A Versatile Segmentation Approach For Diagnosis of Lung Cancer

N.Malligeswari[1],Dr.G.Kavya[2]
malligeswari.n@gmail.com[1],kavyavimal@gmail.com[2]
Department of Electronics and Communication Engineering
SKREC[1],SAEC[2], Chennai

Abstract. The necessary and critical step is to evaluate the development of lung cancer and nodule segmentation. Immobile challenge in the field of segmentation and classification of pulmonary lung nodule is particularly used to identify the small size nodule. To improve and sustain the diagnosis analysis, this paper puts forward and widens a new approach to segmentation and classification method for lung nodule size less than 3mm. In this paper, we examined and proposed a new method based on transition region based P-Tile thresholding and followed by Watershed processing for segmentation. First we reap the ROI from the input CT image and enhance the region of nodule by median filtering algorithm. Second Object contours are obtained by transition region based analysis. Third to extract multiple objects ROI from object contours employ M-Type morphological operation. Fourth prepare the images for segmentation by reducing noise and smoothing operations like weighted average filter. Kuwahara filter is used to smoothen the images and to preserve the edge position. Then we make use of crack code analysis to renovate lung boundaries. Finally the result is obtained by overlap the extracted image with the restored lung mask. To evaluate the novel segmentation method examines 90 lung nodules with 3mm to 9mm diameter from LIDC database. The presented novel approach attain ground truth rate of 86.93% ±0.09 with false positive rate of 15.09% ±0.06.After evaluation and investigation the results of segmentation our proposed method outperformed compared to other literature. .

Keywords: Lung cancer, Transition Region, Lung nodules, MHD, CT image, Crack code, Chain Code.

1 INTRODUCTION

According to the statistical study of American Cancer Society [1] in 2019, there are 245,590 new cases of lung cancer in the United States, and demise because of lung cancer are approximation around 299,080. Amongst all type of cancer, lung cancer is the foremost cause of cancer decease which includes both men and women, i.e. lung cancer cause about one cancer death out of five. The total of breast, prostate and colon cancer are less than the Lung cancer every year. The continued existence rate of lung cancer is four year, which is very unfortunate because the finding of lung cancer is typically done in later stages. Previously the estimate of CT scans was normally done all through manually ensuing very extensive tasks. So the clinical detection and treatment of diseases in managed way are required that would lead to increase of survival rates [2]. Thus a computer-aided diagnosis (CAD) system can be enormously helpful to decrease the manual overhead and to perceive the lung cancer in premature stages so that it can improve the survival rate of lung cancer patient and afford the better healthcare services in

medical imaging field. These facts paying attention of many researchers to develop an efficient and proficient mechanized CAD system that engage steps of processing the image to detect the cancerous cell and proposed analysis to diagnose [3]. The diverse common non invasive imaging modalities are computed tomography (CT), low-dose computed tomography (LDCT); contrast enhanced computed tomography (CE-CT) and Positron emission tomography (PET) for detection and diagnosis of lung cancerous cell.

The flow diagram of common CAD system for the recognition of lung cancer Detetion system is as shown in Figure 1. The key input to the CAD systems is the image obtains by precise imaging modalities. The preprocessing of input image follow the below steps may perform conversion and precise methods of operations and recognition of window size end with thorax part elimination. The lung segmentation is vital and important step; it mainly extracts the segment of lungs from CT chest image for finding the precise ROIs. The next step is to identify and section the nodules in segmented lungs and finally the recognized nodules are classified into cancerous and non-cancerous on the basis of texture, shape, appearance and growth analysis followed by diagnosis.
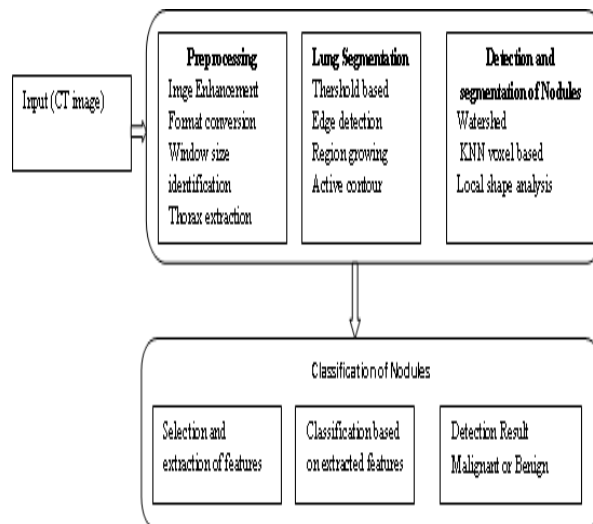

Fig 1.CAD for Lung cancer Detection

2 .EXISTING SYSTEM

The lung segmentation is an important and preprocessing step that reduce the search space and decrease the overhead for additional process i.e. recognition of the nodules. It is procedure of extraction of the lungs in CT chest images and separates the ROIs from, fats, muscles and other fond of pulmonary structures. The exactness of segmentation is very important and well stated by S. G. Armato and W. F. Sensakovic that due to poor and imprecise lung segmentation roughly 6.9 % to 19% true lung nodules are not detected and missed [2]. Though, the existence of noise, low contrast and intensity in homogeneity and in CT images makes the lung segmentation a complicated task but the most challenging problem for researcher

is to identify the nodules attached to pleural surface (i.e. juxtapleural) and to take account of this region in the segmented lungs. Numerous algorithms have been projected in the literature, such as, region growing based segmentation, thresholding-based segmentation, active contour models-based segmentation, segmentation based on fuzzy c-means (FCM) clustering approach, Level set image segmentation, Intensity-based segmentation, Texture-based image segmentation, and so forth. The commonly used lung segmentation techniques are: Active contour-based Segmentation: Minimization of the energy function in every iteration using some dynamic contour is the central idea behind active contour model (ACM). The difficulty in accurate segmentation owes to lack of organ tissue homogeneity in texture and variety of shape of various CT slices of original image. Initialization and broke convergence associated problem also limit the efficiency of active contour technique, which has been deal with many research studies [3, 4].

One of the frequently used segmentation techniques is Otsu thresholding method, but the main disadvantage of this method time consuming in both processing and operation computation. Kim et al [6] proposed an adaptive Otsu thresholding method of segmentation histogram based inspecting the region of bimodality and result improved performance. Kockelkorn proposed user collaborating technique for lung segmentation in CT image with severe irregularity shape nodule where prior shapes are trained using k-nearest-neighbor classifier and on the basis of classification results can be corrected. Kumar et al. [7] presented two stage approaches for automatic lung lobes segmentation wherever fissure region is found then fissure location and curvatures are identified within these regions. Sun et al. [5] proposed novel approach strong active shape model to approximately segment the outline of lungs, then optimal surface finding method is used to further adapt the initial segmentation result. Gill et al. [8] given the feature-based atlas approach for initialization of active shape models for segmentation.

2.1 NODULE SEGMENTATION AND CLASSIFICATION

The lung nodules are white circular shape matter having lower contrast found in lung region. The most challenging tasks for the segmentation of such nodules are their ominous locations. The other type of nodules which are sub-solid in temperament and usually having lower HU values as compared to other nodules are called as "ground-glass opaque" (GGO) nodules. Also, some nodules are small in size but having critical role for the prior detection of cancer in lung region. This method can be implemented in global scenario where firstly complete segmentation of lung from CT slices then the result is agreed to as negative mask to avoid the unwanted portion label reflect on in segmentation results. Morphological operations are the most accepted methods addressed in [9].

Summarizing about the nodules all authors cited here have their opinion that juxta pleural and sub-solid nodules are very tough to characterize and obtainable most demanding task for perfect segmentation. The classification of nodules to be malignant or benign in a computerized system the common approach followed by various authors in number of research studies are based on domain specific features. The most universal features considered for lung nodules are depend upon the nodules, appearance, textures and shape etc.

The steps followed for the classification are: i) selection and extraction of useful features, ii) extracted features are ordered, analyzed and processed by some detailed classifier algorithms e.g. SVM, LDA, ANN, etc., iii) classifier algorithms are designed and implemented based on features, iv) sampled data sets are trained using classifier algorithm allowing for the extracted important features and the outcome are validated with some benchmarks or accuracy v) as a final point in testing part entire data sets are process and nodules are classified into malignant or benign and further the results are observed with ROC analysis and subsequently criteria may be followed for the performance augmentation i.e. fall of false positives.

Gradient vector flow (GVF) as an external force used for analysis as an important answer by Chenyang Xu [10] where both the problems convergence and initialization has been addressed. In this direction additional improvement has been done by Wang et al. [11] with approach called generally biased GVF (NBGVF) as an external force. This technique keeps available chosen property of GVF with frail edges and smoothen the noise. The connected work for improving the convergence has been also addressed in [20]. An additional reason for the poor convergence is the minor searching space. Mumford–Shah function [12] pass over the need of specific limit and doesn't yet entail for smoothening of the original image to detect the object handling. Thresholding Based segmentation: Lungs parts are darker as compared to further anatomical structure in the thorax section so with the selection of finest threshold the segmentation of lungs has been exploiting in number of research studies. Hu et al. [13] acknowledge the iterative threshold with some morphological operations. Gao et al. [14] projected the threshold-based method with a numeral number of preprocessing, morphological smoothing and region growing.

More number of nodule segmentation techniques have been examined still some issues are challenging. Lung parenchyma modification is done by rolling ball algorithm depends on radius because of different size of nodule and need a intervention of human. Lung rectification based on convex hull technique will also fail when the nodules present in the cardiac and mediastinum region. Machine learning based segmentation algorithm is suitable for big data and sensitive based on nodule types[15][16]. In view of confines of the segmentation algorithms in the literatures we proposed segmentation based on transition region analysis and crack code method. Nodule regions are located by drawing rectangular region and implementations based on a rectangular support for the filter. Then the nodules are automatically segmented after the selection of ROI. The proposed segmentation algorithm involves two stages. Transition region based multiple threshold segmentation and crack code method for lung contour modification. To obtain the improved enhanced lung contour Kuwahara filter based smoothing operation is performed before crack code method. The novel segmentation techniques intend to define nodules, unclear boundary and also identify proper parenchyma boundary.

3. PROPOSED SYSTEM

The main aim of this study is proposing and implement a new segmentation method for finding a very small size pleural pulmonary nodule.
The proposed work is organized as follows. Methods, experimental Result,data analysis report, advantages and restrictions of this work. Finally conclusions and future works are discussed.

3.1 METHODS

In Lung CT images the pulmonary or pleural nodule attached to the chest wall. No Intensity difference between the chest wall and adjacent nodules, both are similar and it is very difficult to outline the boundaries. So, the main key problem in segmentation is set the border line between the chest wall and lung nodule. In this proposed method image transition region-based threshold segmentation algorithm and crack code techniques are combined to identify the nodule boundary. First describe the center position and edge of the nodule by analyzing XML. The pixel selection in ROI is based on interactive basis.

3.1.1 PREPROCESSING LUNG CT IMAGE

To remove the annoying noise and edge preprocess the ROI by median filtering. Compared to anisotropic diffusion and spatial filtering linear diffusion it keeps the boundary data and applied before detecting edge. The performance of the median filter is similar to weighted averaging filter as a pixel in output image is replaced by calculating the neighboring pixel average value. This filter is sensitive to obtained mean value. Based on boundary information to smooth the edges linear isotropic diffusion filter is used.

$$c.\Delta u \equiv c(x,y,I).\Delta u$$
$$c(x,y,I) = g(|\nabla\,|(x,y)|^2)$$

If bordersare more distinct in de- noised image apply nonlinear isotropic diffusion because the diffusion tensor depends on x.

$$\frac{\partial x}{\partial y} = div\ (g(|\nabla I|)^2)\nabla u$$

3.1.2 TRANSITION REGION (TR) BASED P-TILE THRESHOLD SEGMENTATION:

The segmentation of lung nodule is obtained by the following procedures. First by calculating the local variance using global threshold values and the transition regions are extracted. If the transition region is greater than width of one pixel, the thinning operation is performed to extract single pixel width. The object contours are obtained by edge linking process. After the object regions are finding by applying morphological operation. Final step is separating the object regions from the background in original image.

3.1.3 TR EXTRACTION:

Region finding in between the object nodule and background chest wall with several pixel width are defined as transition region. Due to recurrent and intensive variation of pixels in TR, local variance is greater compared to regions with small change in gray value. Transition regions are extracted by using local variance as a descriptor. The extracted TR manipulates the conclusion of segmentation output. To select suitable threshold Li et al proposed a better

segmentation algorithm based on salient transition region which suitable for extracting foreground.

Hence after preprocessing the image the foreground extraction is implemented by salient transition region based algorithm. First define the local variance by (i) considering matrix X (ii) define a window size 3x3 (iii) Pad with zeroes on all sides (iv) Find the local mean by sliding 3x3 window on the matrix (v) Find the local mean of the matrix X^2 and (vi) Find local variance =local mean of X^2-(local mean of X)^2

To extract ample number of pixels the α value is select between 0.1 to 0.5. The highest overlap rate is obtained by setting α=0.2.

$$\text{If } Lv(i,j) < \alpha \ Lv_{max} \text{ then } TR(I,j)=0 \text{ else } TR(I,j)=1$$

Sometime the edge generated by the previous methods does not result fully connected pixels. To obtain continuity edge linking process is applied. From discontinuous end points connectivity is checked and junction is encounter and denoted by number. Different labeled junction distance is calculated. If distance is less than 10 link the junctions. The unlabeled pixels are isolated. Finally extract the continuous object contours.

3.1.4 LUNG CONTOUR CORRECTION:

The lung contour gain from the STR methods result with unwanted edge spikes. To remove the spikes morphological filtering operation is used on the object contours. Region filling operation is used to extract perimeter to separate the region from background. Lung contours are repaired to obtain the intact boundary by following steps. (i) Iterative weighted mean filter (ii) crack and chain code analysis.

3.1.5 WEIGHTED AVERAGING MEAN FILTER:

In this stage the novel weighted averaging mean filter is proposed to identify the noisy pixels in the detection stage. In this method the pixels divided in to clean clusters and pixels of noisy. We propose a novel weighted averaging mean filter (NWAMF) remove high level noise. The window size is increased continuously until the values of equal successive windows.The quality of the image is improved by the proposed method. Weighted averaging mean filter is better than weighted median filter. The comparison of PSNR result shows that noisy pixels replaced by weighted mean filter (WMF) give higher than the weighted median filter (WMN).

$$PSNR = 10\log_{10}(255^2/MSE)$$
$$MSE = 1/MN \ E(x_{i,j}-z_{i,j})^2$$
$$\text{Error rate} = \neq(C)/MN \ X \ 100\%$$

By considering standard deviation weight assigned to the largest groups. If $x_{i,j}$ is noisy pixel at i,j

$$X_{i,j} = \text{mean } \{w1 \ \lozenge \ s1, w2 \ \lozenge \ s2\}$$

◊ shows the replication operator known as replication operator. Further improving the filter from the selected groups by consider closest neighbor pixels around central pixel then take as mean value.

3.1.6 CHAIN CODE ANALYSIS:

To facilitate the measurement and manipulation on the object more compacted representations are provided. Chain code and crack code is the techniques used to represent the object by describing its contour.To implement the chain code it has a non object (back ground) pixels four connected neighbors. Consider a is the current pixel position in contour and the codes are defined as

$$
\text{Chain code} =
\begin{matrix}
3 & 2 & 1 \\
4 & a & 0 \\
5 & 6 & 7
\end{matrix}
$$

Get the image and estimate the background, filtration finally convert it to binary. If $P(I,j)=0$ then $I(I,j-1)=0$ else set $I(I,j-1)=1$ The direction of the pixel with 0 in 4 connected is connected to 0.thenIf $P(I,j)=0$ then $I(I,j+1)=0$ else set $I(I,j-1)=1$ the direction is changed from 0 t0 1 The direction of the pixel with 1 in 4 connected is connected to 0. For the each object k find the max and min value in the newly obtained image in X axis. Here the minimum of i is referred as n_{1k} for all $I(I,j)=0$ and maximum of i is referred as n_{2k} for all $I(I,j)=1$. Next consider the image in column wise and ignore the first and last column and repeat the above steps to set for direction 2 If $P(I,j)=0$ then $I(I+1,j)=0$ else set $I(I+1,j)=1$.For direction 3 If $P(I,j)=0$ then $I(I-1,j)=0$ else set $I(I-1,j)=1$. For the each object k find the max and min value in the newly obtained image in y axis. Here the minimum of j is referred as m_{1k} for all $I(I,j)=0$ and maximum of j is referred as m_{2k} for all $I(I,j)=1$.The eight zero connected pixels is detected. Perform the XOR logical operation at every pixel location (I,j).From the above procedure by obtain the corner points the object is extracted. The main advantage of this method is it produces a contiguous with closed border line.

The algorithm is concluded as follows. 1. Transition region of the input CT image is Extracted.2. Morphological operation is used to obtain the edge from the first step. 3. By using edge linking processing methods object contours are obtained. 4. Object region is extorting from the image by filling. 5. To take out the object overlapping the pixels of original and object region.

4. EXPERIMENTAL RESULTS

The proposed algorithm was implemented by using MatLab software. Analysis of our proposed approach statistical measurement of calculated False positive (FP), True Negative(TN), False Negative(FN), MHD,OLR and True Positive(TP) compared to previous methods. Modified Hausdorff distance measurement provides superior image comparison measure. Segmentation

approach is still challenging and difficult to test and getting ground truth segmentation. False positive (FP) is pixels falsely segmented as foreground. These metrics are then used to calculate sensitivity .The sensitivity shows how likely the test is come back positive. This is calculated as TO/(TP+FN) and the accuracy is calculated as (TP+TN)/(TP+FP+TN+FN) and MHD is calculated

$$\mathbf{DT_X}[i,j] = \min_{k,l;X[k,l]=1} d([i,j],[k,l])$$

D(i,j),(k,j) is the Euclidean distance calculated by standard choice. The outcomes of our new approach give up considerably greater performance than the recent studies.
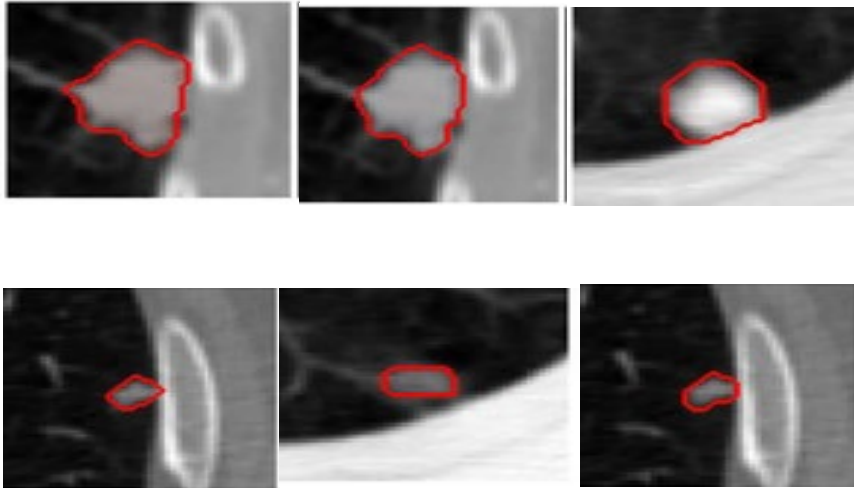
5. SIMULATION OUTPUT

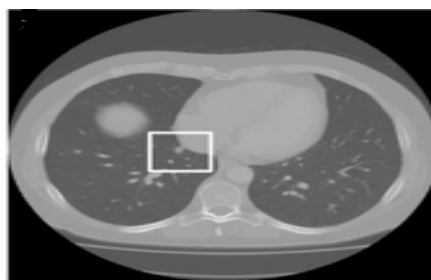**Fig 2.**Segmentation result and Ground truth
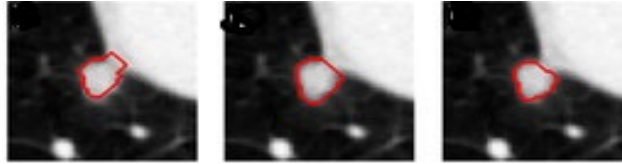

Fig 3 Input Nodule image

Fig4. Segmentation Result and ground truth

The different Performance Metrics (MHD, overlap rate, false positive rate and sensitivity,) for the test data are shown in Table 1. The tabulations are shown in percentage, each column indicates the segmentation of previous and proposed method used and the rows indicate the Metric value.

Table 1: Performance Metrics for test data

		Segmentation		
		Mukhopadhyyay's method	Proposed method	P value
Metrics	FPR	20.30%±0.19	13,56%±0.04	0.02
	OR	66.29%±0.13	77.82%±0.02	0.8×10^{-4}
	sensitivity	77.0%±0.16	87.54%±0.02	0.0003
	MHD	2.42±2.04	1.002±0.43	0.3×10^{-4}

From table 1 it is shown based on experimental result the new methods shows better result compared to previous study. This method attains a lower MME (0.042 vs 0.066) value indicates new method is better than Mukhopadhyay's method
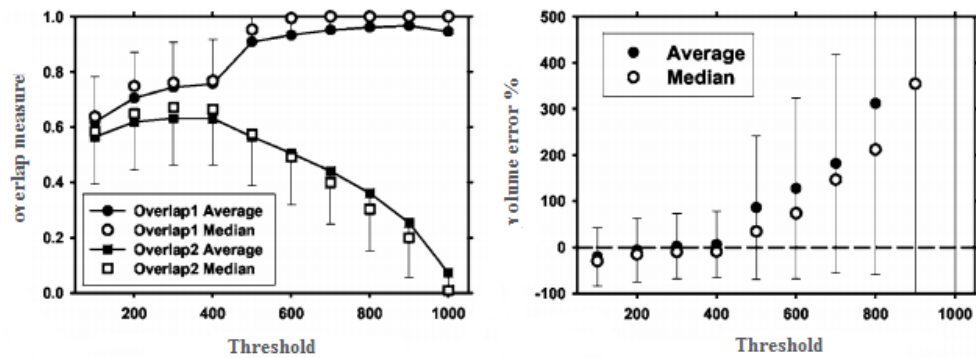

Fig 5. Overlap measures (a) and volume percentage errors (b) at various thresholds for testing of the segmentation using the 32 LIDC nodules.
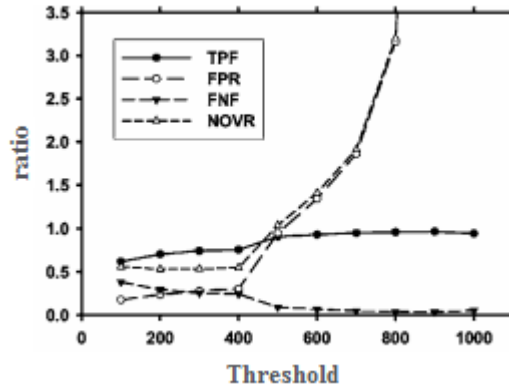
Fig 6. The performance metrics of True positive, false positive, false negative and non overlapping
volume ratio

Fig.6 as resulting from the middling Overlap (A,L) and V_{err} values for the LIDC nodules shown in Fig. 5.

CONCLUSION

Most of the medical images are complex for segmentation and the recent segmentation algorithms obtained more agreeable accuracy but with more error. This is because of weak and omitted artifacts and edges and low SNR. Accurate segmentation is impossible due to ground truth may be missing. For overlapping robust algorithm has been suggested to obtain better result. Our proposed method shows that the new approach statistically reduces the MHD. First the image is converted in to gray scale image and it passed through a filter to augment the needed information. The image is further improved by median filtering. To reduce noise salt and pepper noise is added and restored the damaged edge. To perform segmentation to identify the small nodules TR based analysis and chain code techniques are used. The foreground pixels in region of interest are obtained by applying TR based P-Tile thresholding segmentation. Then contour is corrected by combining iterative weighted averaging median filtering with chain code techniques. The segmentation result is obtained by multiplying the corrected lung parenchyma and the obtained foreground pixels. The new approach of segmentation evades the boundary leakage and inadequate segmentation. In future optimized algorithm is use to strengthen the segmentation with large datasets.

REFERENCES

[1] American Cancer Society, Cancer facts and figures, 2019.

[2] S. G. Armato and W. F. Sensakovic, "Automated lung segmentation for thoracic CT," Acad. Radiology., vol. 11, no. 9, pp. 1011– 1021, 2004.

[3] Preeti Joon, Shalini Bhaskar Bajaj and Aman Jatain, "Segmentation and Detection of Lung Cancer Using Image Processing and Clustering Techniques," Springer Nature Singapore Pte Ltd. 2019.

[4] Laurent D Cohen, "On active contour models and balloons," CVGIP: Image understanding, vol. 53, no. 2, pp. 211–218, March 1991.

[5] Sun, Shanhui, Christian Bauer, and Reinhard Beichel. "Automated 3-D segmentation of lungs with lung cancer in CT data using a novel robust active shape model approach." IEEE transactions on medical imaging 31.2 (2012): 449-460.

[6] Cheol-Hwan Kim, Yun-Jung Lee, "Medical Image Segmentation by Improved 3D Adaptive Thresholding" 978-1-4673- 7116-2/15/$31.00 ©2015 IEEE

[7] Kumar, S. N., and V. Kavitha. "Automatic segmentation of lung lobes and fissures for surgical planning." Emerging Trends in Electrical and Computer Technology (ICETECT), 2011 International Conference on. IEEE, 2011.

[8] Gill, Gurman, Matthew Toews, and Reinhard R. Beichel. "Robust initialization of active shape models for lung segmentation in CT scans: A feature-based atlas approach." Journal of Biomedical Imaging (2014)

[9] W. J. Kostis, A. P. Reeves, D. F. Yankelevitz, and C. I. Henschke, "Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images," IEEE Transactions on Medical Imaging, vol. 22, no. 10, pp. 1259–1274, 2003.

[10],pp. 577–685 [5] Chenyang Xu and Jerry L Prince, "Snakes, shapes, and gradient vector flow," IEEE Transactions on Image Processing, vol. 7, no. 3, pp. 359–369, 1998.

[11] Yuanquan Wang, Lixiong Liu, Hua Zhang, Zuoliang Cao, and Shaopei Lu, "Image Segmentation Using Active Contours With Normally Biased GVF External Force". October 2010.

[12] Alaa Khadidos1, 2, Victor Sanchez1 and Chang-Tsun Li1, "Active contours based on weighted gradient vector flow and balloon forces for medical image segmentation" 978-1-4799-5751-4/14/$31.00 ©2014 IEEE

[13] S. Hu, E. A. Hoffman, and J. M. Reinhardt, "Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images," IEEE Transactions on Medical Imaging, vol. 20, no. 6, pp. 490–498, 2001.

[14] Q. Gao, S.Wang, D. Zhao, and J. Liu, "Accurate lung segmentation for X-ray CT images," in Proceedings of the 3rd International Conference on Natural Computation (ICNC '07), vol. 2, pp. 275–279, 2007.

[15] Vinod Jagannath Kadam, Shivaji rao Manik rao Jadhav, K.Vijayakumar, "Breast Cancer Diagnosis Using Feature Ensemble Learning Based on Stacked Sparse Auto encoders and Soft max Regression", Image & Signal Processing, springer, june 2019.

[16] K. Vijayakumar , K. Pradeep Mohan Kumar ,Daniel Jesline, "Implementation of Software Agents and Advanced AoA for Disease Data Analysis", journal of medical systems, Part of Springer Nature 2019.

[17] Sóñora-Mengana, A., Papavasileiou, E., García-Naranjo, J., et al.: 'Evaluation of data balancing techniques. application to cad of lung nodules using the luna16 framework', Revista Científica de Ingeniería Electrónica, Automática y Comunicaciones, 2018, 39, (3), pp. 57–67, ISSN: 1815-5928

[18] Jue Jiang,'Multiple Resolution Residually Connected Feature Streams for Automatic Lung Tumor Segmentation From CT Images', IEEE Transactions on Medical Imaging, Volume: 38, Issue: 1, Jan. 2019, 134 - 144

[19] X. Ye, X. Lin, J. Dehmeshki, G. Slabaugh, G. Beddoe, Shape-based detection of lung nodules in thoracic CT images, IEEE Trans. Biomed. Eng. 56 (2019) 1810–1820.

[20] Alaa Khadidos1, 2, Victor Sanchez1 and Chang-Tsun Li1, "Active contours based on weighted gradient vector flow and balloon forces for medical image segmentation" 978-1-4799-5751-4/14/$31.00 ©2014 IEEE