum error between the wave form and a Matlab smooth function, Fig. 8. Almost vesicular sounds tested with this subroutine have an estimated error ranging from 0.8 to 1.5; so that they could be distinctly detected and separated from crackles waves Fig. 9.

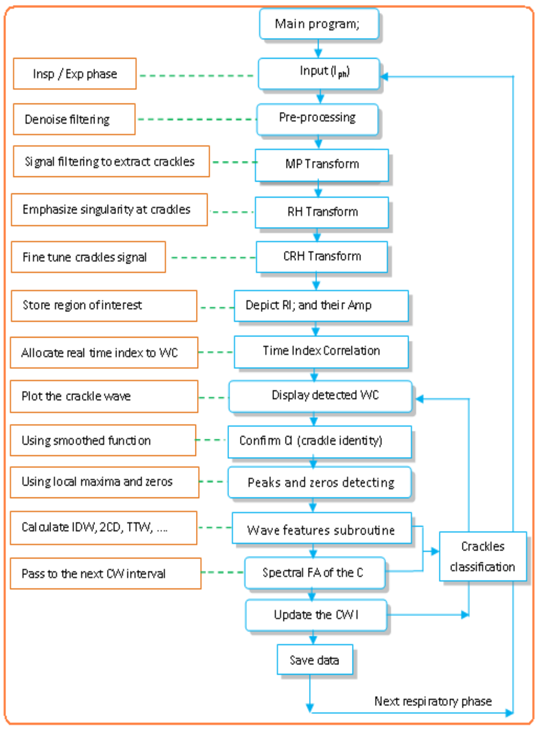

Fig. 3 Flowchart of the automatic crackles detection and classification

Peak detector subroutine was used to allow the separation of peaks associated with crackles. Also, a zero crossing subroutine was used to identify the crossroads between the crackle wave and the time axis, so that it is possible to define precisely all wave features Fig. 7. We think that those features can be very useful in crackles classification step, as they describe crackles pattern that are strictly dependent on the physical morphology of the crackle.

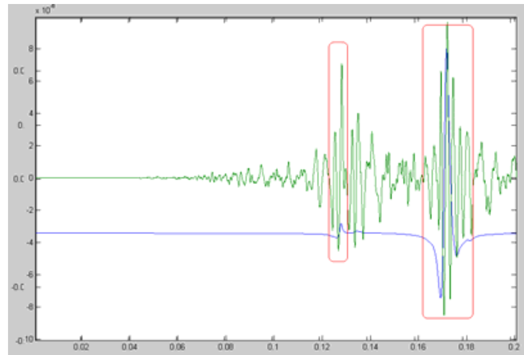

Fig. 4 crackles detection and extraction

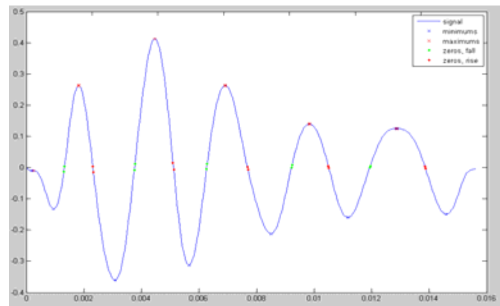

Fig. 5 peaks and zeros detection of the with features crackles calculation. : MDW = 0.0019; TTD = 0.0172; IDW = 9.9773e-004; MCD = 0.0014; TMA = 0.7742; NCD = 5; FQD = 4.9887e-004; NFCD = 0.0031.

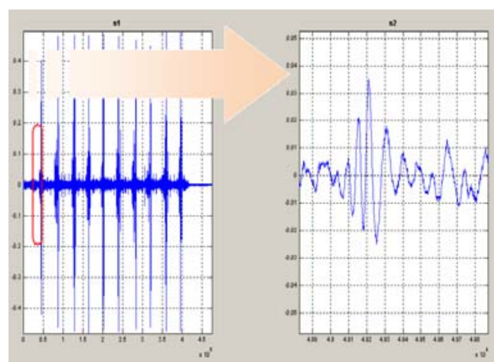

Fig. 6 crackles are plotted to be used in frequency analysis

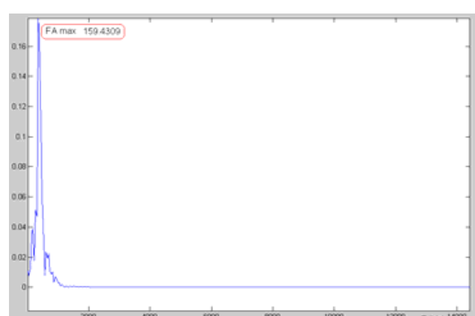

Fig. 7 Frequency spectrum domain of the considered crackle

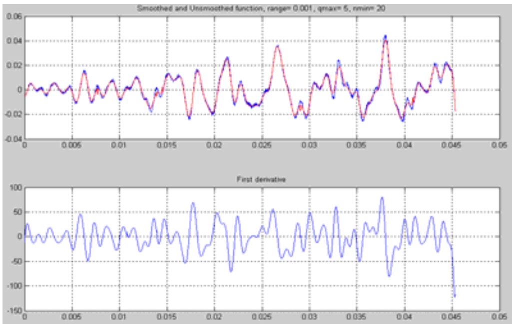

Fig. 8 Identification of the crackle with smooth function (identification max error e = 0.02 %)

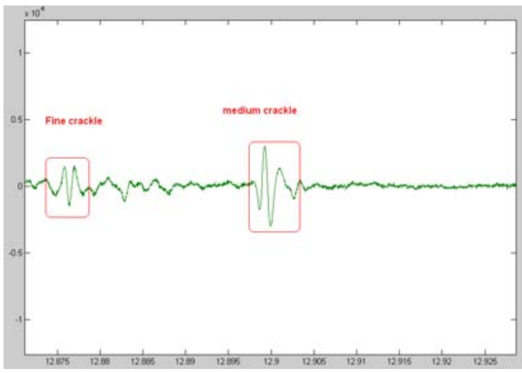

Fig. 9 Depicting and identifying crackles

At this stage the entire respiratory signal is ready to the last phase of crackles depicting, identification and count automatically. The classification of all detected crackles is based on the assessment of the different computed features.

3. Automatic crackles classification

In this section, we are interested with crackles classification using different approaches for data obtained from various databases. In fact, data was collected from Wilkins [14], Lehrer [15], LARALE [16] and our own build database with patients from Charle Nicole Hospital in Tunisia. We have studied 10 patients selected from 15 patients with pulmonary fibrosis diagnosed at Charle Nicole Hospital. The diagnosed patients have similar criteria such as progressive dyspnoea without airway obstruction. Also we have studied 10 patients with chronic bronchitis with sputum production. All patients had cough and sputum for at least 3 months. Our database was build with lung signals recorded using a Littman electronic stethoscope Fig. 10, throughout full inspiration and full expiration at the right posterior chest wall in order to reduce the effect of heart sounds interferences. Then, data was saved in the stethoscope and transmitted to the computer via infrared device. Data was firstly analyzed using the Littman software [17], and then a pre-processing analysis was conducted to reduce noise effect.

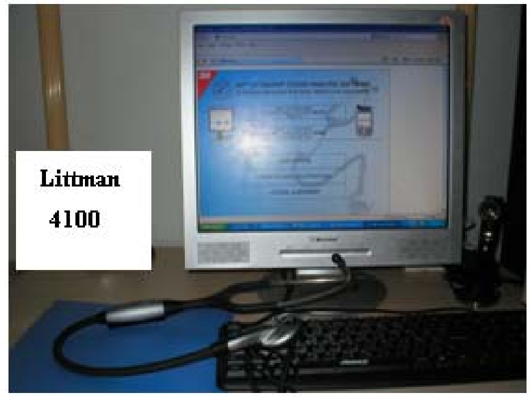

Fig. 10 Littman stethoscope S4100

The different sounds considered from the other databases are heard over the normal chest. Breathing cycle is divided into two stages: inspiration and expiration. Each segment is characterized by rich information (waveform, timing, number and regional distribution of crackles). These segments, in each cycle, are further partitioned into six phases, namely, early, mid, late inspiration and expiration phases. These phases are defined according to the inspired and expired air volume during respiration.

In the case of data collected using a digital Littman Stethoscope, the time duration of recorded signal including multiple respiratory phases was between 30s and 60s, which is a period covering up to 3 to 6 respiratory cycles. After each recording, the data was transferred via an infrared USB device to the computer to be analyzed. All recorded files were first transferred to a lung sound analyzer Littman software after being converted to a digital wave format. The pre-processing procedure was built using Matlab signal processing toolbox. Lung sounds were displayed within waveforms or frequency amplitude domain presentations. Preceding the recording lung sound phases, all patients were identified by name, age, sex and type of pathology. Measurements of different sounds were established by auscultation with the assistance of specialized clinicians. The verification of the number of crackles existing in each respiratory phase was established with the presence of experts in the crackles domain and we have counted crackles from the time expanded waveforms included in a particular phase from each recorded respiratory sound. In fact, the characteristics and wealth of crackles is analogous in every cycle assimilated to a given lung pathology. Total number of crackles is counted from 10 inspiration or expiration phases. Feature vectors, extracted from segments of respiratory cycles from different patients are jointed into a feature set. In fact, the necessity to divide a cycle into phases rises because of the non stationary nature of the data within a cycle. Such a division of a respiratory cycle is consistent with the auscultation terminology since all phases are usually considered as the most informative and distinctive parts of a respiratory sound signal. Thus, the total number of measured feature vectors available in a phase, e.g., in early inspiration, equals to the number of features times the total number of patients. After separation of a cycle feature set into phase features sets, each phase is classified separately. In a second step we have applied this algorithm on signals involving fine and coarse crackles from various data bases and results are presented in Table 1. Where TP, FP and FN are the number of the true positive, false positive and false negative respectively [18]. SG11_C1, SG11_C2, SG11_C3, SG11_C4, are the referred patient with the first, second, third and fourth cycles for the first group of pulmonary disease. The other patients in the two pulmonary diseases are denoted similarly. The true positive indicates the number of crackles correctly detected; false positive indicates the number of the false detections and false negative specifies the number of crackle missed by the developed algorithm. One can notice the versatility of this method.

Table 1

Pat_Grp1	Crackles per cycle	real crackls per patient	TP	FP	FA	Pat_Grp2	Crackles Per cycle	real crackls per patient	TP	FP	FA
SG11_C1	4					SG21_C1	8				
SG11_C2	4					SG21_C2	5				
SG11_C3	4					SG21_C3	5				
SG11_C4	2	14	12	2	0	SG21_C4	5	23	22	1	0
SG12_C1	5			0	0	SG22_C1	5			0	0
SG12_C2	4			0	0	SG22_C2	5			0	0
SG12_C3	5			0	0	SG22_C3	4			0	0
SG12_C4	5	19	21		2	SG22_C4	5	19	20	0	1
SG13_C1	4			0	0	SG23_C1	6			0	0
SG13_C2	4			0	0	SG23_C2	5			0	0
SG13_C3	5			0	0	SG23_C3	5			0	0
SG13_C4	5	18	18	0	0	SG23_C4	8	24	24	0	0
SG14_C1	4			0	0	SG24_C1	7			0	0
SG14_C2	4			0	0	SG24_C2	8			0	0
SG14_C3	4			0	0	SG24_C3	8			0	0
SG14_C4	5	17	17	0	0	SG24_C4	8	31	30	1	1
SG15_C1	6			0	0	SG25_C1	6			0	0
SG15_C2	4			0	0	SG25_C2	9			0	0
SG15_C3	6			0	0	SG25_C3	4			0	0
SG15_C4	5	21	20	1	0	SG25_C4	2	21	22	0	1
SG16_C1	5			0	0	SG26_C1	5			0	0
SG16_C2	5			0	0	SG26_C2	5			0	0
SG16_C3	5			0	0	SG26_C3	5			0	0
SG16_C4	5	20	21		1	SG26_C4	5	20	22	0	2
SG17_C1	5			0	0	SG27_C1	5			0	0
SG17_C2	5			0	0	SG27_C2	5			0	0
SG17_C3	4			0	0	SG27_C3	4			0	0
SG17_C4	5	19	18	1	0	SG27_C4	5	19	19	0	0
SG18_C1	4			0	0	SG28_C1	4			0	0
SG18_C2	5			0	0	SG28_C2	5			0	0
SG18_C3	5			0	0	SG28_C3	5			0	0
SG18_C4	8	22	22	0	0	SG28_C4	8	22	21	1	0
SG19_C1	7			0	0	SG29_C1	7			0	0
SG19_C2	4			0	0	SG29_C2	8			0	0
SG19_C3	8			0	0	SG29_C3	8			0	0
SG19_C4	8	27	26	1	0	SG29_C4	8	31	30	1	0
SG110_C1	2			0	0	SG210_C1	6			0	0
SG110_C2	2			0	0	SG210_C2	9			0	0
SG110_C3	4			0	0	SG210_C3	4			0	0
SG110_C4	2	10	11	0	1	SG210_C4	2	21	20	1	0
total	*187*		*186*	*5*	*3*	*total*	*231*		*230*	*4*	*4*

3.1 Performances of crackles extraction algorithm

To evaluate the performance of the proposed algorithm of crackles detection and extraction, we have performed a comparative study that is well illustrated in references Xiaoguang et al [18], Mete Yeginera et al [19] and Hadjileontiadis [20], these two measures were indispensable to estimate either the separation rate or the noise quantification. To achieve this task, we have computed the ratios described with equations above and results are summarized in table .3. The separation rate is done with equation (1)

$$SR\ (\%) = SN / RN * 100. \quad (1)$$

Where RN is the existent number of crackles in the vesicular sounds, and SN is the correctly separated number of crackles using the developed algorithm.
To evaluate the performances of our algorithm of crackles detection and extraction, we have computed parameters suggested in [19], which are designated by: sensitivity S % and positive predictivity P%.

$$\text{Sensitivity } \% = \frac{TP}{TP+FN} * 100 \quad (2)$$

$$\text{Positive Predictivity } \% = \frac{TP}{TP+FP} * 100 \quad (3)$$

TP_G1 and TP_G2 are designating, the total number of patients belonging to first group and second group of lung pathologies, respectively. Therefore, results of Table 2 display a sensitivity of 98.34 % and a positive predictivity of 97.88 %. This method has performed a high accuracy in detecting crackles automatically for 80 respiratory cycles.

Table 2 synthesis of crackle count

		FP	FA
TP_G1	186	5	3
TP_G2	230	4	4
Total	416	9	7
	S%	98,345	
	P%	97,882	

3.2. Crackles morphology and classification

The aim of this section is crackles classification using different classification algorithms in order to assess the real behaviour of crackles morphology. This step is being extensively important to enhance the automatic lung disease classification. Data used to carry crackles classification is deduced from features characteristics defined in the previous section 2. For every feature extracted from the waveform signal, we tried to conceive a matrix with two columns. The respiratory pathologies considered herein are presented with two groups of patient diseases; pulmonary fibrosis with fine inspiratory crackles and bronchitis with sputum marked with coarse crackles both of them are

common lung disorders. The first column represents the elementary mean characteristic of the feature determined for crackles presented in the lung signal of a designated (inspiration or expiration phase) of patients with chronic bronchitis, and the second colon represents the same elementary mean characteristic of the feature crackles for patients with pulmonary fibrosis.
Classification experiments were achieved on respiratory databases described above, to test the proposed novel classification scheme, assuming that respiratory sound signals are cyclic biological data.

There is distinctly a statistical difference between groups of features belonging to both pathologies considered in this study. We have attempted to present data classification problem, from two strategies. The first one is based on statistics theory and the second involves the fuzzy logic non linear classifiers algorithms.

A matrix of features data vectors is stored as column-wise. Each column represents a feature vector among the features extracted and calculated from crackles in adventitious lung signals. Features vectors are constituted with the following features (T_1, T_2, IDW, 2CD, PF, LDW_1 LDW_2, LDW_3, and Amp) and they are calculated for both lung diseases. The feature vectors representing the same features and resulting from two different groups of lung diseases are located contiguous in the matrix data. So the vectors are defined as follows. [T1_G1, T1_G2, T2_G1, T2_G2, IDW_G1, IDW_G2, 2CD_G1, 2CD_G2, ln(PF)_G1, ln(PF)_G2, LDW1_G1, LDW1_G2, LDW2_G1, LDW2_G2, LDW3_G1, LDW3_G2, Amp_G1,Amp_G2,].
It is noted that, to overcome the problem of large gap of dimensionality between features vectors, we have substitute the peak frequency vector by its logarithm value, so that PF (peak frequency feature vector is substituted by Ln (PF). It is noticeable that all features with indices (_G1) are designating features from patients with pulmonary fibrosis and features with indices (_G2) are representing features from the second group of pulmonary disease that is chronic bronchitis. This build matrix was represented by the statistical graphic tool Tukey's box plot. In fact, the best model that characterizes, obviously central tendency parameters, and dispersion parameters is the Tukey's box plot which captures magnitude of the data series and its dispersion. Actually, we are dealing essentially with a schematic representation of statistical distribution by incorporating parameters of central tendency and dispersion; so that comparisons between different series are manifestly determined Fig. 11.
This graph shows the variation range of each group of feature values. The minimum, maximum and medium values in each vector are well emphasized, thus it is possible to distinguish between crackles that are specified to be fine or coarse. When we analyse the box plot representation, one can see the separation or overlap between two consecutive groups of features representing the same parameter for both pathologies. In fact, it is clear that medians of two consecutive boxes representing the same feature vector and revealing both pathologies are mostly well separated and we distinguish a large discrepancy between those medians.

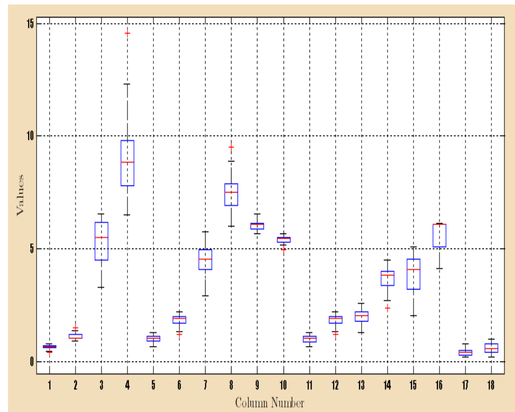

Fig. 11 box plot of all vector features and their statistical parameters. (The column of feature matrix are designated by a numbering ranging from 1 to 18, and the statistical parameters (median, first quartile, third quartile, minimum and maximum parameters, the whiskers extend to the most extreme data points of every feature vector are well indicated)

In few cases the two consecutive boxes are overlapped and have common feature values. This connotation is likely happening for the third large deflection width LDW3; (boxes at column 15 and 16) and maximum amplitude (boxes corresponding to column 17 and 18). Those vectors are sharing some feature values, so that we could have some blurriness about their membership when they are used in a clustering process. This concept can lead to two further forward-thinking. The first one which is most obvious assumes that those particular features cannot be considered alone to conclude about the classification of the crackles into fine or coarse classes. Every comparison between two similar features belonging to each group of lung disease is a complementary information included into a complex decision scheme based on the 18 features [T1_G1, T1_G2, T2_G1, T2_G2, IDW_G1, IDW_G2, 2CD_G1, 2CD_G2, ln(PF)_G1, ln(PF)_G2, LDW1_G1, LDW1_G2, LDW2_G1, LDW2_G2, LDW3_G1, LDW3_G2, Amp_G1,Amp_G2,].
The second thinking is much more like with the presence of an intermediate class of crackles that is called medium and in this case; features characterizing those crackles can be distinct from fine or coarse features definitions. To assess the trustworthiness of both sights we have established a

clustering analysis based on non linear fuzzy classifiers.

4. Fuzzy clustering analysis

It is worthy to notice that a nonlinear classifier is absolutely indispensable to create decision boundaries. There are different ways to build such nonlinear statistics based on hard classifiers. That's why we have opted for four classifiers; denoted by Fuzzy_Clust1, which is a hard partitioning method, Fuzzy_Clust2, and Fuzzy_Clust3, which are fuzzy clustering methods. Those algorithms are based on suitable distances metrics and also they can be straightforwardly used when dealing with relatively small sized data, such as the case in our study. Those fuzzy partitioning methods with different distance norms are well described in [21], [22]. It is proposed to divide the vector features data sets into subsets (called clusters), using hard and fuzzy partitioning, so that transitions between the subsets can be considered as hard or steady. An overview about the theoretical concept of those unsupervised partitioning will be defined in the following sections

4.1 Classification algorithms with fuzzy logic

The fuzzy partition can be viewed as a generalization of the partition; it can achieve real values in the interval [0, 1].
The matrix $U_{N \times c} = [\mu_{ik}]$ represents the fuzzy partition with conditions given by:

$$\mu_{kj} \in [0, 1];\ 1 \le i \le N,\ 1 \le k \le c. \quad (1)$$
$$\sum_{k=1}^{c} \mu_{ik} = 1, \quad 1 \le i \le N,$$
$$0 \le \sum_{k=1}^{c} \mu_{ik} = 1, \quad \le k \le c$$

The i-th column of the matrix U contains the values of the membership function of the i-th fuzzy set of X. the total sum of each column must be equal to 1, and therefore the number of total membership of each x_k in X is equal to one. The distribution of membership among the c fuzzy subsets is not constrained. It should be noted that in the case of a probability score, the sum of degrees belonging to a data point must not be equal to one.

4.2 Fuzzy_Clust1 algorithm

Hard partitioning methods are simple and popular, although their results are not always reliable and these algorithms have numerical problems also. From an n x N data set, the Fuzzy_Clust1 algorithm allocate each data point to one pole of the cluster c to minimize the sum of the square cluster of this group:

$$\sum_{i=1}^{c} \sum_{k \epsilon A_i}^{N} \|x_k - v_i\|^2 \quad (2)$$

Ai is a set of data objects (data points) in the i-th cluster and vi means the average of these points around the cluster i, it is really a standard distance. In Fuzzy_Clust1 algorithm, vi is called the cluster prototype, which is the cluster centre.

$$v_i = \frac{\sum_{k=1}^{Ni} x_k}{Ni} \quad (3)$$

N_i is being the number of objects in A_i.
In Fuzzy_Clust1 algorithm, the cluster centres are the objects closest to the average data in a cluster. V=$\{v_i \in X,\ 1 \le i \le c\}$. It is useful for example when each data point represents a position of a system, so there is no continuity in the data space. In this case, the average of the points in a set does not exist.

4.3 Fuzzy_Clust2 algorithm

The classification algorithm using fuzzy logic and Fuzzy_Clust2 is based on the minimization of an objective function called Fuzzy_Clust2 functional. It is defined by Dunn [21] as:

$$J(X; U, V) = \sum_{i=1}^{c} \sum_{k=1}^{N} (\mu_{ik})^m \|x_k - v_I\|_A^2 \quad (4)$$

With $V = [v_1, v_2, \dots \dots \dots, v_c], v_i \in R^n$

Is a vector of cluster prototypes (centres), which must be determined as follows:

$$D_{ikA}^2 = \|x_k - v_i\|_A^2 \quad (5)$$
$$= (x_k - v_i)^T A (x_k - v_i)$$

It is a standard of the squared distances.
Statistically, the latter distance may be regarded as a measure of the total variance of x_k from the cluster v_i. The minimization of the functional of this Fuzzy_Clust2 is a nonlinear optimization problem that can be solved using a variety of methods, ranging from the minimization of group of coordinates to the genetic algorithm. The most popular method, however, is a simple Picard iteration through the first order conditions for fixed points (stationary) in the above equation, known as Fuzzy_Clust2 algorithm.

4.4 Fuzzy_Clust3 algorithm

Fuzzy_Clust3 extended the standard Fuzzy_Clust2 by using an adaptive distance norm, in order to detect clusters of different geometrical shapes in one data set [23]. Each cluster has its own standard A_i including the matrix, which gives the distance following standard:

$$D_{ikA}^2 = (x_k - v_i)^T A_I (x_k - v_i), \quad (6)$$
$$1 \le k \le N.$$

The matrices A_i are used as optimization variables in the functional Fuzzy_Clust2, allowing each cluster to adapt the standard distance to the surface of each local topological data.
If A is the c-uple of the standard matrices including: A = (A_1, A_2,.. A_c). The objective function of the Fuzzy_Clust3 algorithm is defined by:

$$J(X;U,V,A) = \sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{ik})^m D_{ikA_i}^2 \quad (7)$$
$$\|A_i\| = \rho_i \quad , \quad \rho > 0$$

Where ρ_i is fixed for each cluster. Using the method of Lagrange multiplier, the following expression is obtained for A_i:

$$A_i = [\rho_i \det(F_i)]^{1/n} F_i^{-1} \quad (8)$$

With F_i is the fuzzy covariance matrix of the i^{th} cluster defined by:

$$F = \frac{\sum_{k=1}^{N}(\mu_{ik})^m (x_k - v_i)^T}{\sum_{k=1}^{N}(\mu_{ik})^m} \quad (9)$$

Examples of classification with those classifiers algorithms are presented by Figures 12 to 18.

Recordings from different databases described in the previous section were used, and features data analyzed are collected from 4 respiratory cycles in each patient lung sound record.
As illustrated by Figures 12 to 17, the classification performances on the amplitude feature of both considered lung diseases (chronic bronchitis and pulmonary fibrosis) are presented.

The three decision algorithms, specified by, Fuzzy_Clust1, Fuzzy_Clust2, and Fuzzy_Clust3 have been compared. For two features parameters; crackles amplitude (Amp) and (LDW3) for both groups of pulmonary diseases. The comparison carried in this paragraph is based on standard indexes that are defined as follows.

The first partition index (SC); is defined as the ratio of the sum of compactness and separation of the clusters. It constitutes the sum of individual cluster validity measures normalized through division by the fuzzy cardinality of each cluster as it is defined in reference [24]. This parameter SC is helpful when comparing different partitions that are defined with equal number of clusters. A lower value of SC indicates a better partition. Also we have used another separation index (S); which is opposite to partition index (SC), it uses a minimum-distance separation for partition validity [23]. The optimal number of clusters should minimize the value of this index. It is well noted that the only difference of SC and S is the approach of the separation of clusters. As a final point, we have used Dunn's Index (DI), it is originally proposed to identify "compact and well separated clusters". So the result of the clustering has to be recalculated as it was a hard separating algorithm. In the case of overlapped clusters the values of DI are not really unswerving because of re-division of the results with the hard partition method. The performances of those indexes are shown in Table 3.

It can be observed that all indexes have significant values. The classification index of the feature amplitude is varying in the range of (0.05 to 0.8) for all algorithms, and then in the range of (0.009 to 2.9), for the LDW3 feature. These values are known to be significantly good enough.

One can notice that some algorithms are more robust than others when there is a frugality data problem. Mean values presented in Table 3 are displaying decision clusters for partition indexes constructed after 10 runs of independent initialization parameters that are automatically generated with a random process. We have enhanced performances of different algorithms by running for every case the algorithm up to 10 times and then define the mean value of performances indexes, to unsure that results are being stationary and to minimize errors in indexes values that can be related to the randomized initialisation of calculation parameters. Furthermore, it was recommended to do this step in order to boost the performances of different algorithms. In fact, this number of runs brigs stable results with reasonable time of calculation.

The most advantage of using such fuzzy clustering algorithms is due to the kernel and distance assess used to compute the density distribution of clusters, which cannot operate well while classifying vectors with spares data.
These experiments show that the three Fuzzy_clust1, Fuzzy_Clust2 and Fuzzy cust3 algorithms have overcome this simple sized problem and have carried significant results.

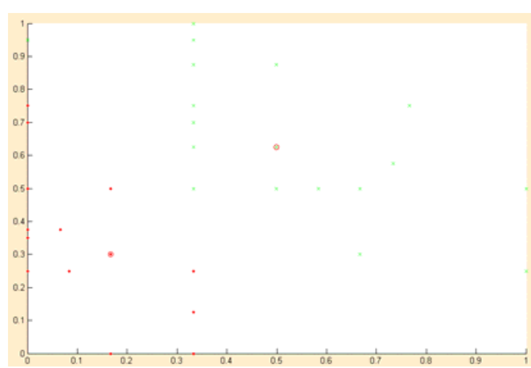

Figure 12 Fuzzy_Clust1 algorithm applied to the amplitude feature, with two clusters separation, c=2

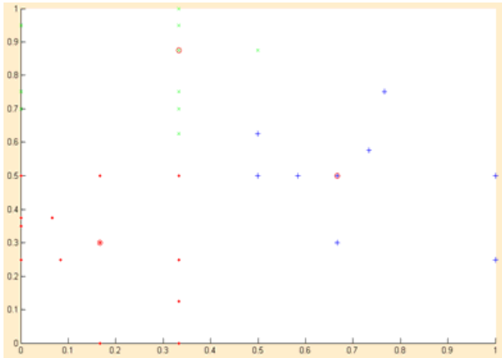

Figure 13 Fuzzy_Clust1 algorithm applied to the amplitude feature, with two clusters separation, c=3

Scatter plots indicate that there is a slight overlap between the two classes, in spite of their separate clustering tendency.

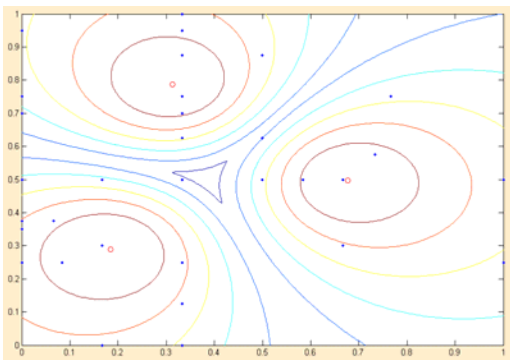

Figure 14 Fuzzy_Clust2 algorithm applied to the amplitude feature, with two clusters separation, c=3

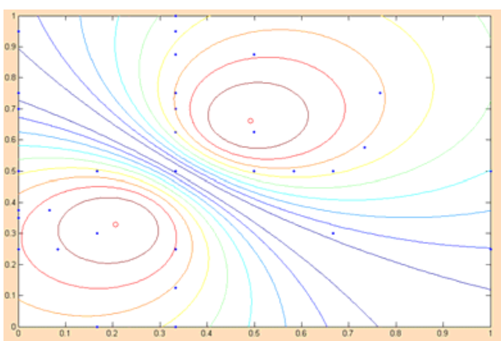

Figure 15 Fuzzy_Clust2 algorithm applied to the amplitude feature, with two clusters separation, c=2

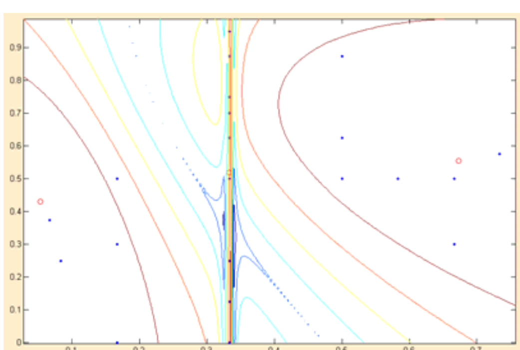

Figure 16 Fuzzy_Clust3 algorithm applied to the amplitude feature, with three clusters separation, c= 3.

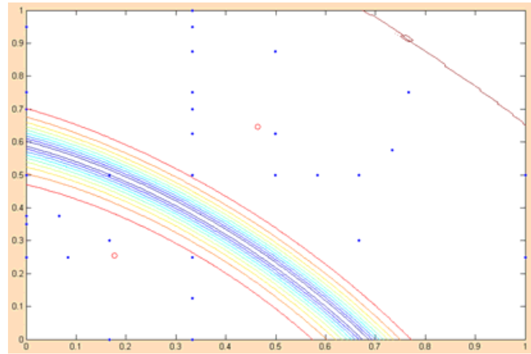

Figure 17 Fuzzy_Clust3 algorithm applied to the amplitude feature, with two clusters separation, c=2.

At this stage, it was possible to measure the clustering algorithms validity using the indices calculated and summarized in the Table 3. Those indices are defined according to the definition of functions assessing individual clusters partition.

Table 3

	Fuzzy_Clust1		Fuzzy_Clust2		Fuzzy_Clust3	
C=2	Amp	LDW3	Amp	LDW3	Amp	LDW3
S					0.0507	0.0090
SC					2.4140	2.9698
DI	0.1778	0.1114	0.1024	0.1114		
C=3	Amp	LDW3	Amp	LDW3	Amp	LDW3
S					0.0589	0.0724
SC					0.8583	0.2359
DI	0.2000	0.4550	0.2000	0.4550		

While calculating validity measure indexes to estimate the goodness of an algorithm for the data sets associated with LDW3 and Amp features, some conclusions can be emphasized. For DI index Fuzzy_Clust1 and Fuzzy_Clast2 are almost similar for both features Amp and LDW3. It is important to notice that the best results are obtained with three clusters for each of the clustering algorithms. The DI indexes are similar in the case of LDW3 for all both algorithms; Fuzy_Clust1 and Fuzzy_Clust2. In the case of Amp feature, Fuzzy_Clust1 seems to be the best classifier as we get the highest DI index.

It is concluded that with all classifiers, the data sets are presenting better performances with three clusters. Thus, according to analysis of Amp and LDW3 features, the existence of three groups of crackles is imminent. In addition, the analysis of Fig. 12, Fig. 15 and Fig. 17 show the existence an overlap between data sets of the two clusters. This fact can be explained by the existence of a third category of crackles which are identified as medium crackles, and then characterized with different feature parameters. This fact confirms results performed in reference [19]. It is concluded actually that a large statistical noteworthy difference exists between the crackles in the two pulmonary diseases analysed in this study; pulmonary fibrosis and chronic bronchitis. But, according to results of features analysis, both of diseases are characterized with different category of crackles that are statistically distinguished as described in the following section.

4.5 Synthesis of crackles classification

Using definitions of features properties defined in section 2, and features parameters established in section 3, we have computed the different category of crackles present in each group of pulmonary disease. A statistical distribution of the number of crackles classified by category for both lung pathologies can lead to the data presented in Table4 below.

In fact, this table summarises the classification results of crackles detected in the two lung pathologies; S_G1, S_G2 designating the data collected from patients belonging to first group and second group of lung pathologies, respectively. The first group is corresponding to pulmonary fibrosis and the second one is corresponding to chronic bronchitis. NFC, NCC, NMC are abbreviations of number of fine crackles, number of coarse crackles and number of medium crackles detected in every set of pulmonary disease respectively.

Table .4

	S_G1	S_G2
NFC	34	4
NCC	4	27
NMC	4	9

5. Conclusion

In this research, a classification scheme has been proposed to classify crackles extracted automatically from two groups of pathologic lung signals. In this scheme we have involved two methodologies; the first one is a statistics based methodology and the second is a fuzzy non linear classifiers methodology. It was demonstrated the existence of three categories of crackles in both signals, but with different proportions and spread. In this study we have pointed out the importance of using a large number of features extracted from crackles to ensure the real category of crackles. Thus, we have selected 9 features to enhance the factual behaviour of crackles and to distinguish between their morphology; features related to the waveform characteristics, such as amplitude, and time duration of the crackle waveform are the most relevant features in crackles characteristics distinction. The 1/4 cycle duration T1, and 9/4 cycle duration T2 defer information that is qualitatively a confirmation of information capitulated from IDW, LDW2 and LDW3. Maximum amplitude and total time deflection of a crackle may represent information coming from pressure inside the pulmonary arteries. So that in the case of patients with chronic bronchitis, obstruction in pulmonary airways is mostly affected by a high amplitude cracking. Conversely, in the case of fine crackling, amplitude of the waveform is much lower than it is in coarse crackles case. The nine characteristics features developed in this study are carrying complementary information about the morphology of crackles detected in 80 respiratory cycles from patients associated to two groups of pulmonary diseases; (pulmonary fibrosis and pulmonary bronchitis). Thus, statistics from features analysis have shown that individual crackles can be separated into three groups (fine crackles, coarse crackles and medium crackles). The number and the spread of crackles that are extracted from the two groups of pulmonary diseases may be significant in the pathology interpretation attendance. These results suggest that spectral and waveforms characteristics of crackles may help to improve the accuracy of pulmonary auscultation and to spread out our knowledge about how crackles are generated.

Acknowledgments

This work was supported by a scientific research grant from the Ministry of Higher Education and Scientific Research, Tunisia attributed to the laboratory of analysis Design and Control Systems - ENIT Tunisia, (LR-ACS-ENIT)/ LR-11-ES20) pport acknowledgments are also placed here.

References

[1] Gavriely N, Palti Y, Alroy G, "Spectral characteristics of normal breath sounds" J Appl Physiol, Vol.50, 1981, pp. 307–314.
[2] Murphy R, Gaensler E, Holford S, Del BE, Epler G, "Crackles in the early detection of asbestosis", Am Rev Respir Dis, Vol. 129, 1984, pp. 375–379.
[3] Holford, S. K, "Discontinuous adventitious lung sounds: measurement, classification and modelling" (Sc.D.thesis), Massachusetts Institute of Technology, Cambridge, MA. 1982.
[4] P. Piiril, A.R.A. Sovij, "Crackles: recording, analysis and clinical significance" Eur. Respir. J, Vol. 8, 1995, pp. 2139–2148.
[5] Mete Yeginer, Yasemin P. Kahya, " Elimination of vesicular sounds from pulmonary crackle waveforms" Computer Methods and Programs in Biomedicine, Vol. 89, 2008, pp. 1-13
[6] Xiaoguang Lu, Mohammed Bahoura, "An integrated automated system for crackles extraction and classification" Biomedical Signal Processing and Control, Vol. 3, 2008, pp. 244-254.
[7] Ayari Fatma, Ali T Alouani, Mekki Ksouri. Wavelets: An Efficient Tool for Lung Sound Analysis. AICCSA08 IEEE., (2008) 875-878.

[8] Murphy RL, Jr. Holford SK, Knowler WC, "Visual Lung-sound Characterization by Time-expanded Wave-form Analysis" N Engl J Med, Vol. 296, 1977, pp. 968–971.

[9] J. Hoevers, R.G. Loudon, "Measuring crackles" Chest, Vol. 98, 1990, pp. 1240-1243.

[10] Matsuzaki M, Homma Y. A, "New method for classifying discontinuous adventitious lung sounds", Jpn J Thorac Dis, Vol.21, 1983, pp. 134-140.

[11] Munakata M, Tanimura K, Ogasawara H, Matsuzaki M, Homma Y, Kawakami Y, "Expiratory crackles in diffuse interstitial lung diseases" Ther Res, Vol.3, 1985, pp. 660-666.

[12] American Thoracic Society Updated Nomenclature for Membership Relation, ATS News, Vol. 3, 1977, pp. 5–6

[13] N. AL Jarad, S. W. Davies, R. Logan-Sinclair and R. M. Rudd, "Lung crackle characteristics in patients with asbestosis, asbestos-related pleural disease and left ventricular failure using a time-expanded waveform analysis a comparative study" Respiratory Medicine, 1994, Vol., pp. 37-46.

[14] Robert L. Wilkins, John E. Hodgkin, Brad Lopez, "Fundamentals of Lung and Heart Sounds", (Book and CD-ROM), C.V. Mosby, 2004.

[15] Steven Lehrer, M.D. Understanding Lung Sounds, (Booklet with Audio CD), W.B. Saunders Company, 2002.

[16] The R.A.L.E, "lung sound", Vol. 3.0, 2003, http://www.rale.ca.

[17]http://solutions.3m.com/wps/portal/3M/en_US/3M-Littmann/stethoscope/

[18] Xiaoguang Lu, Mohammed Bahoura, "An integrated automated system for crackles extraction and classification" Biomedical Signal Processing and Control, Vol. 3, 2008, pp. 244-254.

[19] Mete Yeginer, Yasemin P. Kahya, "Probing the existence of medium pulmonary crackles via model-based clustering", Computers in Biology and Medicine, Vol. 40, No. 9, 2010, pp.765-774.

[20] L.J. Hadjileontiadis, S.M. Panas, "Separation of discontinuous adventitious sounds from vesicular sounds using a wavelet-based filter", IEEE Trans. Biomed. Eng., Vol. 44, No.12, 1997, pp. 1269-1281.

[21] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms" Plenum Press, 1981.

[22] J.C. Bezdek and J.C. Dunn, "Optimal fuzzy partitions: A heuristic for estimating the parameters in a mixture of normal distributions" IEEE Transactions on Computers, pages 835-838, 1975.

[23] D.E. Gustafson and W.C. Kessel. "Fuzzy clustering with fuzzy covariance matrix" Proceedings of the IEEE CDC, San Diego, 1979.pp. 761 -766.

[24] A.M. Bensaid, L.O. Hall, J.C. Bezdek, L.P. Clarke, M.L. Silbiger, J.A. Arrington, and R.F. Murtagh, "Validity-guided (Re) Clustering with applications to image segmentation" IEEE Transactions on Fuzzy Systems, 1996, Vol. 4, pp. 112-123.

Fatma Ayari: is Graduate Research Assistant of Electrical engineering. She is a PhD student in Electrical Engineering at the National School of Engineering, Tunisia. She teaches undergraduate courses in Programming Mathematics and Informatics. She's research areas of interest include experimental and computational signal processing, neural networks and automatics. She has published in journals and proceeding conferences as IEEE proceedings.

Mekki Ksouri: is Professor in the of the Department of Electrical and Computer Engineering automatics and Electrical engineering at the National School of Engineers of Tunis Tunisia. His research interests include distributed computing, neural networks, fuzzy systems, genetic and ellipsoid algorithms bioinformatics and biomimetic and bio-Inspired Computing Machines, author of scientific books. He has published a great number of scientific papers published in various IEEE and other international journals. His recent research has been supported by many international organizations, INRIA, AUF. ..

Ali Alouani: Professor in the Department of Electrical and Computer Engineering Tennessee Technological University, he earned his PhD in Electrical Engineering, University of Tennessee Knoxville, 1986. Dr. Alouani developed and taught many undergraduate and graduate courses in the Systems and Signals areas. Till date he has published in more than 120 technical journal and conference papers. He holds 4 patents. He has been active in many areas of Electrical & Computer Engineering. His theoretical research includes, Stochastic Systems, Sensor Data Fusion, Artificial Neural Networks and Fuzzy Systems, Robust Control, Power Systems Stability & Control, Independent Component Analysis, His applied research work includes, Automatic Prescription Verification System for Pharmaceutical Mail Order (APVS), Tracking and Data Fusion, Tube Leak Detection, Pulmonary Tumor Detection, Autonomous Systems Design (ASRFDE), Various Signal Processing Applications.