A Model: Lung Nodule Detection and Classification by SVM Network

Nuthanakanti Bhaskar[1], Dr. Ganashree T S[2]

[1]Research Scholar, CSE Dept., VTU-RRC, Visvesvaraya Technological University, Belagavi, India.

[1]Assistant Professor, CMR Technical Campus, Hyderabad, Telangana.
[2]Associate Professor, Department of Telecommunication Engineering, Dayanandasagar College of Engineering, Bangalore, India.
[1]bhaskar4n@gmail.com, [2]ganashreeec@gmail.com

Abstract

In this work, we suggest a unique system for pulmonary Nodule Awareness, Segmentation, and Identification of Nodules in CT images. This process consists of 4 phases: Phase-1: Segmentation of lung location by ACM (Active Contour Modeling). Phase-2: Apply of mask strategies. Phase-3: Shift from non-separated nodules to separated nodules. Phase-4: SVM (Support Vector Machine) classifier identifies nodules with the help of 2D hypothetical and 3D anatomical characteristics. Shape of nodules extracted by ACM. Therefore, all cavity and solid nodule segmentation accuracy is high. Moreover, lung tissues divided into 4 groups: a. Lung Wall, b. Parenchyma, c. Bronchioles, d. Nodule. Accordingly, this system implemented in MATLAB Software and accuracy checked with different efficient methods with publicly available data sets (LIDC, LUNA16).

Keywords: *Segmentation, Connectivity Identification, ACM and SVM Classifier.*

1. Introduction

Lung cancer is dangerous for men and women [1]. In the US, annual basis 2,19,000 lung cancer patients diagnosed [2]. Irregularity of lung shows cancer and the growth of a nodule in lung region. Most of the radiologists consider distinct clinical nodule limitations to the diagnosis of lung cancer. Parameter-1 (nodule shape): A jagged structure nodule appears like lung cancer than flatten one. Parameter-2 (nodule texture): If it is a mixture of fatty, bony, watery then the notion degree of lung cancer would be different. Parameter-3 (nodule location): For example, vessel attached nodules are probably to be lung cancer than lonely nodules [3]. Parameter-4 (growth rate of the nodule volume): For instance, lung edge or cracks attached nodules mostly benign with volume increasing twofold time longer than 400 days. Using these parameters to perform Lung detection, segmentation and volumetric procedures by humans is a difficult task, inexact and it takes more time. Accordingly, a complete automated model required for above processing.

2. Related Work

In lung cancer research, every year, numerous methods suggested for Segmenting of Lung and Detection of Nodule. Nevertheless, only a few algorithms are advising for Segmenting of Lungs then recognizing of Nodule. Many techniques effectively work with fewer parameters for the identification of non-isolated nodules that are connected to the Lung region [4, 5]. Several nodule detection algorithms work with satisfactory performance, but they detect only pulmonary lesions [6]. Although some methods are poorer for cavity detection, but helps in solid nodule detection [7]. Very few methods are available for the recognition of nodule or non-nodule in the procedure of segmentation [8, 9]. In addition, some methods classify nodules by spatial positions [3, 10] and some nodules based on texture and outline [11].

Ye et al. [7] suggested a CAD Method based on lung nodule shape for CT scans. Applied flexible 3D thresholding to access inceptive lung mask. Using chain code, the authors estimated the complete lung mask. Moreover, this method detected even smaller nodules which attached lung wall. This method failed when big-sized non-isolated nodules connected to the lung region because it will appear as a piece of the lung region after executing fuzzy thresholding (3D) on the lung CT image. Therefore, this method detection rate for Solid Nodules: 94%, GGO Nodules: 88.2%, and 8.2 value of FP for each scan.

Homma et al. [12] suggested lung segmentation method by ACM (active contour modeling) to identify non-isolated nodules. In this work, researchers identified every slice in the data set is close to its past slice. Consequently, they executed active contour on the upcoming slice by using a lung mask of the present slice as start mask. This segmentation of lung area helps for slices which are non-isolated nodules only. Consequently, it is troublesome for thickened slices. Accordingly, after joining CAD system TP (true positive) value improved from 7 to 10 and FP also increased by 5%.

Taghavi et al. [11] applied F-KNN (Fuzzy k-Nearest Neighbor) for nodule detection and malignancy degree. In this work, researchers calculated the structure and magnitude of nodules by using geometric and intensity-based characteristics. In this, nodules detected competently, but conversely outline of recognized nodules not extracted exactly. Wu et al. [3] Calculated the relationships of the nodule and further shapes for connectivity classification by probability co-occurrence map.

3. Our Approach

In this work, phase-1: a lung partitioning method working ACM with suitable assignment is suggested, In Phase-2: An efficient algorithm for detecting nodules by using the SVM classifier is formulated, In Phase-3, ACM (Active Contour Modeling) provided to segmented for lung and Finally Phase-4: A nodule identification algorithm by classifying and marking lung tissues are suggested.

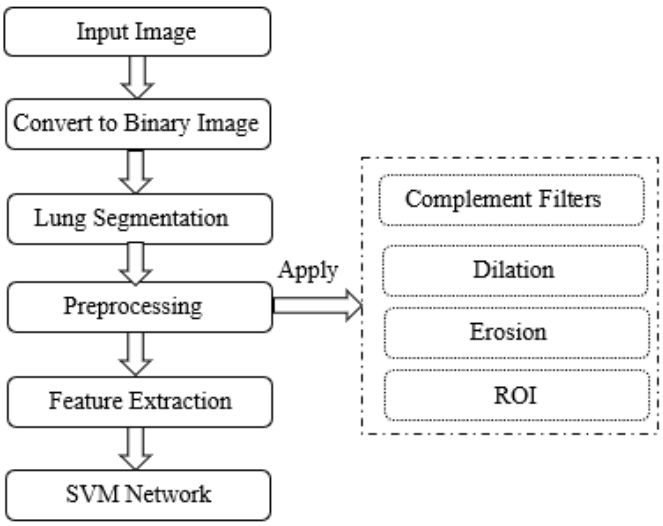

Figure1. Proposed Methodology

Lung segmentation is the main phase of nodules identification in the lung region. During some CT scans, non-isolated nodules enclosed to the breast region, and also which elaborates the lung distribution action.

In this approach, we identified nodules and also classified lung region, bronchioles, and parenchyma (i.e thorax tissues) by using adjacent nodule pixel labels. Consequently, nodule classes identified. Moreover, for validating and assessment of our work, we used 2 datasets: 1-ANODE 09 (http://anode 09.isi.uu.nl) and 2-LIDC (Lung Image Database Consortium).

4. Methodology

4.1 Lung Segmentation

Initially we attained binary lung CT scans by an adaptive fuzzy thresholding, later attained hole-free mask by using different dimensions of 2 windows. By sweeping of revolved window on the full hole-free binary image received the initial lung mask. In this phase, inside of the lung region non-isolated nodules (Small and Large) moved to isolated nodules. At last, automatic active contour model [13] implemented with initial mask. Consequently, lung area segmented accurately. Performance of first and last phases increased by using the latest method extended in [14].

By applying adaptive fuzzy thresholding, the binarization technique executed. The lung pixel intensities are lower than the chest walls. Consequently, we applied the binarization technique, it will assign 0 to lower intensity pixels and 1 to region attached pixels.

$$\mathrm{M_{high}(i)} = \sum_{k=i}^{I_{max}-1} \mathrm{kp(k)}, \quad 0 \le i \le I_{max} - 1$$

$$\mathrm{M_{low}(i)} = \sum_{k=0}^{i} \mathrm{kp(k)}, \quad 0 \le i \le I_{max} - 1 \tag{1}$$

Where p(k): histogram value of gray level k and i

Histogram splits 2 parts by gray level i. So, we calculated mean values of these 2 parts [7]:

$$\mu_{low}(i) = \frac{M_{low}(i)}{T_{low}(i)}, \qquad \mu_{high}(i) = \frac{M_{high}(i)}{T_{high}(i)} \tag{2}$$

$$m_i(t) = \frac{1}{1 + \left(d\left(t, \mu_{low}(i)\right)\right) / \left(I_{max} - 1\right)} \tag{3}$$

Here, mi(t) is the membership measurement function: By the gray level i for each gray level t, it determines one of two regions.

Where $d\left(t, \mu_{low}(i), \mu_{high}(i)\right)$ is enumerated as

$$d = \begin{cases} |\, t - \mu_{low}(i)\,|, & t \le i \\ |\, t - \mu_{high}(i)\,|, & t > i \end{cases} \tag{4}$$

By using d(.) function, we can compute the interval between each gray level (t) from lung region mean value.

$$C_i = \sum_{t=0}^{I_{max}-1} [m_i\,(t)\left(1 - m_i(t)\right)]^2 \tag{5}$$

Here, Ci is the cost function which calculates an optimum threshold at each gray level i.

4.2 Nodule Detection by SVM Classifier

In this, segmented lung classifies into 2 classes i.e nodule and non-nodule. Consequently, we detected nodules in lung images by the following 3 phases:

Phase-1: 2D & 3D Feature Extraction

Features of 2D stochastic are most successful for nodule detection by pixel values. However, the nodule gray level and some portions of bronchioles are divergent from the lung areas. Therefore, pixel mean values are elucidated neighborhood and computed as a feature:

$$M_{ij} = \frac{1}{9}\sum_{k,l=-1}^{+1} L(i+k, j+l) \quad (6)$$

Here, Pixel Coordinates are i, j and the Lung Segmentation Result is L.

Ordinarily, nodules recurrent itself in the same location in the place of before or after. As per the previous clinical analysis declarations, there is no chance of bronchioles in adjacent slices. Some of them may materialize. Moreover, 3D average computed by likeness of objective slice with adjacent slices.

$$M_{ij}^{p} = \frac{1}{9}\sum_{k,l=-1}^{+1} L^{p}(i+k, j+l),$$

$$M_{ij}^{+} = \frac{1}{q}\sum_{p=p+1}^{p+q} M_{ij}^{p}, M_{ij}^{-} = \frac{1}{q}\sum_{p=p-q}^{p-1} M_{ij}^{p},$$

$$3D\ averaging = M_{ij}^{-} M_{ij}^{+} \quad (7)$$

Where index p specifies the label of the slice, M_{ij}^{p} is the mean value of each window in the target slice, M_{ij}^{+} is mean values average of the same coordinates in the succeeding slices, and M_{ij}^{-} is mean values average of the same coordinates in the former slices. For the initial slice in every dataset: M_{ij}^{-} defined as M_{ij}^{+}, and final slice: M_{ij}^{+} defined as M_{ij}^{-}. Variable q defined as the number of considered slices. Moreover, it depends on dataset slice thickness. Accordingly, it defined as:

$$q = \frac{c}{T}, \qquad T: thickess\ of\ slices \quad (8)$$

Phase-2: SVM Classification

Originally, the SVM algorithm proposed by Vapnik [16]. In the binary classification we used SVM algorithms because it discovers the maximum-margin hyper-plane which splits the datapoints.

Let us consider,

$\{x_1, x_2, \ldots, x_n\}$ are training data points with d-dimensions.

$\{y_1, y_2, \ldots, y_n\}$ are labels where $y_i \epsilon \{-1, 1\}$.

In straightforward, the SVM algorithm divides training data by the highest margin because it has hyper-planes. Instances of training which are nearer to hyper-plane known as support vectors. The hyper-plane labeled 3 ways: a). -1: vector lying for one face of hyper-plane. b). +1: vector lying for another face of hyper-plane. c). 0: for Non-nodule, for proper execution 0 labels changed to -1.

Nevertheless, optimal hyper-plane construction required. Consequently, to minimization of error rate SVM uses a QuadProg method for iterative training. Accordingly, the error function defined as:

$$\frac{1}{2} w^{t} w + C\sum_{i=1}^{N} \zeta_i \quad (9)$$

Constraints of the subject:

$$y_i(w^{t}\varphi(x_i) + b) \geq 1 - \zeta_i,\ \ \zeta_i \geq 1, i = 1, \ldots, N \quad (10)$$

Where: C: capacity constant, w: vector of coefficients, b: constant and ζ_i : inputs (parameters for handling non-separable data).

Phase-3: Extraction of Active Contour-based nodule

In this work, proposed lung nodule segmentation by ACM (Active Contour Modeling) [13] for that basic idea and results in [14]. This model efficient for thin edges (boundary between nodule and lung).

So, it has the robustness and tested all types of images (medical and non-medical). Moreover, it provides great performance for noisy images [13,15]. In this ACM, the initial contour utilizes the outcome of the preceding segment works on the original image (I). In this lung segmentation process if small parts not detected it will be segmented by considering I as an input.

Phase-4: Identification of Nodule Connectivity

In most cases, lung tissues attached to the nodule. So, nodule location can identify by the lung tissue. Basically, the image classified into 4 classes: a. Lung Wall (LW). b. Parenchyma (PA). c. Bronchioles (BR). d. Nodules (ND). Consequently, the image will be segmented into 4 classes with high accuracy and the nodule will identify in 3 ways: 1. Nodule attached to Lung Wall: class ND connected to class LW. 2. Nodule attached to Bronchiole: class ND connected to class BR. 3. Solitary Nodule: class ND entirely covered by class PA. However, by considering nodule neighbor pixels, we can examine whether it connected to LW, BR, or Solitary.

Class ND can be using the previous section (Nod) results in a slice. The dilation method executed on the Nod slice because of likely errors in nodule segmentation. Consequently, we took a worthy structure element:

$$\mathrm{ND} = \text{dilate (Nod)} \tag{11}$$

In this Lung Segmentation Process result is called L and applied to the identified LW class which are considered unidentified pixels as the lung region:

$$\mathrm{LW} = \bar{\mathrm{L}} \tag{12}$$

By considering all pixels in image L as PA class and pixels in image Nod as BR class:

$$\mathrm{PA} \cup \mathrm{BR} = \mathrm{L} \cap \overline{\mathrm{Nod}} \tag{13}$$

We applied manageable thresholding to differentiate these 2 classes. Accordingly, considering histogram for sub-images containing only BR and PA classes. Consequently, the suitable threshold decided provably.

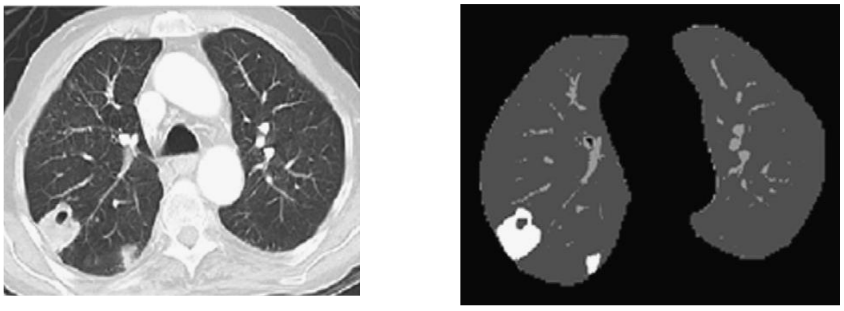

Figure 2. a) Original Image b) Lung tissue labeling

In above Fig.2 shows, Identification of Nodule connectivity, in that studied all gray values of adjacent pixels of the nodule. Consequently, assigned relevant labels to LW, PA, BR, and ND classes.

5. Experimental Results

In this experimentation, we presented outcomes of the recommended system and also some parameter values noted. For this process, we used 2 groups of data sets (LIDC and LUNA16).

This process uses 5 phases: 1) Read the input Lung CT image. 2) Binary format conversion. 3) Segmentation. 4) Preprocessing Techniques (Complement Filter, Dilation, Erosion, and ROI) for feature extraction. 5) SVM Network for Classification of Nodules.

The following process will detect nodule in CT scans by using the ACM Method and SVM Classifier:

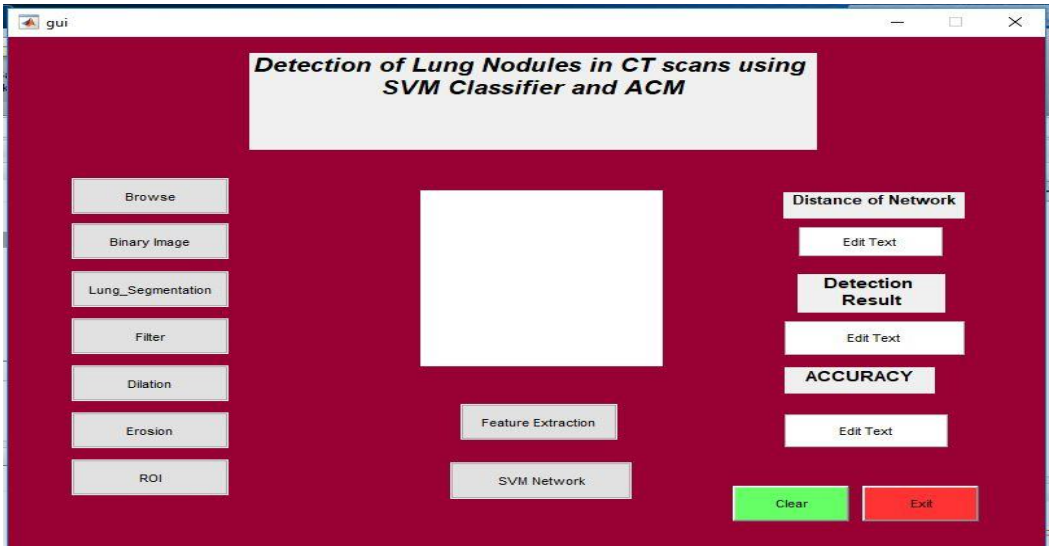

Figure 3. Lung Nodule Detection

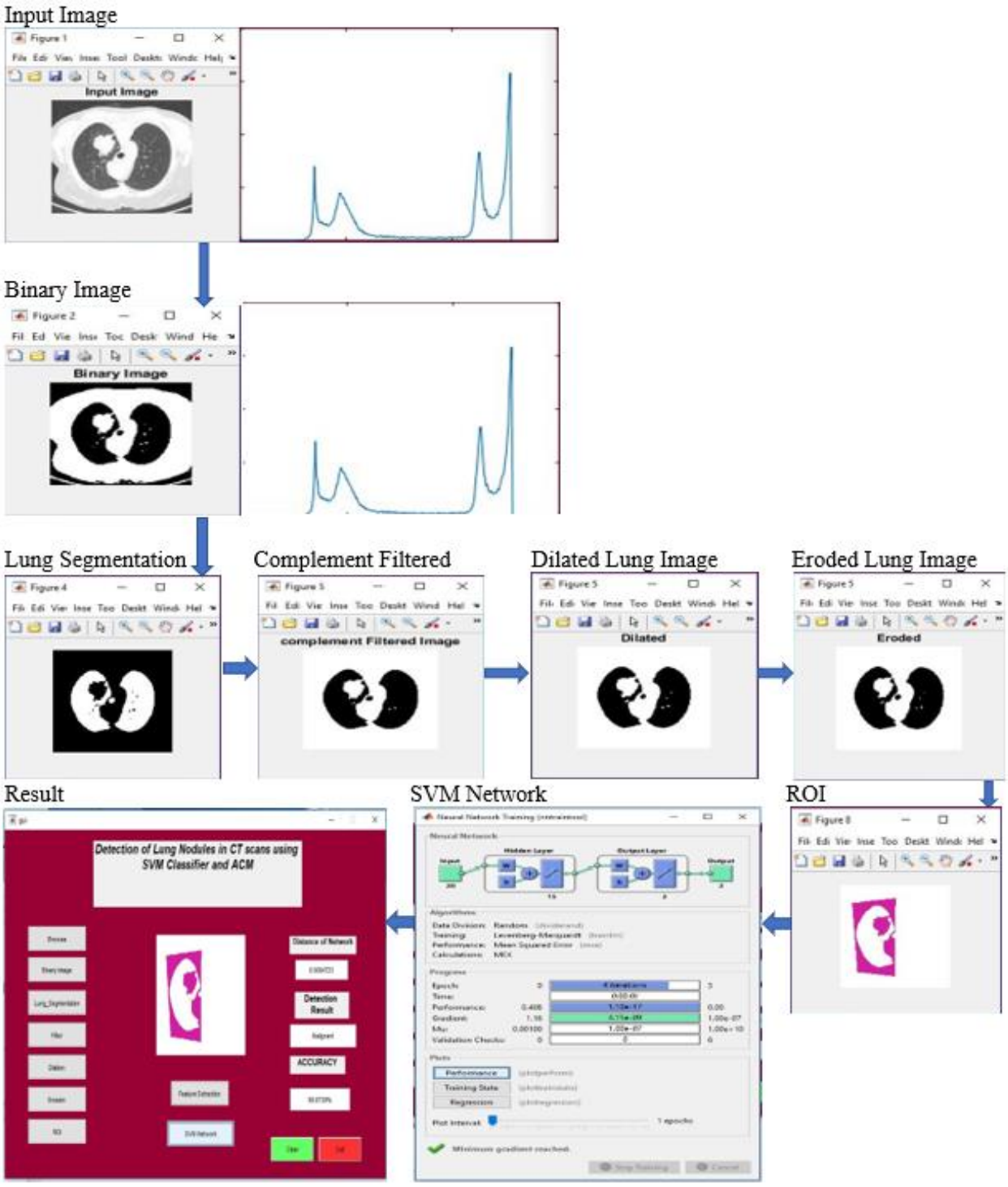

Figure 4. Results of the Recommended System

6. Discussion

In this process, we tested many images based on the analysis of Accuracy, Epoch, Performance, and Validation.

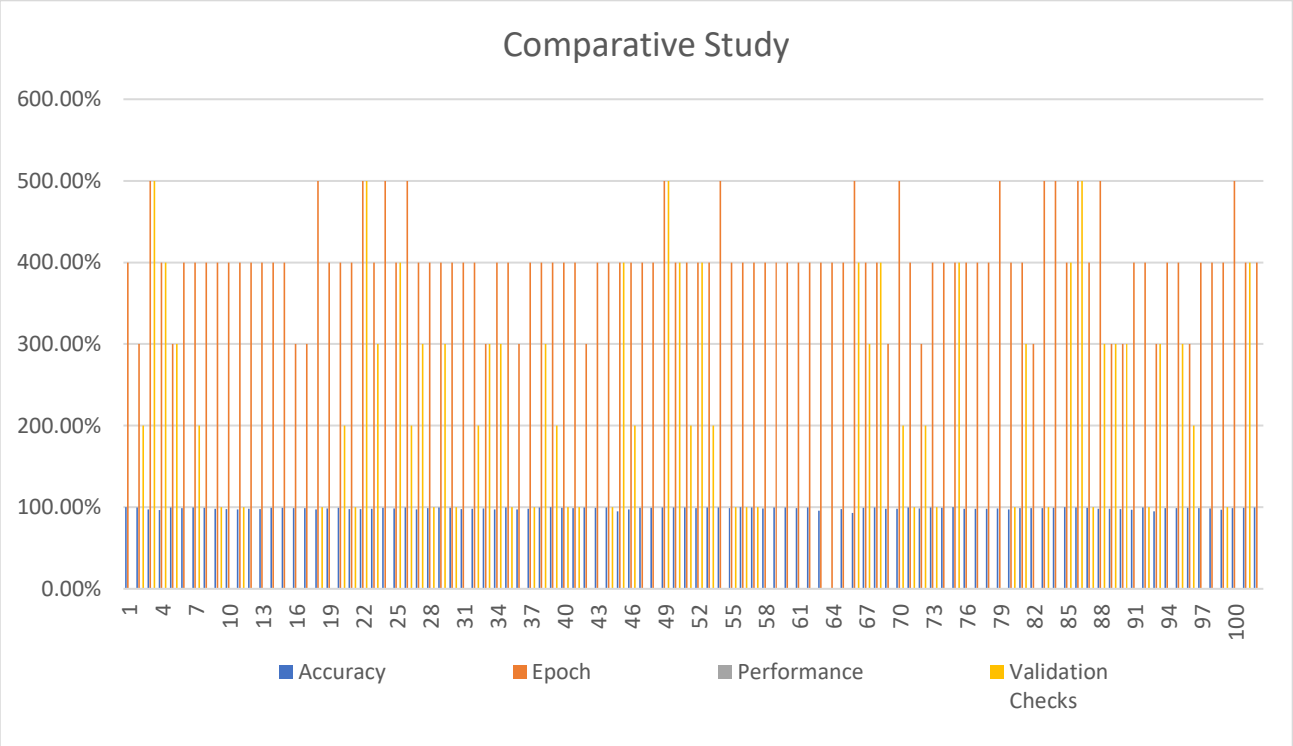

Figure 5. Comparative Study Chart for Accuracy, Epoch, Performance and Validation Checks

The following is some input images and results of our model respectively:

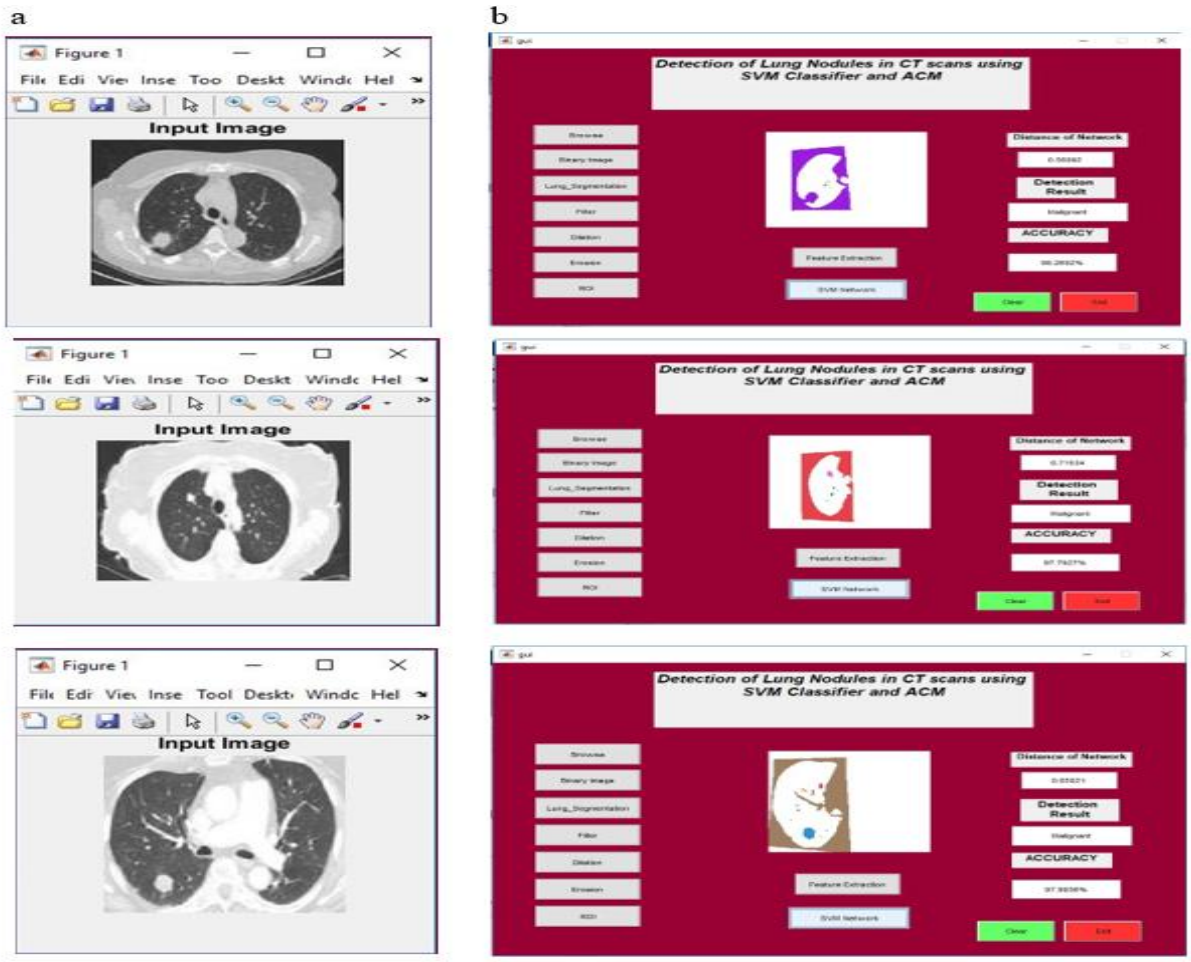

Figure 6. a) Input images of different patients. b) Results of our Model.

7. Conclusion

In this implementation, we proposed a model that detecting nodule automatically, Segmentation, and identification process. This process needs a single input lung slice of a particular patient. In the Segmentation process, lung regions divided into big and little non-isolated nodules based on ACM. For feature extraction, we applied Complement Filter, Dilation, Erosion, and ROI. Accordingly, we assigned relevant labels to LW, PA, BR, and ND classes. Finally, we applied the SVM Network on labels and classified the nodules are cancerous or not.

8. Acknowledgement

The successful completion of any task would be incomplete without an expression of simple gratitude to the people who encouraged our work. Though words are not enough to express the sense of gratitude towards everyone who directly or indirectly helped in this task. I thankful to my guide Dr. Ganashree T S, Associate Professor, Department of Telecommunication Engineering, Dayanandasagar College of Engineering, Bangalore and VTU-RRC, CSE Dept., Visvesvaraya Technological University, Belagavi and also My organization CMR Technical Campus, which provided good facilities to accomplish my work and would like to sincerely thank our Management, Director Dr. A. Raji Reddy, HOD Dr. K. Srujan Raju, my colleagues and parents for giving great support, valuable suggestions and guidance in every aspect of my work.

REFERENCES

[1] *Keshani, Mohsen & Azimifar, Zohreh & Tajeripour, Farshad & Boostani, Reza. (2013). "Lung Nodule Segmentation and Recognition using SVM Classifier and Active Contour Modeling: A*

Complete Intelligent System". Computers in biology and medicine. 43. 10.1016/j.compbiomed.2012.12.004.

[2] *American Cancer Society.Cancer fact and figures 2009. Available @ http://www.cancer.org.*

[3] *Wu, Dijia & Lu, Le & Bi, Jinbo & Shinagawa, Yoshihisa & Boyer, Kim & Krishnan, Arun & Salganicoff, Marcos. (2010). "Stratified Learning of Local Anatomical Context for Lung Nodules in CT images". Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2791-2798. 10.1109/CVPR.2010.5540008.*

[4] *Lei, Cao & Xiaojian, Li & Jie, Zhan & Chen, Wufan. (2008). "Automated Lung Segmentation Algorithm for CAD System of thoracic CT". Journal of Medical Colleges of Pla. 23. 215-222. 10.1016/S1000-1948(08)60045-3.*

[5] *Silveira, Margarida & Marques, Jorge. (2006). "Automatic Segmentation of the Lungs using multiple active contours and outlier model". Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference. 1. 3122-5. 10.1109/IEMBS.2006.260185.*

[6] *Kuhnigk, Jan-Martin & Dicken, Volker & Bornemann, Lars & Bakai, Annemarie & Wormanns, Dag & Krass, Stefan & Peitgen, Heinz-Otto. (2006). "Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans". IEEE transactions on medical imaging. 25. 417-34. 10.1109/TMI.2006.871547.*

[7] *Ye, Xujiong & Lin, Xinyu & Dehmeshki, Jamshid & Slabaugh, Greg & Beddoe, Gareth. (2009). "Shape-Based Computer-Aided Detection of Lung Nodules in Thoracic CT Images". IEEE transactions on bio-medical engineering. 56. 1810-20. 10.1109/TBME.2009.2017027.*

[8] *Dehmeshki, Jamshid & Amin, Hamdan & Valdivieso, Manlio & Ye, Xujiong. (2008). "Segmentation of Pulmonary Nodules in Thoracic CT Scans: A Region Growing Approach". Medical Imaging, IEEE Transactions on. 27. 467 - 480. 10.1109/TMI.2007.907555.*

[9] *Farag, Aly & El-Baz, Ayman & Gimel'farb, Georgy & Falk, Robert. (2004). "Detection and Recognition of Lung Abnormalities Using Deformable Templates". 738-741. 10.1109/ICPR.2004.1334634.*

[10] *Sofka, Michal & Stewart, Charles. (2010). "Location registration and recognition (LRR) for serial analysis of nodules in lung CT scans. Medical image analysis". 14. 407-28. 10.1016/j.media.2010.02.006.*

[11] *Taghavi, Sarah & Moghaddam, Hamid & Jafari, R. & Esmaeilzadeh, Mohammad & Gity, Masoumeh. (2010). "Automated detection and classification of pulmonary nodules in 3D thoracic CT images". Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics. 3774-3779. 10.1109/ICSMC.2010.5641820.*

[12] *Homma, Noriyasu & Shimoyama, Satoshi & Sakai, Masao & Ishibashi, Tadashi & Yoshizawa, Makoto. (2009). "Auto-detection of non-isolated pulmonary nodules in X-ray CT images". 396-401.*

[13] *Lankton, Shawn & Tannenbaum, Allen. (2008). "Localizing Region-Based Active Contours". IEEE Transactions on Image Processing. 17. 10.1109/TIP.2008.2004611.*

[14] *Keshani, Mohsen & Azimifar, Zohreh & Boostani, Reza & Shakibafar, Alireza. (2010). "Lung nodule segmentation using active contour modeling". 10.1109/IranianMVIP.2010.5941138.*

[15] *Yezzi, Anthony & Tsai, Andy & Willsky, Alan. (2002). "A Fully Global Approach to Image Segmentation via Coupled Curve Evolution Equations". Journal of Visual Communication and Image Representation. 13. 195-216. 10.1006/jvci.2001.0500.*

[16] *V. N. Vapnik and V. Vapnik, Statistical learning theory vol. 2: Wiley New York, 1998.*

[17] *Schölkopf, Bernhard & Smola, A.J.. (2001). Smola, A.: "Learning with Kernels - Support Vector Machines, Regularization, Optimization and Beyond". MIT Press, Cambridge, MA.*

[18] *Zheng, Sheng & Liu, Jian & Tian, Jin. (2004). "A new efficient SVM-based edge detection method. Pattern Recognition Letters". 25. 1143-1154. 10.1016/j.patrec.2004.03.009.*

[19] *Ginneken, Bram & Armato III, Samuel & Hoop, Bartjan & Vorst, Saskia & Duindam, Thomas & Niemeijer, Meindert & Murphy, Keelin & Schilham, Arnold & Retico, Alessandra & Fantacci, Maria & Camarlinghi, Niccolò & Bagagli, Francesco & Gori, I. & Hara, Takeshi & Fujita, Hiroshi & Gargano, Gianfranco & Bellotti, Roberto & Tangaro, Sabina & Bolaños, Lourdes & Prokop, Mathias. (2010). "Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: The ANODE09 study". Medical image analysis. 14. 707-22. 10.1016/j.media.2010.05.005.*

[20] *http://www.via.cornell.edu/databases/lungdb.html.*

[21] *Okada, Kazunori & Akdemir, Umut & Krishnan, Arun. (2005). "Blob Segmentation Using Joint Space-Intensity Likelihood Ratio Test: Application to 3D Tumor Segmentation". Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2. 437-444. 10.1109/CVPR.2005.92.*

[22] *Bellotti, Roberto & Carlo, F & Gargano, G & Tangaro, Sabina & Cascio, Donato & Catanzariti, E & Cerello, Piergiorgio & Cheran, Sorin & Delogu, Pasquale & De Mitri, Ivan & Fulcheri, Christian & Grosso, Daniele & Retico, Alessandra & Squarcia, S. & Tommasi, E & Golosio, Bruno. (2008). "A CAD system for nodule detection in low-dose lung CTs based on region growing and a new active contour model". Medical physics. 34. 4901-10. 10.1118/1.2804720.*

[23] *Cerello, Piergiorgio & Cheran, Sorin & Bagagli, Francesco & Bagnasco, Stefano & Bellotti, Roberto & Bolanos, Lourdes & Catanzariti, Ezio & De Nunzio, Giorgio & Fiorina, Elisa & Gargano, Gianfranco & Gemme, Gianluca & Torres, Ernesto & Masala, Gian & Peroni, Cristiana & Santoro, Matteo. (2008). "The Channeler Ant Model: Object segmentation with virtual ant colonies". 3147 - 3152. 10.1109/NSSMIC.2008.4775019.*

[24] *Retico, Alessandra & Bagagli, Francesco & Camarlinghi, Niccolò & Carpentieri, Carmela & Fantacci, Maria & Gori, I.. (2009). "A voxel-based neural approach (VBNA) to identify lung nodules in the ANODE09 study". Proceedings of SPIE - The International Society for Optical Engineering. 7260. 10.1117/12.811721.*

[25] *Murphy, Keelin & Ginneken, B & Schilham, Arnold & Hoop, B & Gietema, Hester & Prokop, Mathias. (2009). "A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification". Medical image analysis. 13. 757-70. 10.1016/j.media.2009.07.001.*

[26] *R.Wiemker, P.Rogalla, T.Blaffert, D.Sifri, O.Hay, E.Shah, R.Truyen, T.Fleiter, "Aspects of computer-aided detection(CAD) and volumetry of pulmonary nodules using multi slice CT". (2005). The British journal of radiology. 78 Spec No 1. S46-56. 10.1259/bjr/30281702.*

[27] *Dehmeshki, Jamshid & Ye, Xujiong & Lin, Xinyu & Valdivieso, Manlio & Amin, Hamdan. (2007). "Automated detection of lung nodules in CT images using shape-based genetic algorithm". Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society. 31. 408-17. 10.1016/j.compmedimag.2007.03.002.*

[28] *Farag, Amal & Graham, James & Elshazly, Salwa & Farag, Aly. (2010). "Statistical modeling of the lung nodules in low dose computed tomography scans of the chest". Proceedings / ICIP ... International Conference on Image Processing. 4281-4284. 10.1109/ICIP.2010.5651832.*