Classification of Lung Nodules into Cancerous and Non-Cancerous in Computed Tomography (CT) Images

[1]Madhu Choukimath,[2]Prof.Savitha S K,

[1]Student, Department of Computer Science and Engineering, [2]Assistant Professor, Department of Computer Science and Engineering, Banglore Institute Of Technology,, Banglore Institute of Technology,

V V Puram, Banglore-560004

Abstract:— The Biomedical Image Processing is a growing and more demanding field which contains different types of imaging methods such as CT scans, X-Ray and MRI etc. Computer Aided Diagnosis (CAD) is becoming one of the most popular and effective method for diagnosing many diseases including cancer. Lung cancers are the most common diseases which cause mortality worldwide. Developing a most effective computer-aided diagnosis (CAD) system for detecting lung diseases is of great clinical importance and can increase the patient's chance of survival. In lung cancer, detection of a nodule is a fundamental problem.However, detection of early stage lung cancer in computed tomography (CT) scans is challenging and time-consuming. Radiologists will experience heavy pressure and workload considering the large number of scans they have to analyse on a daily basis. In the present paper, various steps namely pre-processing, segmentation, feature extraction followed by neural networks classification that aims to increase the speed and accuracy that decrease the time involved in cancer detection are given, Here various pre-processing methods such as contrast enhancement method such as histogram equalization, gabor filter method, dilated gradient mask are used. In segmentation, image is partitioned from other anatomic structures by binary thresholding After thresholding, the background (the outside of the body) is eliminated by suppressing all components adjacent to image boundary by flood-filling. This gives lung mask a meaningful region and the result of image segmentation is a set of segments that collectively cover the entire image and all pixels in the segmented region which are similar with respect to some characteristic such as colour, intensity, texture etc. Different features of the nodule such as area, eccentricity, equivDiameter, Euler number, perimeter, solidity, orientation are considered, A comparative study of Support Vector Machine and Neural Networks i.e back propagation algorithm for classification of lung nodules is applied.

IndexTerms–Computer Aided Diagnosis (CAD), Lung nodules, Support Vector Machine (SVM)[8], Multi-Layer Perception (MLP)[9].

I. INTRODUCTION

Image processing in the medical field concentrates on the capture of images for both diagnostic and therapeutic purposes. Snapshots of in vivo physiology and physiological processes can be garnered through advanced sensors and computer technology. Biomedical imaging technologies utilize either x-rays, Computed Tomography (CT), ultrasound, Magnetic Resonance Imaging (MRI), radioactive pharmaceuticals (nuclear medicine: SPECT, PET) or light (endoscopy, OCT) to assess the current condition of an organ or tissue and can monitor a patient over time over time for diagnostic and treatment evaluation.

With the advancement of medical image processing, the modern healthcare system has started adopting upcoming technologies in order to give better quality of diagnosis of disease. Proposed study is more particular about the lung nodule detection from various other types of radiological study available in image processing [1].

Various diseases that are evaluated using lung CT images are pneumonia, nodule developments, tuberculosis, bronchiectasis, cystic fibrosis, inflammation or the other diseases of the pleura, interstitial and chronic lung diseases, congenital abnormalities, internal haemorrhage within the chest wall etc [3]. Lung nodules are automatically detected and localized in CT images using Computer Aided Detection (CAD) tool. These are helpful to assist the radiologists in the process of lung nodule detection. A major problem in these CAD tool is the large number of false positives, so to achieve sensitivity and low number of false positives the machine learning models are applied in the developed CAD systems.

II. REVIEW OF EXISTING WORK

Literature review in a report is the section which shows the various analyses and research made in the field of our interest and the results are already published taking into account the various parameters of the project and the extended of the project. It is the most important part of report as it gives a direction in the area of the research related to Lung nodule detection in CT images based on the important data from the current or the previously collected data. It helps us to set a goal for analysis. The reviews of the literatures are carried out with respect to disease detection, segmentation, classification and preprocessing.[3].

Computer Aided Detection (CAD) systems used in lung nodules detection generally consists of mainly four main stages: Pre-processing, Segmentation, Feature extraction and Classification.

The architecture of the proposed system is shown in Fig.1. The phase-1 pertains to review of literatures where the existing survey papers as well as techniques pertaining to Lung-related disease detection, segmentation, preprocessing and classification are critically investigated. Research gap is explored along with problem identification. The phase-2 focuses on developing a multi-level image enhancement scheme for Lung CT images. The next phase-3 pertains to the novel segmentation process where indicative Content-based image retrieval techniques will be used for ensuring the accuracy in the segmentation process of the Lung CT images. The last phase-4 will be based on performing classification of the abnormalities / diseases pertaining to Lung nodules.

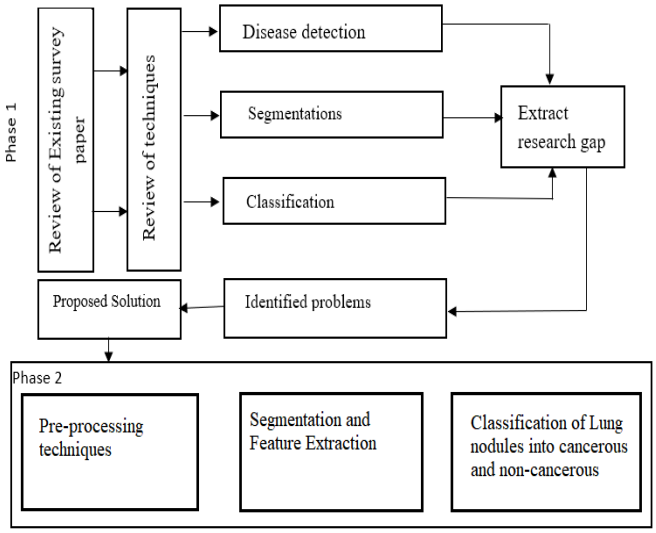

Fig 1. Proposed Architecture.

Phase-1 Of Study: Investigation Of Research Gap

Medical image processing has always posses a greater deal of challenges to find the most satisfactory solutions to the complex problems. The various techniques of diagnosis of the chest radiographs are discussed. The discussion of prior literature is done with respect of problem identification, techniques used, performance observed[1], and limitations accompanied by it. In a nutshell, every research work discussed does have certain potential accomplishment as well as limitations. However, that shouldn't be bigger scale problems as every other research too has certain limitation. However, it is essential to understand some of the problems that do have existed in past and inspite of rigorous attempts, the problem is still unsolved. Therefore, we are more interested to understand such unsolved problems rather than discussion limitations. Hence, the identified research gaps explored in our investigation process are as follows:

- **Only Focused on Similar Clinical Aspect**: As per our clinical review, the meaning of chest radiograph basically pertains to radiological investigation of thoracic cavity, where there are various organs e.g. rib cage, lung, heart, diaphragm, etc.[5] The proportion of the work on tuberculosis is very higher than amount of the research work done on lung cancer or pneumonia or internal hemorrhage etc.
- **Less Focus on Disease Classification**: The existing system of disease detection from the chest radiographs just results in precise segmentation process. In a nutshell, segmentation technique is highly emphasized in existing system. However, the segmentation technique is not found with enhanced classification system. Hence, we have also explored the implication of data mining technique, neural network technique, and support vector technique in order to mitigate the problems of classification. However, there are some of the major contradictions in this that doesn't make the existing system much applicable for classification viz.
- **No Significant Work on Enhancement**: It is highly essential that every radiograph images should be preprocessed in order to retain maximum useful information required in next steps of processing. Chest radiographs are in various formats, types and colors. Hence, it is required to implement a technique to retain maximum resolution of the input image.
- **Expensive Machine Learning Process**: In existing system, there are some work on chest radiograph towards disease classification done using data mining, neural network, and support vector machine. However, there are some issues observed in existing implementation e.g. i) no discussion on number of training required to get the elite outcome, ii) no illustration of validity of the condition set for identification of diseases in terms of preciseness, iii) training and learning process usually takes time of bigger size of medical image in DICOM format and its accuracy of disease detection and classification have higher dependency on size of trained data and hence not reliable, iv) implication of supervised technique e.g. SVM is more concern for enhancing the attributes of optimization to fit into the framework being selected. However, such approach is highly sensitive to over-fitting. v) The feature extraction process is not much emphasized in this process.
- **Less focus on Algorithm Complexity**: Automatic detection of the abnormalities and diseases pertaining to the chest-radiograph is a complex process owing to involvement of various subtask e.g. pre-processing, feature extraction, and classification of the sophisticated formats (e.g. DICOM). Hence, the algorithms to perform such task will also require to be highly cost efficient with respect to time and space complexity. Till date, none of the existing techniques have any evidence of time and space complexity compliance.

Phase-II Of Study: Pre-Processing With Enhancement

The main goal of the proposed work is to collaborate the different pre-processing techniques and segmentation processes and applying the different classificatiobn techniques that gives different sensitivity and accuracy results. Applying the different preprocessing techniques leads to highly enhancement of lung CT image to enhance the detection rate of the lung disease. Our major aim lies in enhancing the signal quality in order to reduce the outliers when the lung CT image is used for analysis by the radiologists. It should be noted that presented work is focused on pre processing the lung CT images using multiple ranges of operations. Although, there are various implementations in past pertaining to image enhancement, but we bring out contribution / novelty by setting the following research objectives that have been addressed in this manuscript:Capability to process lung CT image: Lung CT images are normally associated with complexities (poor illumination, presence of radio-opaque objects) and the proposed system is designed to perform enhancement using 5 different enhancement operations.

III. METHODOLOGY

CT images are collected from The Cancer Imaging Archive (TCIA) database[15]. The images are stored in DICOM format and for the convenience the DICOM image is converted into jpg format. The dataset is available through the National Cancer Institute's (NCI) The Cancer Image Archive (TCIA). This dataset consists of 70 thoracic CT scans (10 for a calibration and 60 for test), including the DICOM format images, spatial coordinates of the nodule locations and the diagnosis for each nodule in the calibration and test datasets. Dimensions of whole CT images dataset are 512 x 512 pixels, with a bit depth of 12 bits.

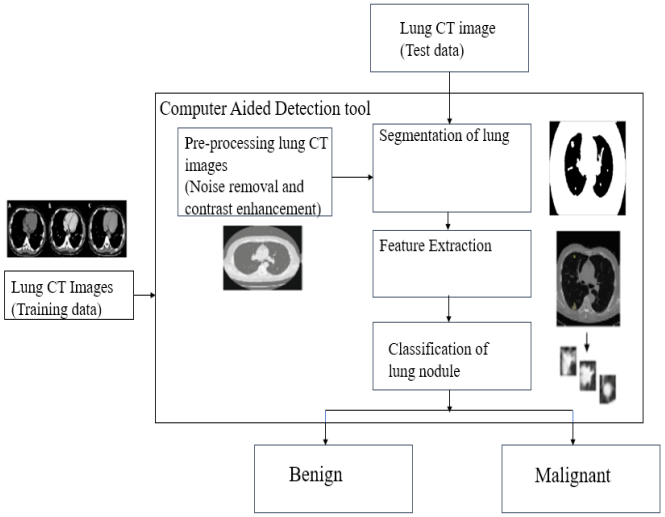

Fig. 2 Lung Nodule Detection tool.

A. Preprocessing

The pre-processing stage is the stage of improving the quality of the image, It reduces the noise and improves the images visibility,[7] CT images mostly prone to salt and pepper noise. The median filter proves to be the best solution in reducing the impulse noise as it also preserves all the edges in the image.Noise removal and the Contrast enhancement are the two techniques used to improve the image quality in this paper. The Contrast enhancement improves the visibility of minute structures by enhancing the brightness difference between objects and background and also Median filter and Gaussian filter method are used to increase the clarity of CT image. Median filter is used to remove the salt and pepper noise, under certain conditions it preserves edges while removing noise.

B. Segmentation

The main goal of segmentation is to isolate the lung tissue from the chest wall and heart segments[7]. The image segmentation method uses thresholding which means it uses a threshold to partition an image into foreground and background pixels. The threshold is computed using the Otsu's method[4]. This method chooses the threshold value that minimizes the intra-class variance.[6].

C. Feature Extraction

After knowing the region of interest we now identify the nodules based on the features such as area, eccentricity, equivDiameter, Euler number, perimeter, solidity, orientation and so on. In medical field texture provides better and more detailed information about the images. Hence the size of the nodule is taken into consideration if the size of the nodule is 3cm or less in diameter is called a pulmonary or benign nodule. These types of nodule are noncancerous and pulmonary nodules are the characterization of early stage of lung cancer. Another type of cancer nodule whose size is larger than 3cm is in diameter is called as a lung mass. This type of nodule is likely to be cancerous and needs to be detected as early as possible.

D. Classification

Here two methods are used for classifying the lung nodules they are Support Vector Machine algorithm and Multi-Layer Perception(MLP). The SVM Algorithm in machine learning, support vector machinesthat analyze data used for classification and regression analysis[2]. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilisticbinarylinear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). MultiLayer Perceptron: AMLP is a class of feedforwardartificial neural network. An MLP consists of, at least, three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called backpropagation for training.[1][2] Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable.

IV. RESULTS AND ANALYSIS

The GUI created for nodule detection is given as

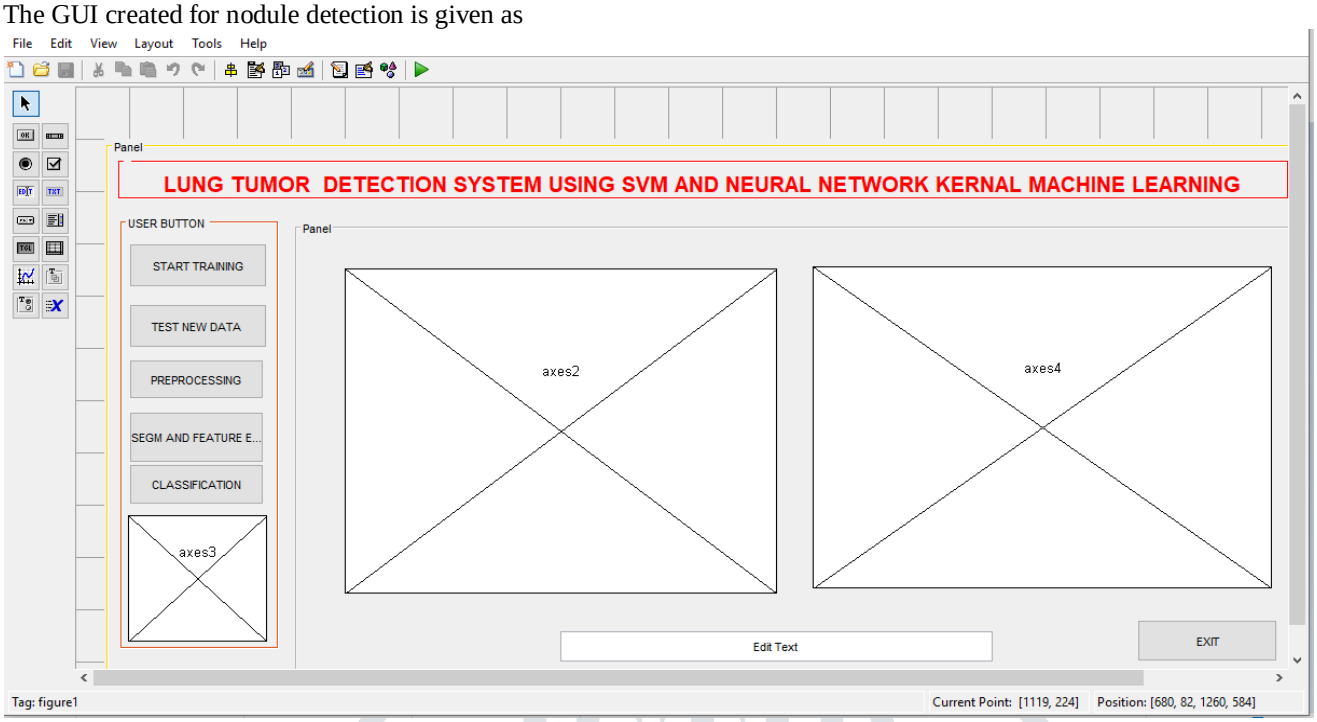

Fig. 3 GUI for nodule detection.

The simple GUI is created with five buttons, one for training the machine with 133 CT images, Test New Data button asks the used to choose among few CT images.

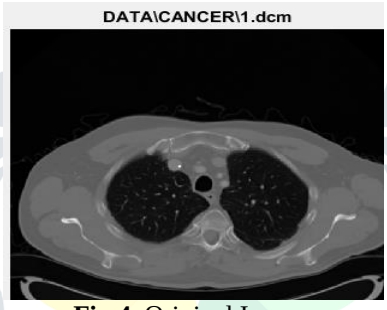

Fig 4. Original Image.

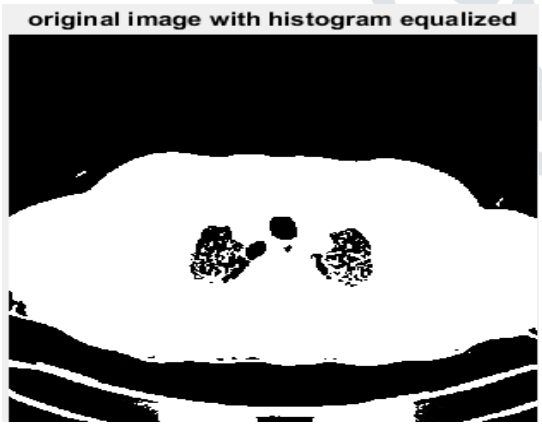

Fig 5. Original image with histogram equalized

Fig 6. Image with dilated gradient mask

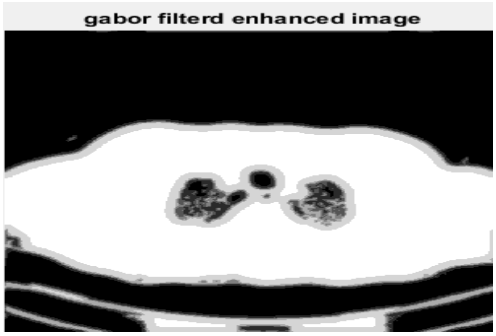

Fig 7. Gabor filtered enhanced image

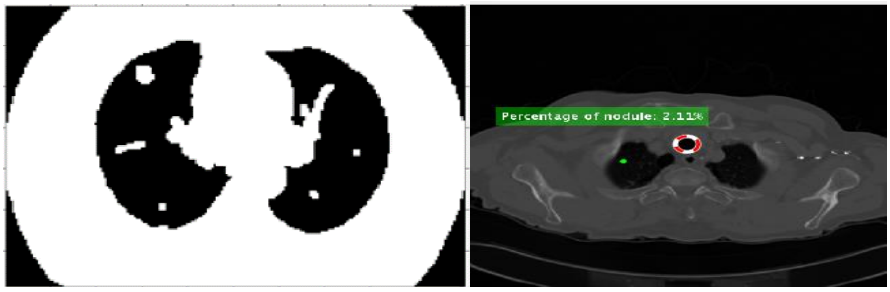

Fig 8. Segmentation and feature extraction.

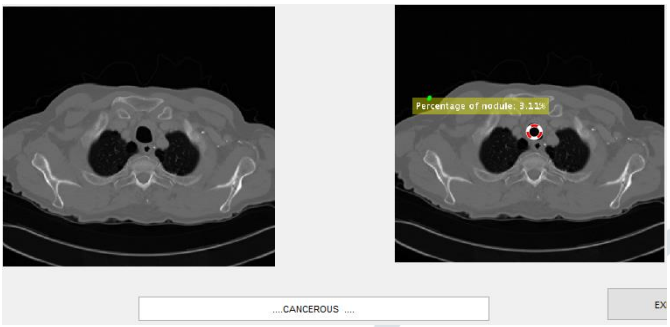

Fig 9. Classification of nodules into cancerous and non cancerous.

Confusion matrix

The confusion matrix is a table in machine learning that depicts the performance of the classification algorithm. The below figures shows the confusion matrix for a lung nodule detection after applying the SVM algorithm,[8]

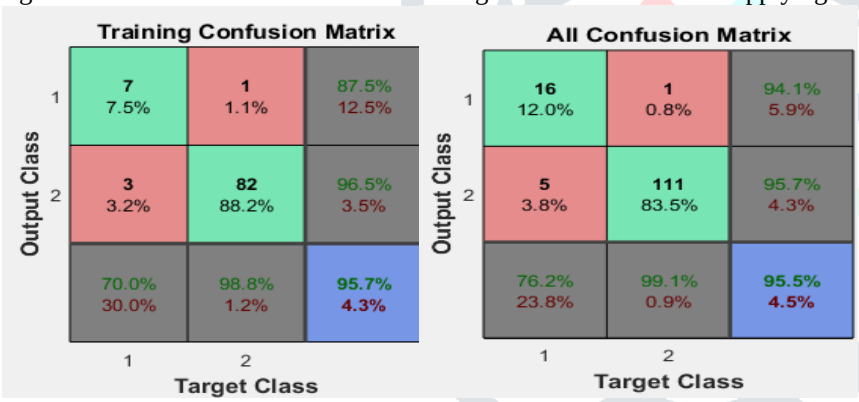

Fig 10 Training confusion matrix **Fig. 11**Test confusin matrix

As seen in fig 7.1 total of 226 images used, out of which 93 images are used for trainingand accuracy of 95.7% is achieved for training. For testing 133 images were used, out of which 6 images were miscalssified yielding an overall accuracy of 95.5%, the sensitivity thus calculated found to be 99.1% and specificity was 79.2%[9].

TABLE I. EVALUATION OF SVM

Samples	True Nodules	Accuracy	Sensitivity	Specificity
112	16	88.2%	76.363%	89.4%
90	20	91.2%	78.264%	97.6%

V. CONCLUSION

Lung cancer is one of the most harmful disease in the world. Proper diagnosis and early detection of lung cancer can increase the survival rate of the patients, Computer Aided Diagnosis (CAD) involving Image processing technique for nodule detection helps in diagnosis of cancer, I this paper, Both computer aided detection (CAD) systems (SVM and MLP) have shown that they were able to detect lung nodules that are not attached to the thoracic wall (non-juxtapleural nodules). These systems have achieved a low number of false positives (high precision) while having a high sensitivity for non-juxtapleural nodules detection. However,[12]. this doesn't hold for sleuthingrespiratory organ nodules that areconnected to the pectoral wall (juxtapleural nodules).The developed CAD systems have difficulties in detecting juxta-pleural nodules. Therefore, these systems should be further developed in juxtapleural nodules detection. Despite that CAD systems could make huge improvements in lung cancer detection in the future, these systems should not replace radiologists or be used for final interpretation. The reason is that experience and expertise are always needed. Therefore, CAD systems should always remain as second opinion for the radiologists. CAD systems are supposed to assist radiologists and not supposed to replace them.

VI. REFERENCES

[1] Website: https://www.cancerimagingarchive.net

[2] L. Liang; Y. Si, "Medical image enhancement using sliding weighted empirical mode decomposition," IEEE International Conference on Information and Automation, pp.3145-3148, 8-10 Aug. 2015

[3] Sneha Potghan, R. Rajamenakshi, Dr. Archana Bhise, “Multi-Layer Perceptron based Lung tumor classification” Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018) IEEE Conference Record # 42487; IEEE Xplore ISBN:978-1-5386-0965-1.

[4] Ravindranath K, K Somashekar, “Early Detection of lung cancer by nodule extraction- A Survey”, 2017 International Conference on Electrical, Electronics,Communication, Computer and Optimization Techniques (ICEECCOT), 978-1-5386-2361-9/17/$31.00 ©2017 IEEE.

[5] Akahisa TANAKA, Noriaki MIYAKE, Huimin LU, Joo Kooi TAN, Hyoungseop KIM, “Detection of Lung Nodules on Temporal Subtraction Images Using 3D Sparse Coding” 2017 17th International Conference on Control, Automation and Systems (ICCAS 2017) Oct. 18-21, 2017 in Ramada Plaza, Jeju, Korea.

[6] May Phu Paing ; Somsak Choomchuay , “A Computer Aided Diagnosis System for Detection of Lung Nodules From Series of CT Slices”, 2017 14th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), vol. 12, 978-1-5386-0449-6/17/$31.00 ©2017 IEEE.

[7] Moumita Mukherjee, Pradyut Kumar Biswal, “Segmentation of lungs nodules by iterative thresholding method and classification with Reduced Features”, Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018) IEEE Xplore Compliant - Part Number: CFP18BAC-ART; ISBN:978-1-5386-1974-2\

[8] Nejla Jbeli ; Rekka Mastouri ; Henda Neji ; Saoussen Hantous-Zannad ; Nawres Khlifa , “Detection and Characterization of Subsolid Juxta-pleural Lung Nodule from CT Images”, 2018 5th International Conference on Control, Decisionand Information Technologies (CoDIT’18), 978-1-5386-5065-3/18/$31.00 ©2018 IEEE.