Lung Cancer Diagnosis based on Ultrasound image processing

Achim Cristian, Rusu-Both Roxana, Dulf Eva-Henrietta
Automation Department
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
achimcristian@yahoo.com, roxana.both@aut.utcluj.ro

Chira Romeo Ioan
I[st] Medical Clinic, "Iuliu Haţieganu" University of Medicine and Pharmacy Cluj-Napoca
Cluj-Napoca, Romania
romeochira@yahoo.com

***Abstract*— Cancer represents one of the leading causes of mortality in this century. Patients suffering from lung cancer (LC) have an average of 5 years life expectancy after diagnosis. This is due usually to late detection. The most common methods to detect lung cancer are computed tomography (CT) and magnetic resonance imaging (MRI). These investigations are invasive to the human body and are performed only based on a doctor's recommendation, which comes usually after patients develop symptoms – hence the late detection of cancer. Nowadays transthoracic ultrasonography (TUS) and US-guided biopsy have gained a larger field in the management of patients with peripheral pulmonary nodules or masses. From the multiple ultrasonography (US) advantages which made it a well-established solution in the management of tumoral and non-tumoral abdominal pathology, few stand out: it's non-invasive and has reduced costs. However, until now transthoracic US is underused for lung cancer diagnosis because it can be hard to interpret to determine an accurate diagnosis. In this paper we aim to develop a software application for lung mass classification based on ultrasound image processing. This could be an important step in towards early detection of lung cancer, by introducing the transthoracic ultrasonography in the regular annual check-up, improving the life expectancy of patients or even complete recuperation.**

Keywords—lung cancer, ultrasonography, tumor detection, image processing

I. INTRODUCTION

Lung cancer (LC) represents a leading cause of cancer-related mortality worldwide. Recently published data estimates an increase of LC deaths worldwide from 1.6 million in 2012 to 3 million in 2035, [1]. It is the second most common cancer in both men and women due to the tobacco consumption. This is the main cause, but not the only one, because lung cancer is a disease that can affect anyone. In some cases, previous lung diseases (chronic bronchitis, emphysema, pneumonia, and tuberculosis) could have a role in the development of lung cancer [2, 3].The assessment of patients with pulmonary focal abnormalities continues to represent a major problem in LC diagnosis.

Unfortunately, the early stages of lung cancer may not show symptoms. As tumors grow in size, they can produce a variety of symptoms. There are patients who show certain symptoms, even if some are nonspecific, which causes them to contact the doctor. Symptoms may be the consequence of the primary tumor, which can cause local compression or can spread through metastasis to the rest of the body. In some cases, the primary tumor does not have a pulmonary location, but it appears secondary, which is why the clinical picture can vary greatly from patient to patient. Cough is the most common symptom of lung cancer. However, many smokers (those who smoke for a long time) have chronic cough, so it is very important for a person with chronic cough to go to the doctor's control if they notice that the cough is changing or getting worse.

There are two main types of lung cancer: small cell lung cancer and non-small cell lung cancer. The latter is the most common type and it can spread anywhere in the body through a process called metastasis. There are various methods for diagnosing lung cancer and the most common used are: computerized tomography (CT) scan, positron emission tomography (PET) scan, magnetic resonance imaging (MRI) scan. In the case of PET scan [3, 4], it involves radioactive tracers, but the exposure to radiation is minimal.

Ultrasonography (US) [3, 5] has been developed in order to offer a lot of useful information about the tumors such as the tumor structure, vascularization, the stage of parietal invasion and sometimes lymph node invasion. The US is less expensive and comparatively safer technique than CT scans. More than that, it is not recommended for the human body to perform a CT, rather than by the doctor's recommendation, because for a person it would take 3.3 years to get the same amount of background radiation that a CT delivers in less than a minute [3, 6]. Even considering the advantages of the US, it is not used for lung cancer detection. The main reason is that is very difficult to interpret by lung specialists that are not used to perform transthoracic ultrasonography on patients. Plus, the lung is not visible on US if it is healthy and full of air.

Taking into consideration the multitude of advantages the US has to offer, we imagined an application which helps doctors diagnose lung cancer. Hence, the main contribution of this work consists in the development of a lung cancer (LC) diagnosis software application able to accurately and non-invasively detect non-small cell lung cancers as part of a research project. Until now, we developed a preliminary

978-1-5386-4444-7/18/$31.00 ©2018 IEEE

software application for intrathoracic tumor recognition of ultrasonography imaging, addressed to peripheral LC diagnosis, based on US principles of definition of LC concerning echogenicity, contours, echo structure, interfaces with adjacent structures, and also on diagnostic criteria for differential diagnosis (lung condensations, cysts, benign tumors and pleural pathology). This would help physicians to identify patients with LC and choose the next appropriate method of investigation – lung percutaneous biopsy, bronchoscopy, echo-endoscopy, surgical treatment or surveillance.

The applied method uses the edge-based active contour algorithm used for segmentation, similar to Geodesic Active Contour [7] and it is a very flexible and powerful method. It can segment many types of images, including some that are very difficult to segment with other methods based on thresholding or gradient. We will outline the steps in which the application runs, the algorithm for segmenting images, and examine performance on some US images.

II. PROPOSED METHODOLOGY

Driven by the fact that there are reports in which the greyscale histogram was used for breast tumors, we performed a pilot study as part of a national research project, in order to evaluate the possibility to build a computerized model to assist young doctors or ultrasound technicians in recognizing peripheral lung tumor and differentiate them from benign condensations.

The first stage of the pilot study consisted in the identification of potential cases, their partial enrollment in the study, the collection of blood samples as well as the acquisition of ultrasound images. As result, we analyzed two populations of patients – with peripheral lung cancers and with lung condensations. We performed transthoracic US, recording different parameters – diameters, contour, echogenicity, distribution of the vascularization, invasion of the adjacent structures, etc. Images were recorded and transferred for computerized analysis. The development of lung cancer diagnostic software could improve the quality and accuracy of the diagnosis, enhance the success of therapy by early detection of cancer, avoid unnecessary biopsies and reduce radiologist's interpretation time.

Hence, the second stage of the pilot study consisted in the elaboration of the diagnostic software specifications that accurately detects cancer cells without any confusion, in order to achieve the previously presented goals. The main features that allow differentiation between benign tumors and malignant tumors are related to the shape of the nodules, the echogenicity of the different regions and the regularity of their contours. An ellipsoid form or the presence of two or three well-defined lobs is considered benign finding. On the other hand, the presence of ramifications or a form that is higher than wider are considered to be malignant findings. Moreover, when examining the smoothness of the contour, if it has micro-lobes, angular or spicular edges, they are interpreted as malignant findings as opposed to a rounded and well-defined outline. Finally, hyperechogenicity and the presence of an echogenic thin capsule increase the likelihood of benignity, while calcifications and hypoechogenicity are associated with malignant nodules.

Based on the clinical experience of the physicians we were able to identify the main features that we need to extract from the ultrasonography image in order to be able to help the diagnostic process or even give a probabilistic diagnostic. Based on their importance for the diagnostic process the main features of the determined area of interest are: A) the irregularities (spiculation); B) the shape; C) the transition rate of the contour pixel intensities; D) the number/area of the whiter spots inside the contour and E) the aspect ratio of the contour shape. In this step, in the development of the diagnostic software application was implemented a feature for global analysis and local contour characteristics, which can act as support for the radiologist and highlight the locations of suspect areas so that the overall sensitivity is high. The proposed method for implementation consists of 2 processing steps, Fig.1. The first step is to preprocess the ultrasound image and extract the region of interest (ROI) from the whole image. The second step involves detecting the contour of the tumor in the region of interest in order to classify it. The last step is to extract the previously presented key features.

Fig. 1. Lung cancer tumor detection algorithm

At this stage, the edge-based active contour algorithm was chosen for implementation and testing for separating the suspicious region from its background to be used further for tumor classification stage. Tumor contouring is an important criterion in classifying them into cancers or non-cancers.

III. LUNG CANCER SOFTWARE APPLICATION

The acquisition stage of the ultrasound images (the first stage in the lung cancer detection algorithm, Fig.1) was performed using a General Electric LOGIQ S6 in .bmp or .jpg format.

The ultrasound image preprocessing step consists in detecting the region of interest. This is done through the user application. The following processing steps will apply only to the region of interest determined by the user (physician).

The entire software application was created in Matlab, using its native image processing capabilities and complex functions on top of which we implemented our own algorithms for feature. The algorithms were conjoined through the visual composition and temporal behavior of a graphical user interface, in order to enhance the efficiency and ease of use for the underlying logical design of the program.

After the application started, the user has to load an ultrasound image and define a relatively accurate mask by selecting the region of interest, from which the contour algorithm will find its exact real margins, Fig.2. When acoustic shadow is present on the image, a signal void behind structures like bones that strongly absorb or reflect ultrasonic waves, the user also has the possibility to add one or more white lines to the image in order to separate those irrelevant areas. Having a high intensity, the line stands in contrast with its surroundings, acting like a barrier. However, they only appear for this sole purpose, and not for the followed algorithms.

The similar to Geodesic edge-based active contour algorithm evolves the segmentation using an iterative process, performing the number of iterations given by the user, 200 being the default. The algorithm starts from the given mask, a binary image that specifies the initial state of the active contour, and then grows outwards, expanding until it finds object boundaries. In order to obtain faster and more accurate segmentation results, an initial contour position that is close to the desired object boundaries should be specified.

Fig. 2. Diagnostic application users interface (In red is the defined starting mask and in green is the obtained contour after the segmentation)

The results obtained using the implemented key features extraction algorithms are displayed in the left lower part of the detection system user interface. Having the exact uneven outline of the real region of interest, the centroid is easily computed as well as the distances (D) between it and all the contour pixels. For the five key factors that had to be analyzed in order to determine if the region of interest has stronger characteristics of a cancer or a non-cancer, the main algorithmic program was structured accordingly.

$$D(i) = Distance(ContourX(i),ContourY(i),CentreX,CentreY)$$

A. Contour Algorithm

The more non-superficial irregularities the contour has, the more it indicates towards a tumorous nature. Therefore, once the contour was determined, we developed an algorithm to determine the smoothness of the lesion's contour by computing the differences between the peaks and valleys of the contour in respect of the centroid. The output of the algorithm is a coefficient obtained from the mean of all these differences. Perfect smoothness, meaning no differences between all the supposed D fluctuations, would indicate not only the absence of peaks and valleys, but also the shape of the contour being a circle of radius D. Consequently, the Contour Algorithm also plays a part in determining the shape of the studied area.

$$ContourCoeff = mean(abs(difference([peaks; valleys])))$$

B. Shape Algorithm

Because the shape and the smoothness of the lesion contour are the most important features in distinguishing its nature, another approach to the deduction of its shape is the comparison of its surface to that of the circumscribed circle and the inscribed triangle (which can indicate pneumonia). To determine the inscribed triangle, three points and their coordinates are needed in such a way to obtain the largest possible area for the triangle. The first point, specifically the farthest away from the centroid point, is plainly the largest of all D distances. The second point, as well as the third, must be situated as relatively opposite to this first one and to each other as possible. Therefore, the second point has to be located on the contour not closer than the sixth of the entire contour length, by either side of the first point. Likewise, the third one must be at a length/6 distance of each of the other two points. This algorithm assures the largest distance between the three corner points of the inscribed triangle, offering its largest area. The circumscribed circle is then constructed exactly on these three vital points, encapsulating the entire contour. If the difference between this circle and the contour is larger than the triangle-contour difference, the shape tends towards a more circular one, or triangular otherwise.

C. Fade Algorithm

Another one of the five key features that has to be studied is the differences of pixel intensities at the lower part of the contour. This feature is also very important because a hard in depth transition of the pixel intensities indicates the cancerous nature of the lung lesion in opposition to a smooth in depth transition. To obtain a coefficient in this sense, two different brightness averages were computed from the pixels around the outline of the shape. One mean from the points found inside the contour and the other for the ones outside, but only around its lower outline, the top part not being of interest in this fade determination algorithm, Fig.3. The final result is obtained from the difference between these two means. If this coefficient is relatively small, meaning close to zero, the fade is a slow one, barely visible, compared to the sudden intensity change when the large result is obtained, the contour representing the border of a high contrast in brightness.

$$FadeCoeff = (FadeCoeffExt/FadeCoeffInt-1)$$

Fig. 3. Fade algorithm example

D. White Spots Algorithm

The conglomerates of higher intensity pixels inside the contour represent the white spots. From a diagnostic point of view the presence of whiter spots indicates the presence of healthier areas inside the tumor's contour. The desired result is one regarding their computed area compared to the whole lesion contour surface. If the computed area of the brighter pixels is high compared to the formation shape surface, it indicates the tumorous nature of the lesion. To find and obtain the number of the brighter pixels, all points inside the contour are compared to the average intensity of that surface. The ones brighter than that threshold are added up and compared to the others through a ratio. In order to not count in also the brighter pixels which are on the contour line, as part of the sudden fade,

the area studied is taken slightly smaller than that of the shape's outline.

BrightPix = sum(Img(ROImg >= mean2(ROImg(shape==0)=0 > 0)))

E. Aspect Ratio Algorithm

The aspect ratio of the lesion contour is defined as the ratio between the largest horizontal length and the largest vertical length, specifically the anteroposterior and latero-lateral axes. This feature was implemented to identify if the shape is a triangular one with the farthest point away from the surface, like in the case of pulmonary condensation. To obtain this ratio, each pair of opposite pixels must be found and the distance between them computed. For the horizontal line, the y coordinates range only by incrementing and the x coordinates are the ones of the contour points at that precise index, Fig.4. From all the distances between these coordinates, the largest one is divided by the vertical line, obtained analogous.

AspectRatio = abs(max(verticals(ii))/max(horizontals(ii)))

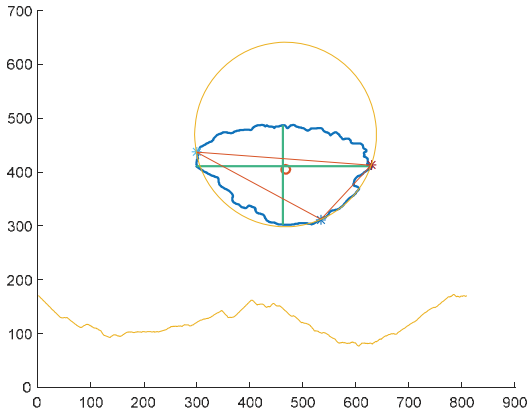


Fig. 4. Shape and aspect ratio algorithm example

IV. Results and Discussion

Using the developed application, a series of tests were performed to determine the accuracy and degree of detection of the contour of the lung lesions using ultrasound images of confirmed malignant tumors (cancers) and other non-cancers.

To this end, the first test scenario was to use over 30 ultrasound images of both confirmed malignant tumors and other non-cancers to determine each individual feature coefficients using the developed diagnostic system. Based on the performed tests we determined the nominal range value for the feature coefficients in both cases, table I.

TABLE I. Feature Coeff. Nominal Range Values

Confirmed diagnostic / Feature	Cancers	Non-cancer
*A. **Contour***	< 14	> 14
*B. **Shape***	circular	triangular
*C. **Fade***	> 5	< 5
*D. **White spots***	> 600	< 600
*E. **Aspect ratio***	0.5-1.5	∉ 0.5-1.5

The second test scenario was to use the developed diagnostic system on new ultrasound images of both confirmed cancers and other non-cancers to test if based on the determined nominal features coefficients range value we will obtain the confirmed diagnostic. The second test scenario was performed on multiple ultrasound images, some results examples being presented in Fig.5 and Fig.6.

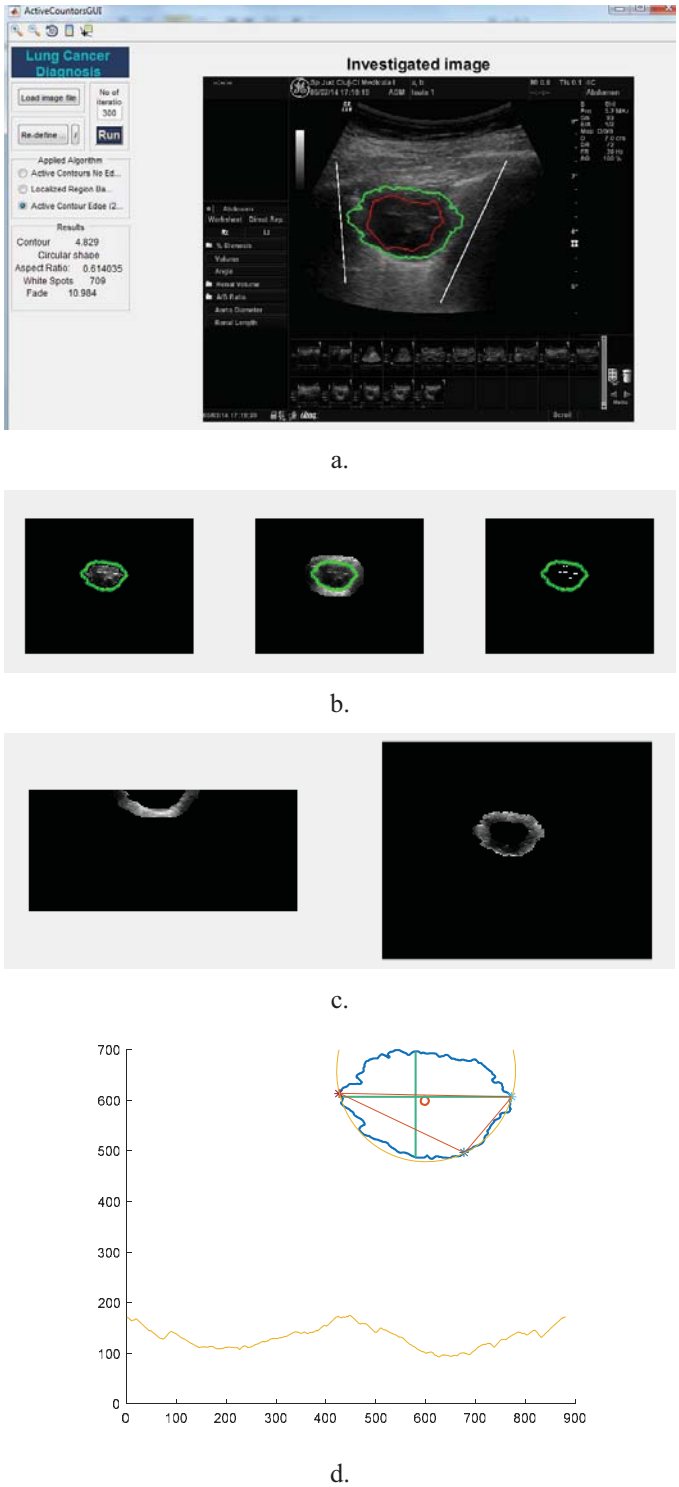
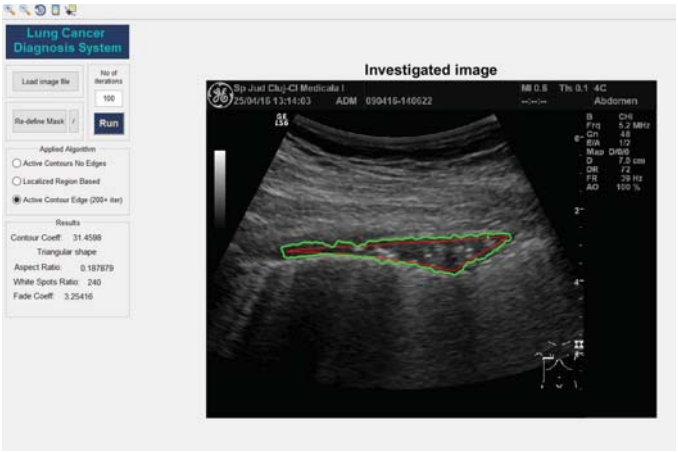


Fig. 5. Key feature coefficients for necrotic tumor: a) contour coeff; b) white spot coeff; c) fade coeff; d) shape and aspect ratio coeff.

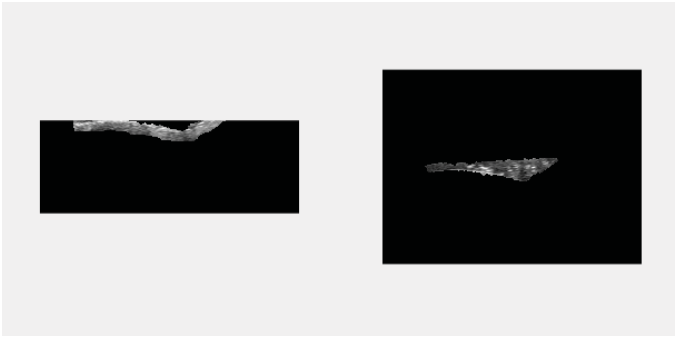
As it can be observed in Fig.5 the developed diagnostic system is tested on a confirmed malignant tumor ultrasonography. In Fig.5.a can be observed the determined contour of the tumor is represented with green color. The contour represented with red color line is the contour of the region of interest defined by the user (physician). The straight white line that can also be observed are also defined by the user just to eliminate the non-important areas for the contour detection algorithm. Fig.5.b shows the sequential results based on which the white spots coefficient is computed. Fig.5.c. actually presents the dimension of the area outside the tumor contour which is taken into consideration for fade coefficient computation. Lastly in Fig.5.d is presented the circumscribed circle, the inscribed triangle for the determined tumor contour and also at the bottom side is presented the alignment of D, specifically all the distances from the centroid to each pixel on the contour, used for finding the peaks and valleys of the outline. Based on these results, on the bottom left side of the diagnostic system main window, Fig.5.a, are displayed the determined key feature coefficients. As it can be observed based on the obtained results and the key feature nominal coefficient range value, Table I, the tumor can be classified as malignant as the confirmed diagnostic.

In Fig.6, the obtained results after testing the developed diagnostic system on a confirmed pulmonary condensation are presented. In Fig.6.a can be observed the determined contour of the suspicious region of interest. As in the previous case the contour of the region of interest defined by the user (physician) is represented with the red contour. In a similar manner Fig.6.b shows the sequential results based on which the white spots coefficient is computed. As it can be observed in this case the number of the brighter pixels is greater. In Fig.6.c. are presented the obtained results by applying the fade algorithm and in Fig.6.d are presented the obtained results by applying the shape and aspect ratio algorithms. As it can be observed in the diagnostic system main window, Fig.6.a, in this case the diagnostic system indicated shape is a triangular one. Considering this results, together with the other key features coefficients the suspicious region of interest can be classified as benign which corresponds with the confirmed diagnostic.

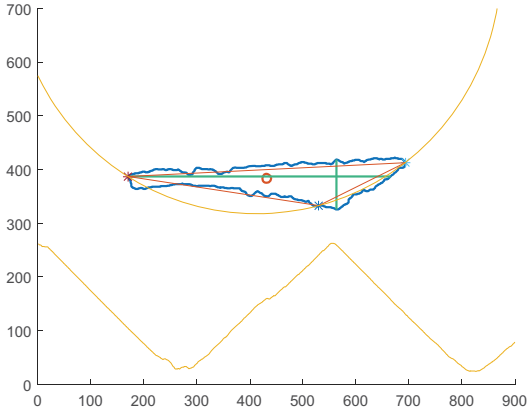


a.

b.



c.



d.

Fig. 6. Key feature coefficients for pulmonary condensation: a) contour coeff; b) white spot coeff; c) fade coeff; d) shape and aspect ratio coeff.

Based on the results presented so far we can state that the developed lung cancer diagnostic system proved to be effective by showing promising results in the test scenarios. The next step in its development is to determine a mathematical model based on the key feature coefficients in order to be able to provide a possible diagnostic. A possible approach is to use machine learning, neural networks in order to provide a possible diagnostic. A minus of this approach is the need of a large database with lung ultrasonographies with confirmed diagnostic.

For now, there is a lot of room for improvements and tests. The accuracy of the system must also be tested on the special case of lung lesions, but for now we do not have such cases in our database. From a theoretical point of view, this may result in another set of features that will need to be extracted from the lesion contour which we consider for further developments.

V. Conclusions

The work presented in this paper is the startup of a national research project which promises to improve the life expectancy of lung cancer patients by early diagnosis. The main cause that prevents early diagnosis of lung cancer is that it can be diagnosed only by CT of MRI which cannot be performed every year (as a check-up) without a doctor's recommendation, being invasive tests and also expensive. An alternative is represented by transthoracic ultrasounds which is a non-invasive test and reduced costs. The only drawback is that the results depend on the experience of the physician. In this paper, we propose a novel approach for automatic detection lung tumors based on ultrasound images, to assist young doctors or ultrasound technicians in recognizing lung tumor and differentiate them from condensations. Experimental results prove that the effectiveness of the developed system.

Acknowledgment

This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS – UEFISCDI, project number PN-III-P2-2.1-PED-2016-0425, contract number 178PED/2017.

References

[1] J. Didkowska, U. Wojciechowska, M. Manczuk, J. Lobaszewski, "Lung cancer epidemiology: contemporary and future challenges worldwide", in Ann Transl Med. 2016;4(8):150. DOI: 10.21037/atm.2016.03.11

[2] DR. Brenner et all., "Previous lung diseases and lung cancer risk: a pooled analysis from the International Lung Cancer Consortium.", in Am J Epidemiol. 2012 Oct 1;176(7):573-85. Epub 2012 Sep 17.

[3] R.I. Chira et. *all*, „Transthoracic Ultrasonography: Advantages and Limitations in the Assessment of Lung Cancer", InTech, 2017

[4] A. Sebei et. all, "Hybrid PET/MRI co-segmentation based on joint fuzzy connectedness and graph cut", in Comput Methods Programs Biomed. 2017 Oct;149:29-41. doi: 10.1016/j.cmpb.2017.07.006.

[5] T. Nelson et. *all*, Three-dimensional ultrasound, Lippincott Williams & Wilkins, Philadelphia, 1999

[6] LF. Herbert, "Drawbacks and Limitations of Computed Tomography", in Tex Heart Inst J. 2004; 31(4): 345–348.

[7] V. Caselles, R. Kimmel, G. Sapiro, Geodesic active contours. International Journal of Computer Vision, Volume 22, Issue 1, pp. 61-79, 1997.

[8] L.K.Wee, E. Supriyanto, W. M. H. B. W. Mahmud, "Ultrasound Image Processing: And Its Application Using Matlab", Lap Lambert Academic Publishing GmbH KG, 2011, ISBN 3845417838, 112 pp

[9] F. Boutaouche and N. Benamrane, "Diagnosis of breast leasions using the local Chan-Vese model, hierachical fuzzy partitioning and fuzzy decision tree induction", in Iranian Journal of Fuzzy Systems 2017

[10] S. N. Acho and W. I. D. Rae, „ Dependence of shape-based descriptors and mass segmentation areas on initial contour placement using the chan-vese method on digital mammograms", in Computational and Mathematical Methods in Medicine, 2015, 1-16.

[11] P. Rahmati et. *all*, "Mammography segmentation with maximum likelihood active contours", in Med. Img. Analysis, 2012, 1167–1186.