Early Detection of Cancerous Lung Nodules from Computed Tomography Images

Senthil Kumar T K *tkseneee@gmail.com*
Research Scholar,
Anna University,
Chennai, 600025, India

Ganesh E.N *enganesh50@gmail.com*
Principal,
Saveetha Engineering College,
Thandalam-602105, India

Umamaheswari R *umavijay.iitm@gmail.com*
Associate Professor*,*
Velammal Engineering College,
Chennai 600066, India

Abstract

This work is developed with an objective of identifying the malignant lung nodules automatically and early with less false positives. 'Nodule' is the 3mm to 30mm diameter size tissue clusters present inside the lung parenchyma region. Segmenting such a small nodules from consecutive CT scan slices are a challenging task. In our work Auto-seed clustering based segmentation technique is used to segment all the possible nodule candidates. Effective shape and texture features (2D and 3D) were computed to eliminate the false nodule candidates. The change in centroid position of nodule candidates from consecutive slices was used as a measure to eliminate the vessels. The two-stage classifier is used in this work to classify the malignant and benign nodules. First stage rule-based classifier producing 100 % sensitivity, but with high false positive of 12.5 per patient scan. The BPN based ANN classifier is used as the second-stage classifier which reduces a false positive to 2.26 per patient scan with a reasonable sensitivity of 88.8%. The Nodule Volume Growth (NVG) was computed in our work to quantitatively measure the nodules growth between the two scans of the same patient taken at different time interval. Finally, the nodule growth predictive measure was modeled through the features such as tissue deficit, tissue excess, isotropic factor and edge gradient. The overlap of these measures for larger, medium and minimum nodule growth cases are less. Therefore this developed growth prediction model can be used to assist the physicians while taking the decision on the cancerous nature of the lung nodules from an earlier CT scan.

Keywords: Computed Tomography, 3-D Image Segmentation, 3-D Image Features, Volume Growth, Lung Nodule Classifier.

1. INTRODUCTION

Worldwide lung cancer causes maximum cancer deaths than any other cancer[1]. Lung cancer diagnosis is also critical compared to other cancer diagnosis procedure[2]. The radiologist first analyzes the Computed Tomography(CT) scan on the patient and suggest for one another CT scan, if any symptom of lung cancer is spotted. Physicians will not recommend for biopsy unless they found the reasonable evidence from the CT scan. This procedure of wait and watch approach makes the lung cancer diagnosis time critical. Nodule is a 3mm to 30mm roughly round tissue clusters exists inside the lung parenchyma region. These nodules are the primary indicator of lung cancer. Visualizing these nodules in 3D and measuring the growth of it from earlier and a

later scan of the same patient makes the lung cancer diagnosis efficient. This work is concentrating on developing a mathematical model to predict the nodules which have the potential to grow in future.

The computer-aided approach for detecting lung nodule consist of three fundamental steps: 1.Lung nodule candidate segmentation [3,4,5], 2.Nodule feature extraction[6,7,8], and 3.Malignant nodule classification [9,10]. Later some work on lung nodule growth was reported in the literature. W.J.Kostis et al.[11], developed 3-D methods for the segmentation and characterization of nodules imaged using CT. A volumetric growth characterization was formulated in this work, and this was the first such a nodule growth based system reported for clinical use. Yuanjie Zheng et al.[12] applied a 2D graph-cut algorithm on the three-dimensional lung and tumor datasets. This method eliminates the possible volume variations of the nodule and estimates the tumor growth accurately. M Hasegawa et al.[13] computed a volume doubling time (VDT) based on the exponential model using follow-up CT images. All the above works concentrating on measuring the nodule growth accurately, but not much on predicting the nodule growth from earlier CT scan. In this work, we are proposing a structural model to predict the nodules growing nature from an earlier CT scan.

2. MATERIALS AND METHODS

2.1 Database

VOLCANO and LIDC are the two open source databases used for this work [14,15]. 34 patient cases from VOLCANO database and ten patient CT scan images from LIDC database are used in this work. For each patient in VOLCANO database, an earlier and follow-up scan CT images are available. The CT scans were taken using GE Medical and Philips CT scanners with different X-ray tube current and exposure. The time span between the scans is not specified for the VOLCANO database; however, information on nodule location is available. A total of 2425 cross sections of CT images from 44 cases were analyzed in this work. Totally 48 nodules are suspected to be malignant in this database by radiologists.

2.2 Auto Center Seed K-means Clustering Based Lung Nodule Segmentation

Segmenting a lung nodule from the lung parenchyma region is a complex task as the size, shape and the position of the nodules are differs much. Well-circumscribed, juxta-pleural, pleural tail and vascularized are the four types of nodules present in the lung parenchyma region. Segmenting all these nodules with the common algorithm is a challenging task. We already carried out some works on segmenting these nodules using threshold, auto-seed region-grow, and level set evolution methods[16,17,18,19]. In this work, we are using the concept of data clustering to segment those nodules from each slice of CT scan. Block histogram processing is carried out in this work to choose the initial cluster center seed automatically. In this research 3-D segmentation of nodule is carried out by consecutive 2-D segmentation of nodule structure from each CT scan image.

The proposed auto center seed k-means clustering based segmentation algorithm steps are:

Step 1: Compute the histogram (H) of lung CT image.

Step 2: Split the histogram into five blocks.

$$H = \{H1, H2, H3, H4, H5\} \quad (1)$$

Step 3: Compute the pixel value corresponding to the maximum peak of each histogram block.

$$Mi = maximum(Hi), \quad (2)$$

$$Pi = find_position(Hi == Mi) \quad (3), \text{ where i=1 to 5}$$

Step 4: Choose the mean of first two histogram blocks maximum peak pixel position as cluster 1 (C1) seed and the mean of the last two blocks maximum peak pixel position as cluster 2 (C2) seed.

$$C1 = \frac{P1+P2}{2} \quad (4) \qquad C2 = \frac{P1+P2}{2} \quad (5)$$

Step 5: Group the image pixels which have minimum distance with C1 and C2 separately, without changing its spatial positions.

$$d1 = \sqrt{(I - C1)^2} \quad (6)$$
$$d2 = \sqrt{(I - C2)^2} \quad (7)$$

$$K1\ (i, j) = I\ (i, j), \text{ if } d1 < d2 \quad (8)$$
$$K2\ (i, j) = I\ (i, j), \text{ if } d2 < d1 \quad (9)$$

Step 6: Find the new cluster seed $C1_{new}$ and $C2_{new}$ using the equation 10 and 11. Repeat from step 1 until the two consecutive cluster seeds (C1 and C2) are equal.

$$C1_{upd} = \frac{1}{L}\sum_{i=1}^{m}\sum_{j=1}^{n}(K1(i,j) \neq\neq 0) \quad (10)$$

$$C2_{upd} = \frac{1}{L}\sum_{i=1}^{m}\sum_{j=1}^{n}(K2(i,j) \neq\neq 0) \quad (11)$$

Step 7: Perform morphological closing/opening operation and border clearing morphology on a cluster 2 reference image to segment nodule candidates.

Step 8: Repeat the steps 1 to 7 for all the CT images slices.

Step 9: Stack the each nodule candidates segmented from each CT slice to form a 3-D visualization.

For a CT lung image with juxta-pleural nodule (nodule connected to the lung wall), the morphological closing process was performed. Since juxta-pleural nodules are located in the parenchyma and connected with lung wall, the hole (black region) was created in place of juxta nodule as shown in Figure 2(d). This juxta-hole region is usually connected with the background black region; hence it should be closed before filling the parenchyma region. Morphological closing process with disk structural element was performed on an image segmented from step 6. The left and right lungs were separated using connected component analysis to avoid connection between two lung lobes during the closing process. Finally, after closing the juxta nodule holes both the left and right lung masks were added as shown in Figure 1(e).

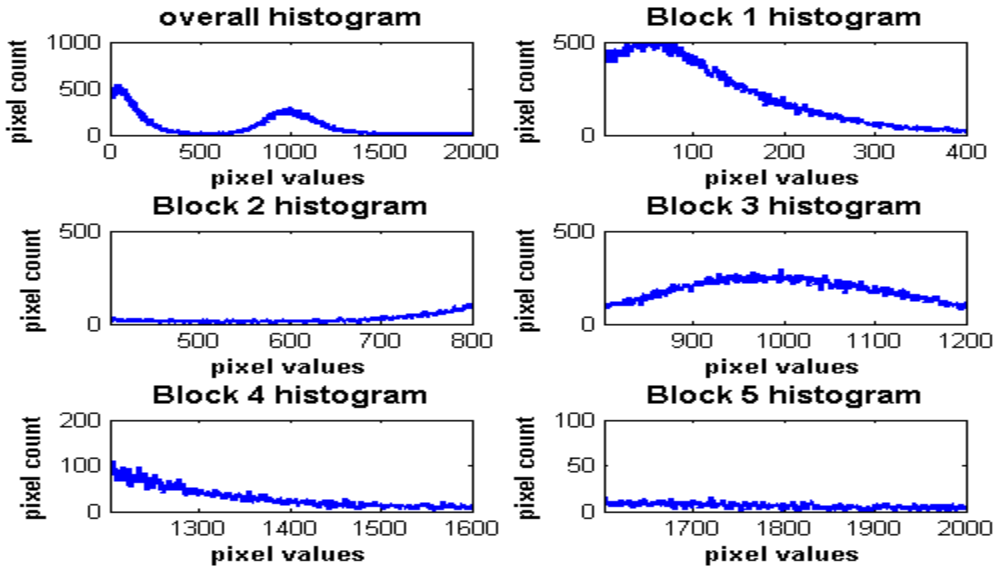


FIGURE 1: Block histogram to select initial seed.

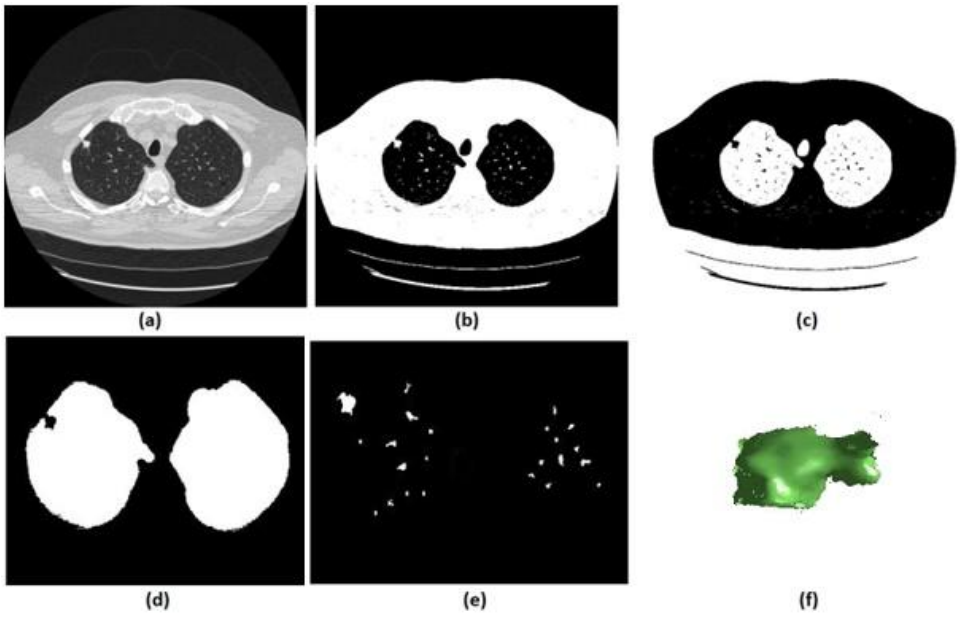


FIGURE 2: (a) Original CT 2-D slice, (b) cluster-2 seed mask, (c) cluster-1 seed mask, (d) Lung mask with juxta nodule hole, (e) Segmented nodule candidates, (f) 3-D visualization of one segmented nodule.

The proposed segmentation algorithm segmented 30 to 50 potential nodule candidates from each slices of CT image. Many non nodule patterns, calcifications and vessels are also segmented as nodules candidates. The three dimensional visualization of these nodule candidates will help to differentiate the nodules from the vessel like structures. Also 3-D visualization of the nodule candidates will help to identify the cancerous nature of it qualitatively. In the latter stage of this work, the qualitative information of nodule structure modeled quantitatively, which helps to find the cancerous nature of the nodule at early stage.

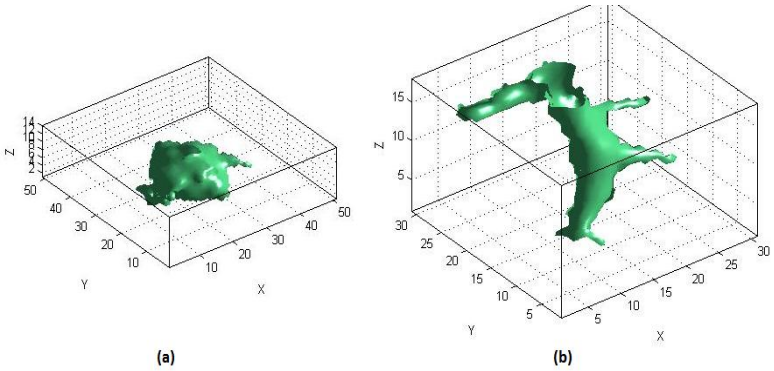


FIGURE 3: Three dimensional visualization of (a) nodule (b) vessel.

2.3 Nodule Feature Extraction

Most of the segmented nodule candidates are not malignant. To classify the malignant and benign nodules, different shape, and texture based features are extracted on the segmented nodule candidates [20, 21]. The quantitative features must give a discriminative value for malignant and benign nodules. In this work shape and texture features are computed in both 2D and 3D planes of nodule candidates. All the shape features are extracted from the binary version of nodules and the texture features are obtained from the original DICOM version of the nodule portion.

The probability of lung nodule being benign increases with the regularity in its shape across consecutive CT slices. The more regular the shape of the nodule, the more likely it is benign. In this work, the variation in shape features in the consecutive slices are quantitatively modeled through the features such as a change in centroid, change in area and change in equivalent diameter of the nodule. Edge sharpness is also one another measure which helps to discriminate malignant nodules from benign as the benign nodules have smooth edges. Frequency domain feature of DC constant also computed in this work, which is used to identify the calcifications. As the calcifications are uniform throughout its region, the DC constant value of it is on the higher side, and also the calcifications have a higher value of homogeneity and autocorrelation due to its uniform surface nature. The expressions for the features used in this work are tabulated in table1.

Area	$\sum_{x=1}^{M}\sum_{y=1}^{N} fb(x,y)$ Where fb = binary image of nodule candidate M= Number of rows in fb N= Number of columns in fb
Diameter	$\sqrt{\frac{4 * Area}{\pi}}$
Centroid	$\left[\frac{\sum_{i=1}^{M}\sum_{j=1}^{N} i * f(i,j)}{\sum_{i=1}^{M}\sum_{j=1}^{N} f(i,j)}, \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} j * fb(i,j)}{\sum_{i=1}^{M}\sum_{j=1}^{N} fb(i,j)}\right]$
Elongation	$\frac{\max(x_{length}, y_{length}, z_{length})}{\min(x_{length}, y_{length}, z_{length})}$

Volume	$\sum_{x=1}^{m}\sum_{y=1}^{n}\sum_{z=1}^{p} fb(x,y,z)$
Edge Sharpness	$\frac{\text{Magnitude(gradient(f(i,j))}}{length(x_{gradient})}$
Compactness	$\frac{\text{radius of sphere with equivalent radius}}{\text{root mean square distance from center}}$
Bounding box dimension rate	$\frac{\min([\text{xLength, yLength, zLength}])}{\max([\text{xLength, yLength, zLength}])}$
Compactness 2	$\frac{Surface\ Area^3}{Volume^2 * 36 * \pi}$
x-y plane projection compactness	$\frac{4 * \pi * A}{P^2}$
Homogeneity	$\sum_{i=1}^{m}\sum_{j=1}^{n} \frac{P[i,j]}{1+\lvert i-j\rvert}$
Correlation	$\sum_{i=1}^{m}\sum_{j=1}^{n} \frac{(i*j)*P[i,j]-(\mu x*\mu y)}{(\sigma x*\sigma y)}$
mean contrast	$\frac{Mean\ Inside - Mean\ outside}{Mean\ Inside + Mean\ outside}$
moment 6	$\frac{\sum_{x=1}^{m}\sum_{y=1}^{n} y^6 * f(x,y)}{\sum_{x=1}^{m}\sum_{y=1}^{n} f(x,y)}$
moment 7	$\frac{\sum_{x=1}^{m}\sum_{y=1}^{n} y^7 * f(x,y)}{\sum_{x=1}^{m}\sum_{y=1}^{n} f(x,y)}$
3D variance	$[\frac{1}{n}(\sum_{i=1}^{n}(xi-x_mean)^2 \sum_{y=1}^{n}(yi-y_mean)^2 \sum_{z=1}^{n}(zi-z_mean)^2)]^{\frac{1}{2}}$
3D Skewness	$\frac{E[(x-x_{mean}),\ (y-y_{mean})\ (z-z_{mean})]^3}{3D_variance^3}$

Fourier DC constant	fc(x,y) = (-1)(x+y) * f(x,y) F(u,v)= $\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} fc(x,y) * \exp[-j2\pi(\frac{mu}{M}+\frac{nv}{N})]$ DC constant =F(M/2, N/2)

TABLE 1: List of features extracted on a nodule.

2.4 Two-Stage Hybrid Classifier

In this section, a two-stage hybrid classifier is developed to classify the malignant and benign nodules using the selective features extracted in the previous section. The rule-based scheme followed by ANN based classifier is utilized in this chapter to classify the lung nodules.

The first stage rule-based classifier is framed with the following rules:

Nodule diameter	dia_min to dia_max (3mm to 30mm)
Nodule Area	$\pi*(\frac{dia_min}{2})^2$ to $\pi*(\frac{dia_max}{2})^2$
Nodule Volume	$3*\frac{\pi}{4}*(\frac{dia_min}{2})^3$ to $3*\frac{\pi}{4}*(\frac{dia_max}{2})^3$
Nodule Elongation	$\frac{\max(x_{length}, y_{length}, z_{length})}{\min(x_{length}, y_{length}, z_{length})}$ <4 mm
Nodule round degree	$\frac{4*\pi*area}{perimeter^2}$ >(1/6) mm
Nodule Centroid change	Centroid change in consecutive three slices < 10 pixels

TABLE 2: Rules of the first stage classifier.

After performing the first stage rule-based classifier, most of the vessel structures, a line like irrelevant structures are removed from nodule candidate list. The nodule elongation criteria as 4mm and round degree criteria as (1/6) mm are fixed after applying and analyzing these rules on a training image database considered for this work. All the vessels like structures and highly non-linear irrelevant tissue clusters are removed successfully from nodule candidates group, but still more non-nodule and benign tissue clusters exist in the output of the first stage of classifier.

All the shape and texture features are calculated on the nodule candidates remained after the first stage of classifier. As the radiologist report on training dataset is available, the correlation analysis is performed for each feature variable against the target (malignant or benign). The correlation coefficient between each feature and the target is computed. From the correlation matrix, the features which have high correlation coefficient are picked as input features to the ANN classifier. Finally, a hybrid input feature vector is formed with twelve features (diameter, edge sharpness, circularity, compactness, bounding box dimension rate, compactness2, x-y plane projection compactness, mean contrast, moment 6, moment 7, 3D variance and 3D skewness).

2.5 Lung Nodule Growth Measurement and Prediction

One major factor that the physicians are using in their qualitative analysis to make a call on malignant nodule is its growth. Generally, the doctors instruct the patient to undergo one another CT scan after a time span of 3 to 18 months if they suspect any symptoms in their earlier CT scan. Then by comparing the two CT scans, based on the growth of the suspected nodules, they may take a final call of its malignancy and instruct the patient to go for a biopsy. In this work, we quantitatively measure the nodules growth which helps the physicians to diagnose. The growth of the lung nodule must be measured in terms of its volume than area. Any small change in the 2D area of the nodule may result in a huge change in the 3D volume. The volumes of real final

nodules are computed from both the scan of each patient taken at different time intervals. The Nodule Volume Growth (NVG) is calculated using the equation below.

$$\mathrm{NVG} = \frac{\mathrm{V2}-\mathrm{V1}}{\mathrm{V1}} \quad (12)$$

where V1 and V2 are the volumes of nodules segmented from earlier and follow-up scan respectively.

A total of 34 real nodules were segmented from VOLCANO and NVG was computed which is given in Table 3.Three cases showed an increase in volume of more than double. The corresponding NVG value for case no 1, 3 and 44 was 14.6, 1.46, and 0.96 which was found to grow approximately by 15, 1.5 and 1 times, respectively, compared with the initial scan. This noticeable growth confirms that these nodules are malignant in nature. There were 14 cases for which the RGN ranged between 0.2 and 0.8, suggesting a reasonable growth that requires repetition of a scan after few days to confirm malignancy. Of the remaining 17 cases, 7 had NVG of less than 0.2 , i.e., almost no change and 10 cases showed minimum nodule growth which may require follow-up scans.

The nodules from VOLCANO database are analyzed further for its growth measurement and prediction, as this database have two CT scan series for each patient which taken at different time intervals. In figure 4, the nodules segmented from earlier and later scan are depicted for three patient cases. In the first row of figure 4(a) and (b), the nodule segmented from an earlier scan is shown. These nodules surface is not uniform, and few tissue parts are coming out of the nodule surface in a radial direction, whereas in figure(c), the tissue components which coming out of nodule surface is minimum and the surface looks uniform. From the follow-up scans of all these patient cases, it has been proven that first two cases depicted in 4(a) and 4(b) are grown in future, on the other hand, the nodule showed in 4(c) not grown. These characteristics of nodules segmented from earlier scans are used as a quantitative measure to predict the nodule growth. In this work, the nodule growth is predicted using four measures: tissue excess (TE), tissue deficit (TE), isotropic factor(IF) and edge gradient (EG). The template and expression of these measures are tabulated in Table 4 and these values are computed and tabulated in table 5.

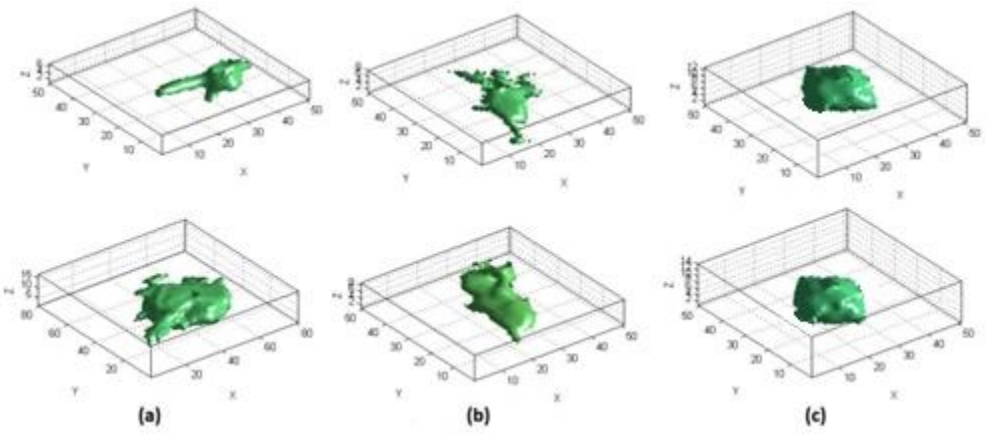


FIGURE4: Nodules from earlier and later scan (a) with large growth (b) with medium growth (c) No growth.

Case No	Database Case ID	Volume Scan1(pixels)	Volume Scan2(pixels)	Volume Diff	VNG
1	1	630	9840	9210	14.619
2	2	2108	2576	468	0.222
3	6	95	234	139	1.4632
4	7	149	181	32	0.2148
5	8	90	91	1	0.0111
6	10	1490	1508	18	0.0121
7	11	5640	5970	330	0.0585
8	13	8197	8746	549	0.067
9	15	447	646	199	0.4452
10	16	284	318	34	0.1197
11	18	3709	3839	130	0.035
12	20	1340	1353	13	0.0097
13	21	2239	2233	6	0.0027
14	23	2807	3031	224	0.0798
15	24	930	901	29	0.0322
16	25	1485	1961	476	0.3205
17	28	956	1340	384	0.4017
18	29	64	66	2	0.0313
19	32	112	124	12	0.1071
20	33	643	649	6	0.0093
21	34	106	160	54	0.5094
22	36	1028	1098	70	0.0681
23	37	1818	2430	612	0.3366
24	38	1742	1835	93	0.0534
25	39	1920	2533	613	0.3193
26	40	773	954	181	0.2342
27	41	1236	1566	330	0.267
28	42	611	850	239	0.3912
29	44	690	1356	666	0.9652
30	45	866	1223	357	0.4122
31	46	136	149	13	0.0956
32	47	1688	2111	423	0.2506
33	48	653	381	272	0.7139
34	49	4280	4950	670	0.1565

TABLE 3: Lung nodule growth measurement.

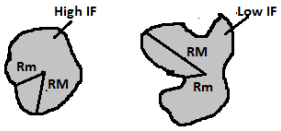
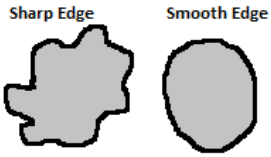
Tissue Deficit (TD)	Tissue Excess(TE)	Isotropic Factor(IF)	Edge Gradient (EG)
$TD = \frac{Sc - A}{Sc}$	$TE = \frac{A - Si}{A}$	 $IF = \frac{Rm}{RM}$	 $\frac{\text{Magnitude(gradient(f(i,j)}}{length(x_{gradient})}$

TABLE 4: Nodule growth prediction model.

		Earlier CT Scan			
Case No	Database Case ID	Tissue Deficit (TD)	Tissue Excess (TE)	Isotropic Factor (IF)	Edge Gradient (EG)
1	1	0.308	0.668	0.400	0.820
2	2	0.271	0.395	0.600	0.534
3	6	0.532	0.732	0.324	0.610
4	7	0.201	0.278	0.740	0.634
5	8	0.152	0.143	0.895	0.342
6	10	0.093	0.072	0.920	0.123
7	11	0.093	0.173	0.912	0.321
8	13	0.189	0.188	0.820	0.432
9	15	0.198	0.306	0.718	0.540
10	16	0.215	0.301	0.720	0.632
11	18	0.097	0.132	0.944	0.240
12	20	0.117	0.211	0.775	0.234
13	21	0.176	0.128	0.759	0.211
14	23	0.183	0.192	0.812	0.500
15	24	0.102	0.152	0.783	0.267
16	25	0.283	0.452	0.740	0.654
17	28	0.195	0.214	0.800	0.701
18	29	0.125	0.141	0.856	0.345
19	32	0.203	0.200	0.670	0.458
20	33	0.042	0.135	0.750	0.345
21	34	0.345	0.655	0.654	0.610
22	36	0.172	0.182	0.764	0.590
23	37	0.214	0.534	0.670	0.467
24	38	0.048	0.115	0.785	0.340
25	39	0.276	0.521	0.723	0.587
26	40	0.243	0.319	0.562	0.454
27	41	0.341	0.627	0.540	0.620
28	42	0.321	0.678	0.730	0.560
29	44	0.478	0.878	0.263	0.712
30	45	0.234	0.403	0.884	0.532
31	46	0.194	0.196	0.830	0.432
32	47	0.301	0.347	0.630	0.534
33	48	0.107	0.509	0.556	0.640
34	49	0.341	0.432	0.693	0.576

TABLE 5: Nodule growth prediction measure values.

3 RESULTS AND DISCUSSION

In this work 24 patient lung CT scan series from VOLCANO database and five patient cases from LIDC database are used as a training image set. VOLCANO cases contain 25 to CT image slices per scan as an average, and LIDC scans have 120 to 280 CT slices per scan. In this work, a total of 1450 CT cross-sectional images (slices) are used for training. The auto center seed k-means clustering based segmentation was effectively applied on all these 1450 CT slice images to segment all the possible nodule candidates.

The nodule candidates segmented from these 1450 CT slices are applied to the rule-based classifier (first stage classifier). The nodule candidates, which are not satisfying the rules mentioned in Table 2 are eliminated. The training database set used in this work contains 32 malignant nodules. The first stage of classifier output preserving all the 32 nodules, which marked in radiologist report. Along with these 32 nodules, 362 other nodules also remained in the output of the rule-based classifier. Hence, the True Positive (TP) (the nodules detected which matched with radiologist report) at the output of classifier first stage is 32, and the False Positive (FP) (the nodules detected, which not matched with radiologist report) is 362. As all the nodules

marked by the radiologists report are detected, the False Negative (FN) is zero. The performance of the first stage rule-based classifier is shown in Table 6.

Statistic	Formula	Value
Sensitivity	$\frac{TP}{TP + FN}$	100%
False Positive	FP/scan	12.48

TABLE 6: Performance measure of the first stage of classifier.

The first stage rule based classifier producing a poor (high) false positive of 12.5 per patient case. Artificial Neural Network based second stage of classifier is used to reduce the false positive. The input feature matrix size for an ANN is 12 x 394, where 12 is the total number of finalized features and 394(362+32) is the number of nodule candidates remained at the output of the rule-based classifier. The target matrix of size 2 x 394 is framed with the help of radiologist findings. The target output of [0 1] and [1 0] is fixed for benign and malignant nodules respectively. The BPN based ANN classifier is trained towards the minimum error. The trained net input and output weights are saved for testing the new cases.

10 cases of VOLCANO and 5 cases of LIDC are tested using the parameters of trained classifier. This testing set contains 18 malignant nodules. Total of 975 cross-sectional CT images are present in the testing database. After the segmentation process, all the suspected nodule candidates are applied under the rule based classifier. Total of 204 nodule candidates remained at the output of the rule-based classifier. The 12 features discussed previously are computed for each nodule and the feature vector of size 12x204 is applied to a trained ANN system. The benign nodules feature vector are converged very near to [0 1], and the malignant nodules are to [1 0]. The output of our designed lung nodule detection system is compared with the radiologist report. Algorithm developed in this work correctly detected 16 malignant nodules out of 18 and missed to detect 2, with a TP of 16 and FN of 2. This algorithm successfully eliminated 170 nodule candidates and wrongly identified 34 nodule candidates as a malignant nodule. Therefore, the TN is 170; FP is 34; and the FP per patient scan is 2.26. The overall performance measure of the developed algorithm is given in Table 7.

Statistic	Formula	Value
Sensitivity	$\frac{TP}{TP + FN}$	88.8%
Specificity	$\frac{TN}{TN + FP}$	83.3%
FP/patient scan	Total FP / Number of patient scan	2.26
Disease prevalence	$\frac{TP + FN}{TP + FN + TN + FP}$	8.1%

TABLE 7: Statistical Results.

The nodules from VOLCANO database are analyzed further for its growth measurement and prediction, as this database have two CT scan series for each patient which taken at different time intervals. The nodule growth measures are tabulated in table 5. The nodules which had larger growth (in follow-up scan) having TD > 0.3, TE > 0.6, IF < 0.4 and EG>0.6 in its earlier scan. The nodules which had average growth having TD in between 0.2 to 0.4 , TE in between 0.2 to 0.7, IF in between 0.5 to 0.75, and EG in between 0.45 to 0.7. Finally, the nodules which had very minimum or no growth have the TD <0.2, TE < 0.2, IF >0.65, and EG <0.6. The overlap

of these measures for larger, medium and minimum nodule growth cases are less. A comparison of our method with other standard works are tabulated in table 8.

Authors	Sensitivity	False Positive	Nodule Diameter
Dehmeshki et al.(2007) [5]	90	15.5	3-20mm
Opfer and Wiemeker (2007) [22]	74	4	>4mm
Ozekes et al (2008) [23]	100	13.4	>5mm
Golosio et al (2009) [24]	79	4	>4mm
Choi WJ et al(2012) [25]	95	2.3	>1.5mm
Alilou et al (2014) [26]	80	3.9	>4mm
Lu L et al (2015) [27]	85	3.13	>5mm
Our proposed work	**88.8**	**2.26**	**>3mm**

TABLE 8: Performance comparison.

4 CONCLUSION

An efficient and automatic lung nodule detection system was successfully implemented in this work. Auto center seed clustering based segmentation was successfully implemented in this work. Block histogram processing technique was used in this study to initialize the cluster center seed value. After analyzing many shape and texture features, six features were used to form the rules for the first stage of classifier. Other twelve features were computed on the nodules remained at the output of the first stage of classifier. The BPN-ANN network was used as a second stage of classifier to classify the malignant and benign nodules. This algorithm proposed in this paper produced a false positive of 2.26 per patient scan with a reasonable sensitivity of 88.8%. The nodules growing nature was modeled in this work using four shape measures computed from earlier CT scan, which can be used to assist the physicians while taking the decision on cancerous nature of lung nodules at an early stage. In future, this work can be extended by modeling the nodule growth prediction algorithm, more accurately by training and testing with more patient cases.

5 REFERENCES

1. Torre, L.A., Bray, F., Siegel, R.L., Ferlay, J., Lortet Tieulent, J. and Jemal, A., 2015. Global cancer statistics, 2012. CA: CA Cancer J Clin, 65(2), pp.87-108.

2. Chheang, S. and Brown, K., 2013, June. Lung cancer staging: clinical and radiologic perspectives. Semin Intervent Rad, Thieme Medical Publishers, 30(2),pp. 099-113.

3. Armato, S.G. and Sensakovic, W.F., 2004. Automated lung segmentation for thoracic CT: impact on computer-aided diagnosis1. Acad Radiol, 11(9), pp.1011-1021.

4. Leader, J.K., Zheng, B., Rogers, R.M., Sciurba, F.C., Perez, A., Chapman, B.E., Patel, S., Fuhrman, C.R. and Gur, D., 2003. Automated lung segmentation in X-ray computed tomography: development and evaluation of a heuristic threshold-based scheme1. Acad Radiol, 10(11), pp.1224-1236.

5. Dehmeshki, J., Amin, H., Valdivieso, M. and Ye, X., 2008. Segmentation of pulmonary nodules in thoracic CT scans: a region growing approach. IEEE T Med Imaging, 27(4), pp.467-480

6. Mehtre, B.M., Kankanhalli, M.S. and Lee, W.F., 1997. Shape measures for content based image retrieval: a comparison. Inform Process Manag, 33(3), pp.319-337

7. Han, F., Wang, H., Zhang, G., Han, H., Song, B., Li, L., Moore, W., Lu, H., Zhao, H. and Liang, Z., 2015. Texture feature analysis for computer-aided diagnosis on pulmonary nodules. J Digit Imaging, 28(1), pp.99-115.

8. Choi, W.J. and Choi, T.S., 2014. Automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor. Comput Meth Prog Bio, 113(1), pp.37-54.

9. Shen, W., Zhou, M., Yang, F., Yang, C. and Tian, J., 2015, June. Multi-scale convolutional neural networks for lung nodule classification. In International Conference on Information Processing in Medical Imaging (pp. 588-599). Springer International Publishing.

10. Chan, H.P., Sahiner, B., Hadjiyski, L., Zhou, C. and Petrick, N., The Regents Of The University Of Michigan Technology Management Office, 2003. Lung nodule detection and classification. U.S. Patent Application 10/504,197.

11. Kostis, W.J., Reeves, A.P., Yankelevitz, D.F. and Henschke, C.I., 2003. Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images. IEEE T Med Imaging, 22(10), pp.1259-1274.

12. Zheng, Y., Steiner, K., Bauer, T., Yu, J., Shen, D. and Kambhamettu, C., 2007, October. Lung nodule growth analysis from 3D CT data with a coupled segmentation and registration framework. In 2007 IEEE 11th International Conference on Computer Vision (pp. 1-8). IEEE.

13. Hasegawa, M., Sone, S., Takashima, S., Li, F., Yang, Z.G., Maruyama, Y. and Watanabe, T., 2000. Growth rate of small lung cancers detected on mass CT screening. Brit J Radiol, 73(876), pp.1252-1259.

14. Armato III, S.G., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Zhao, B., Aberle, D.R., Henschke, C.I., Hoffman, E.A. and Kazerooni, E.A., 2011. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Med Phys, 38(2), pp.915-931.

15. Reeves, A.P., Jirapatnakul, A.C., Biancardi, A.M., Apanasovich, T.V., Schaefer, C., Bowden, J.J., Kietzmann, M., Korn, R., Dillmann, M., Li, Q. and Wang, J., 2009, September. The VOLCANO'09 challenge: Preliminary results. In Second international workshop of pulmonary image analysis (pp. 353-364).

16. Kumar, T.S. and Ganesh, E.N., 2013. Proposed technique for accurate detection/segmentation of lung nodules using spline wavelet techniques. Int. J. Biomed. Sci, 9(1), pp.9-17.

17. Krishnamurthy, S., Narasimhan, G. and Rengasamy, U., 2016. Three-dimensional lung nodule segmentation and shape variance analysis to detect lung cancer with reduced false positives. P I Mech Eng H, 230(1), pp.58-70.

18. SenthilKumar Krishnamurthy, Ganesh Narasimhan, Umamaheswari Rengasamy .(in press) "Lung nodule growth measurement and prediction using auto cluster seed k-means morphological segmentation and shape variance analysis", Int J Biomed Eng Technol.

19. Reed, T.R. and Dubuf, J.H., 1993. A review of recent texture segmentation and feature extraction techniques. CVGIP: Image understanding, 57(3), pp.359-372.

20. Zhang, D. and Lu, G., 2004. Review of shape representation and description techniques. Pattern recognition, 37(1), pp.1-19.

21. Mukundan, R. and Ramakrishnan, K.R., Moment functions in image analysis: theory and applications (Vol. 100). Singapore: World Scientific, 1998.

22. Opfer R, & Wiemker R. Performance analysis for computer-aided lung nodule detection on LIDC data. Medical Imaging, International Society for Optics and Photonics. 2007, (pp. 65151C-65151C).

23. Ozekes, S., Osman, O. and Ucan, O.N., 2008. Nodule detection in a lung region that's segmented with using genetic cellular neural networks and 3D template matching with fuzzy rule based thresholding. Korean journal of radiology, 9(1), pp.1-9.

24. Golosio, Bruno, et al. "A novel multithreshold method for nodule detection in lung CT." Med Phys 36.8 ,2009: 3607-3618.

25. Choi W. J, & Choi T. S. Automated pulmonary nodule detection system in computed tomography images: A hierarchical block classification approach. Entropy, 2013: 15(2), 507-523

26. Alilou M, Kovalev, V, Snezhko E, & Taimouri V. A comprehensive framework for automatic detection of pulmonary nodules in lung ct images. Image Analysis Stereology, 2014: 33(1),13 27.

27. Lu, L., Tan, Y., Schwartz, L.H. and Zhao, B., 2015. Hybrid detection of lung nodules on CT scan images. *Medical physics*, *42*(9), pp.5042-5054.