EUROMEDITERRANEAN BIOMEDICAL JOURNAL

for young doctors

Original Article

AUTOMATIC LUNG NODULE SEGMENTATION USING AUTOSEED REGION GROWING WITH MORPHOLOGICAL MASKING (ARGMM) AND FEATURE EXTRACTION THROUGH COMPLETE LOCAL BINARY PATTERN AND MICROSCOPIC INFORMATION PATTERN

T.K.Senthil Kumar [1], E.N.Ganesh [2], R. Umamaheswari [3]

Summary

An efficient Autoseed Region Growing with Morphological Masking(ARGMM) is implemented in this paper on the Lung CT Slice to segment the 'Lung Nodules',which may be the potential indicator for the Lung Cancer. The segmentation of lung nodules carried out in this paper through Multi-Thresholding, ARGMM and Level Set Evolution. ARGMM takes twice the time compared to Level Set, but still the number of suspected segmented nodules are doubled, which make sure that no potential cancerous nodules go unnoticed at the earlier stages of diagnosis. It is very important not to panic the patient by finding the presence of nodules from Lung CT scan. Only 40 percent of nodules can be cancerous. Hence, in this paper an efficient Shape and Texture analysis is computed to quantitatively describe the segmented lung nodules. The Frequency spectrum of the lung nodules is developed and its frequency domain features are computed. The Complete Local binary pattern of lung nodules is computed in this paper by constructing the combine histogram of Sign and Magnitude Local Binary Patterns. Local Configuration Pattern is also determined in this work for lung nodules to numerically model the microscopic information of nodules pattern.

Introduction

Diagnosing the Lung cancer from CT images using Image Processing algorithms is a vital research area. Lung Cancer is diagnosed generally by analyzing a tissue cluster formation called 'Nodule' inside the lung. The size of the nodules varies from few millimeters to 5 centimeters. Based on the shape, the nodules are classified as Well-Circumscribed, Pleural tail, Vascularised and Juxta Pleural. Well Circumscribed nodules are placed well inside the Lung parenchyma region, Pleural Tail nodules are connected to the pleural surface by a thin tail like structure, Vascularised nodules are attached with the vessels present inside the lung region, and the Juxta Pleural nodules are con-

Address of the authors:

1. Department of ICE, Anna University, Chennai, Tamilnadu, India
2. Rajalakshmi Institute of Technology, Tamilnadu, India
3. Department of Electrical and Electronics, Velammal Engineering College, Chennai.

Send correspondence to: *T.K.Senthil Kumar , tkseneee@gmail.com*

Received: *17th April, 2015 — **Revised:** 13h May, 2015 — **Accepted:** 21th May, 2015*

nected significantly to the wall surface inside the lung.The Kostis classification framework phantom model of nodule [1] is shown in the figure 1. As the shape and size of the Nodules are varying, the process of segmentation is really a challenging task. The nodules segmented initially based on its shape, size, texture or position, are only suspected nodules (tissue clusters). Only few of this suspected nodules lead to cancer, therefore different mathematical features both shape and texture should be calculated for the suspected nodules, based on those values the cancerous nature of the nodules can be analyzed. Developing an algorithm, which gives the discriminative feature value for cancerous and non-cancerous nodules, is really a challenging research task.

Objective of the work

There are two main objectives for this work

1. Segment allthe nodule like structures from the CT scan. This work concentrates especially on segmenting even a small few millimeter size nodule shape structure, which help diagnosing the cancer at very early stage.
2.Extract shape and Texture based feature for segmented nodules, which is fed as an input classifier to classify the cancerous and non-cancerous nodules. Local Binary Pattern (LBP) based texture features are used specifically in this work along with other basic texture features.

The importance of segmenting the small nodules is to find the cancerous nature of the nodule at an early stage. Presently, physicians diagnose lung cancer by comparing the Lung nodules from CT scans which taken within an interval of 6 months or 1 year. Based on the growth rate of nodule like structures in CT scan, physicians instruct the patient to go for biopsy[2][3]. The drawback to this procedure is that the patient has to wait 6-12 months before making the decision for doing a biopsy. This work is concentrating on making effective decisions in the early stage itself, by analyzing the smaller nodules from initial CT scans [4] [5].

The novel Automatic Region Growing with Morphological Masking (ARGMM) algorithm, which is completely automatic, is developed in this paper. Also, Multi Threshold and Active contour based segmentations are implemented in this paper, and the results of all three segmentations techniques are compared. Effective and reliable Shape based and Texture based features are calculated in this paper on the segmented nodules, through which the cancerous and non-cancerous nodules are classified efficiently. Specifically, this work concentratesas well on Modified and Efficient Local Binary Pattern (LBP), based texture features namely Complete Local Binary Pattern (CLBP) and Microscopic Information Pattern(MIP).

Database used

LIDC (Lung Image Database Consortium) is the public open access online database [6], which is used in this work. The segmented algorithms developed in this paper are tested for about 30 different patient images and the results are satisfactory.

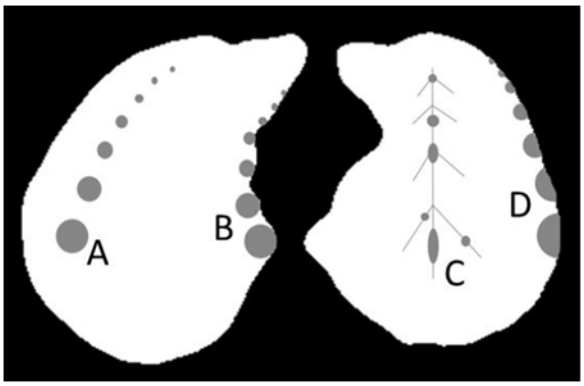


Figure 1: Kostis' Phantom model of Nodule, A - Well-Circumscribed, B - Pleural tail, C - Vascularised, D - Juxta Pleural

Nodule segmentation

The CT scan of lung cross section contain Heart muscle and part of chest wall, along with the main lung lobes as shown in the figure 2. The first objective of the segmentation process is to segment the lung lobes alone. Once the lung lobes are segmented, then the vessels inside the parenchyma region need to be removed. Lastly, the "nodule" candidates can be segmented from within the lung lobe [7]. Several segmentation methodologies were discussed in different literatures[8] [9][10].

1. Multi Threshold Based Segmentation

Conventional Lung Nodule segmentation consist in 4 steps

1. Extraction of Heart Muscle and Large Airway
2. Segmentation of Lung Region
3. Separation of Right and Left lungs
4.Segmentation of Nodule

The pixel intensity of CT scan is measured in Hounsfield Units(HU). The HU value of Lung lobe tissue is in between -910 to -500 [11]. In this work the DICOM format CT image has been converted in to JPEG format,which occupies less space. The JPEG conversion will not affect the segmentation result, as the pixel values in the JPEG format is discriminative enough between different main regions of Lung CT. After initial thresholds and region grow, the binary morphology operations are performed to segment the nodules, which doesn't depend on gray values. The Pixel intensity range of Lung tissue in JPEG format is in between 80 to 100. The CT Lung slice and its Histogram are shown in figure 3. In figure 3 the majority of pixels are distributed in three

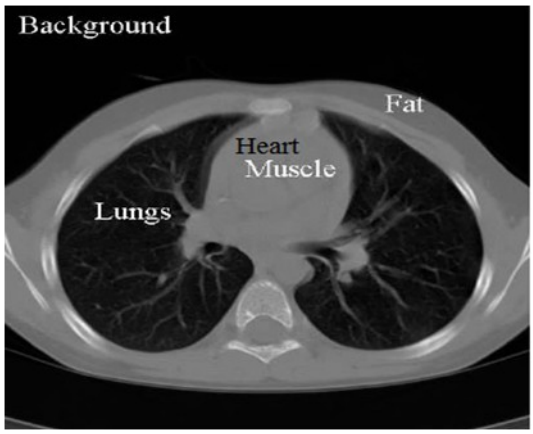


Figure 2: Lung CT Slice

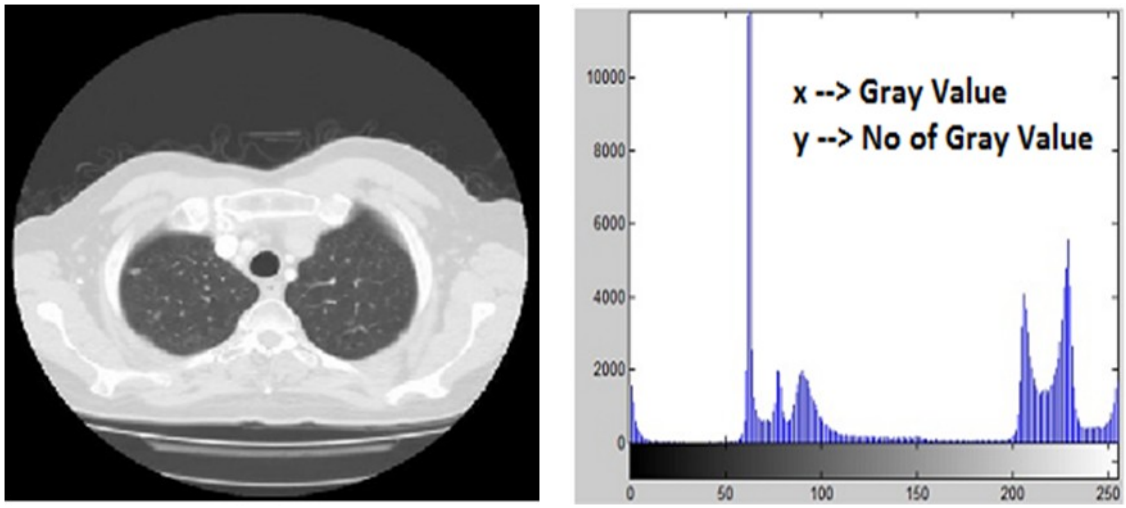


Figure 3: CT lung slice and its Histogram

clusters (0 to 15), (55 to 120) and (200 to 255). Picking the threshold for lung lobe region is a bit difficult because background pixels overlap with the lung region. The pixel values of the heart muscle and the wall region of the lung are between (200 to 255), which is discriminative completely from the lung lobe region as shown in figure 3. The threshold of 80 is chosen to remove the background, and only the lung region is segmented as shown in figure 4. The histogram of the lung slice without the background is shown in figure 4. The threshold of 120 was chosen once again to remove the muscle and fat portion of the lung slice, so that only the lung lobes can be segmented from the CT slice image. The segmented lung lobes and its histogram are shown in figure 5. Irrespective of choosing fine threshold values, there are unwanted clusters of pixels that get segmented along with lung lobe region. Connected pixel concepts need to be applied to remove the pixel groups in which lesser number of pixels are connected. Morphology operations can be carried out on binary images easily, rather than gray scale image. Hence, the segmented lung lobe image converted in to binary image and then the morphology operations are performed on the binary image to remove the unwanted small clusters. The clusters in which the number of pixels is less than 2200 are removed, so that all the unwanted clusters are removed and

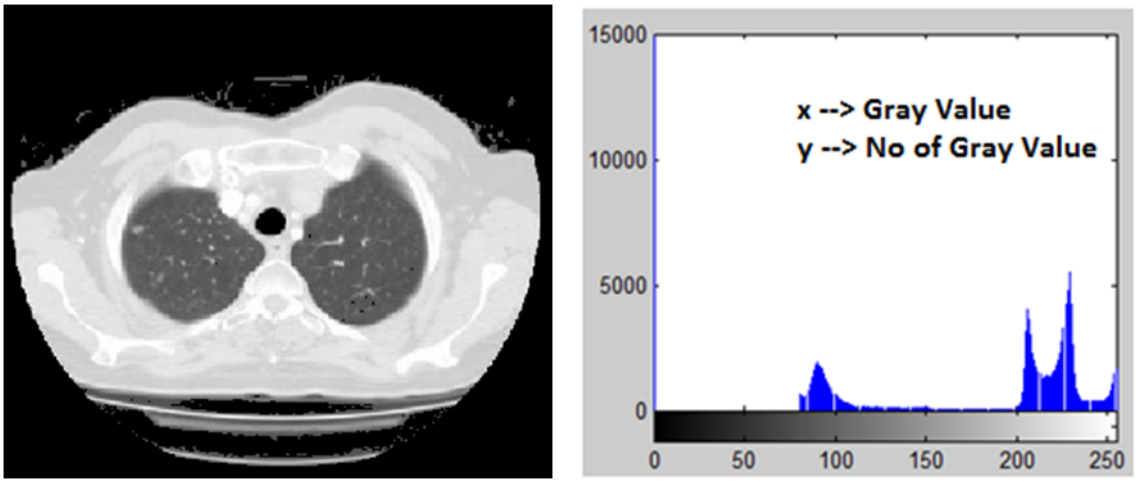


Figure 4: CT lung slice without background and its Histogram

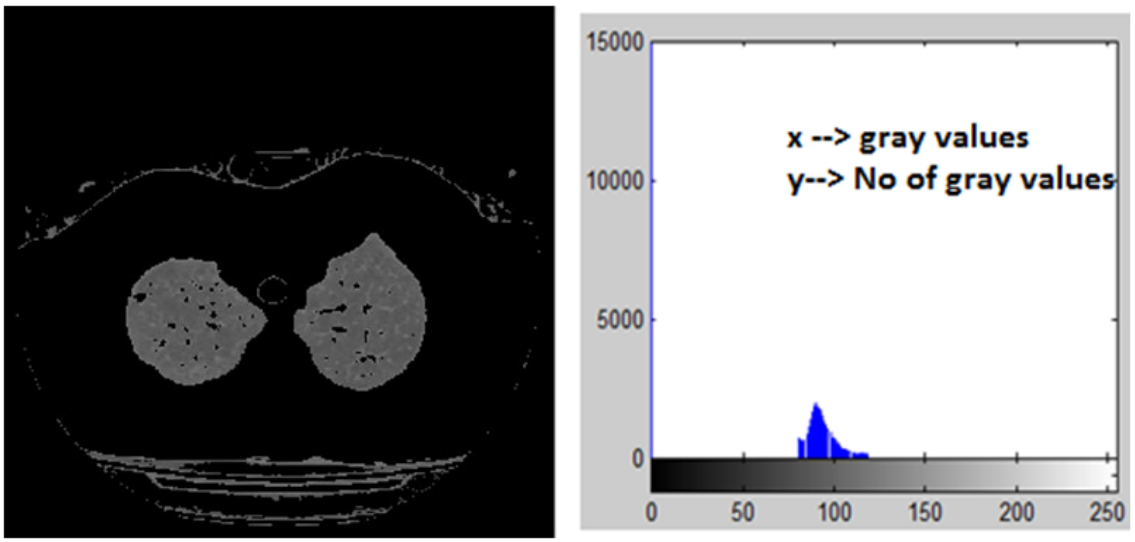


Figure 5: Segmented lung lobes and its Histogram

only the main lung lobe is segmented. The binary image before and after the morphological operation are shown in figure 6. The black regions inside the lung lobes in figure 6a are the suspected nodules. The binary lobe image is inverted to render the nodule region white, which is our region of interest as shown in figure 7. The white regions in figure 7 are our regions of interest, which could be nodules. Morphological operations were performed again on the lung mask to clear the border as shown in figure 8. Each white cluster in figure 8 is saved as separate image, and different shape and texture features can be determined for classification of its cancerous nature.

Choosing the threshold value to remove the background is the main drawback of this method. The background pixels and the main lung lobe tissue pixels overlap, and hence, we need to perform morphological opening and closing operations to remove the small clusters. Again, for this we need to set some user defined values to decide up to what size clusters need to be removed. Hence, the Thresholding method is not fully automatic, and also this method is not universal, as the pixel intensity values vary from one CT machine to another.

2. Autoseed Region Growing With Morphological Masking Segmentation

The drawbacks previously discussed re-

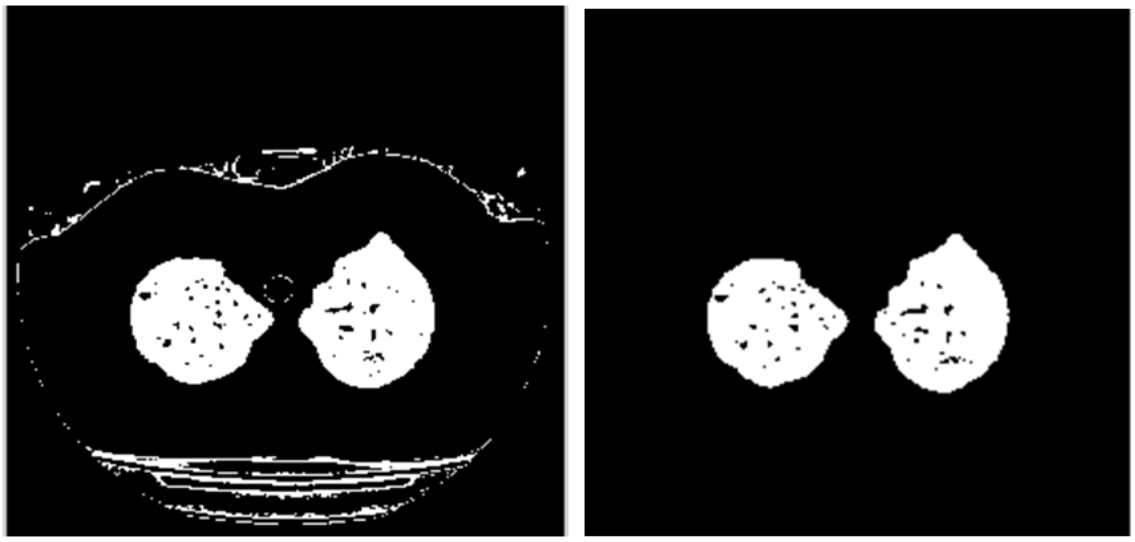


Figure 6 a (left) and 6 b (right) : Before and After Morphological operation

Figure 7: Lung Nodule Mask

Figure 8: Suspected Nodules from Multi Threshold method

garding the Threshold based method, can be overcome using this Autoseed Region Growing with Morphological Masking (ARGMM).

The steps for ARGMM are the following:

Step 1: The mask created by choosing a seed point from the outer background region of the Lung CT scan. This seed point selection can be universal for all the Lung CT images taken from any CT machine. The region growing [12] output with the seed point are chosen from the outer region of CT lung scan is shown in figure 9.

Step 2: At all the positions where there are '0' pixels, the mask is replaced by the original CT image pixels, and the positions where there are '1' pixels, the mask is replaced by '0'. Hence, the lung region can be segmented as shown in figure 10.

Step 3:

After step 2, it is very easy to automatically fix the threshold using Ostu's method[13]. The output image after applying Ostu's algorithm is shown in figure 11.

Step 4: Create a nodule mask by inverting the image in figure 11. Clear the border of the inverted image by applying Morphological operations to the segment with the suspected nodules as shown in figure 12.

The ARGMM segmentation method is completely automatic and can be universally adapted for the Lung CT images taken from any machines, as the seed point chosen from the outer region and

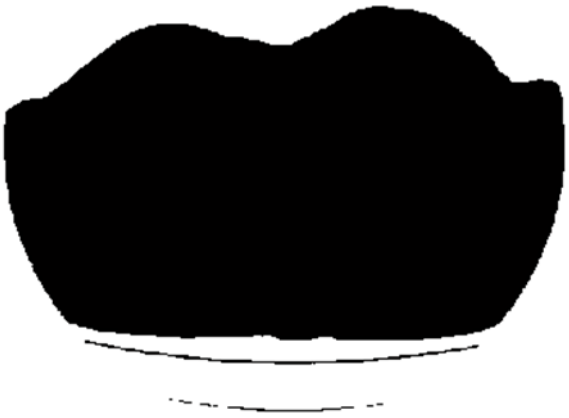


Figure 9: Autoseed Region Growing

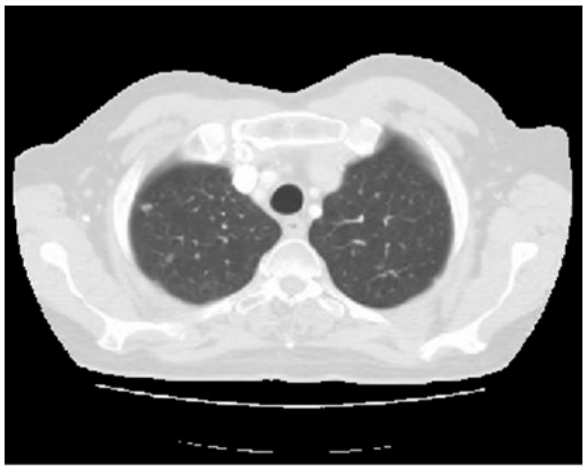


Figure 10: Segmented Lung region using Mask

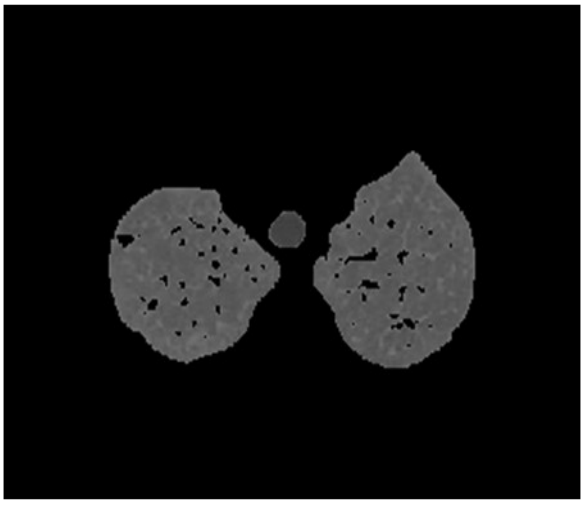


Figure 11: Segmented Lung Region

Figure 12: Suspected Nodules from ARGMM

also Ostu's automatic threshold method is used in step 3. In step 4, the ARGMM segmentation method provided the better-segmented result on the images taken from different CT scans with different X-ray doses.

3. Level Set Evolution Based Lung Nodule Segmentation

The Level Set technique is a powerful tool to segment the region of interest from an Image. Initially, Active Contours were used by Kass [14], in which the curves explicitly move to extract the region of interest in an image. In the active contour model, the selected points on the initial curve (at t=0) are moved with the velocity in the normal direction. The curve deformation in this active contour mode depends on how many points we choose. More points lead to instability, and fewer points lead to an incorrect and deformable shape in the curve.

To overcome this drawback, the Level Set proposed by Osher and Sethian [15] implicitly represents the curve by the zero level of high dimensional function.

The implementation of the Level Set algorithm has three fundamental steps:

1.Initialization of Curve function φ(x,y)
2.Evolution of curve
3.Stopping condition for curve evolution

The curve can be mathematically defined as:

$$\mathbf{C(p)= \{(x,y) \mid \varphi(x,y) =0\}} \quad \text{----- (1)}$$

The point p (x,y) canbe included inthe curve if the function φ(x,y) = 0. The function φ(x,y) >0 inside the curve and φ(x,y) <0 outside the curve.

Generally, the initial curve function can be just a rectangle, which needs to evolve consecutively to deform in such a way to fit in to the region of interest. To evolve the curve, it is essential to compute the velocity and curvature.

The deformation/evolution of curve will move in the normal direction, therefore, the velocity should be defined in the normal direction. The curve also has velocity in tangential direction, but the velocity in the tangential direction doesn't change the shape of the curve. Normal and tangent component of φ is defined as:

$$\vec{N} = -\frac{\nabla\varphi}{|\nabla\varphi|} \quad \text{----- (2)}$$

$$\vec{T} = \frac{\vec{\nabla}\varphi}{|\nabla\varphi|} \quad \text{----- (3)}$$

$\nabla\varphi$ is gradient of the curve that gives the maximum rate of change. The gradient along the curve, in the tangent of gradient is zero along the curve, which is known as Level set.

$\langle \frac{\nabla\varphi}{|\nabla\varphi|}, T \rangle$ **=0 --->** $\frac{\nabla\varphi}{|\nabla\varphi|}$ is perpendicular to T.

The curvature is the second order differentiation of the curve C in the normal direction.

$$\mathbf{C_{ss}=k}\vec{n} \quad \text{-----(4)}$$

where curvature k=div ($\frac{\nabla\varphi}{|\nabla\varphi|}$)

The velocity of deformation of initial

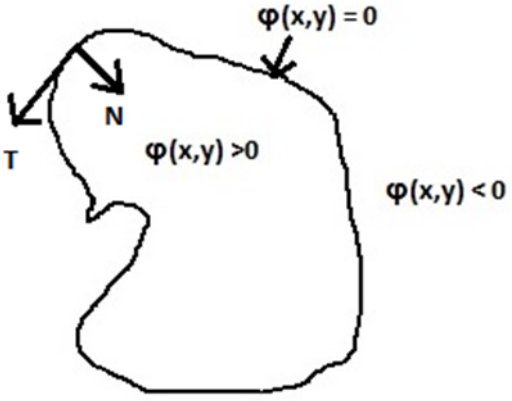


Figure 13: Level set function

curve is the first order differentiation of curve in the normal direction.

$$\frac{dC}{dt} = V\vec{N} \quad \text{-----(5)}$$

Change of curve function with Velocity V is expressed as:

$$\frac{d\varphi}{dt} = V|\nabla\varphi| \quad \text{-----(6)}$$

The evolution of the curve needs to be stopped once the curve fits the region of interest. The condition to stop the curve deformation is modeled as shown below:

$$\frac{dC}{dt} = (g(x,y)k - \{\nabla g(x,y), \vec{n}\})\vec{n} \quad \text{(7)}$$

where $(g(x,y)k - \{\nabla g(x,y), \vec{n}\})$ is nothing but velocity V, which needs to become zero to stop deformation / curve evolution.

$$g(x,y) = \frac{1}{\nabla I} \quad \text{-----(8)}$$

where ∇I is the gradient of the Image. The gradient of an image is the rate of change of the image pixel values. For a smooth portion of the image the gradient(rate of change of pixel values) is decreased and at edges the rate of change is increased, i.e. gradient is greater. The function g is minimal when the gradient of the image I is increasedi.e. at the edges. If g(x,y) is less (close to zero), the velocity decreases, and therefore, thecurve evolution stops. If the the g(x,y) is not zero for smooth edges, then another gradient term of g is introduced to stop the curve evolution.

As the curve re-initialization process is time consuming and complex, Kaihua Zhang, Lei Zhang proposed a Re-initialization free Level set evolution using Reaction Diffusion [16], which provide better results on segmenting nodules from CT scan in this work. The Level Set segmented output of CT image is shown in figure 14

The comparative analysis of the three segmentation methods implemented in this paper is shown in Table 1. Multi threshold method segmented time is less but it is not automatic. The user needsto fix the thresholds first to segment the lung lobe region from background, then the parenchyma region from the lobes, and finally the nodules from the parenchyma, which make this method unreliable, since more human inputs are required. ARGMM method is completely

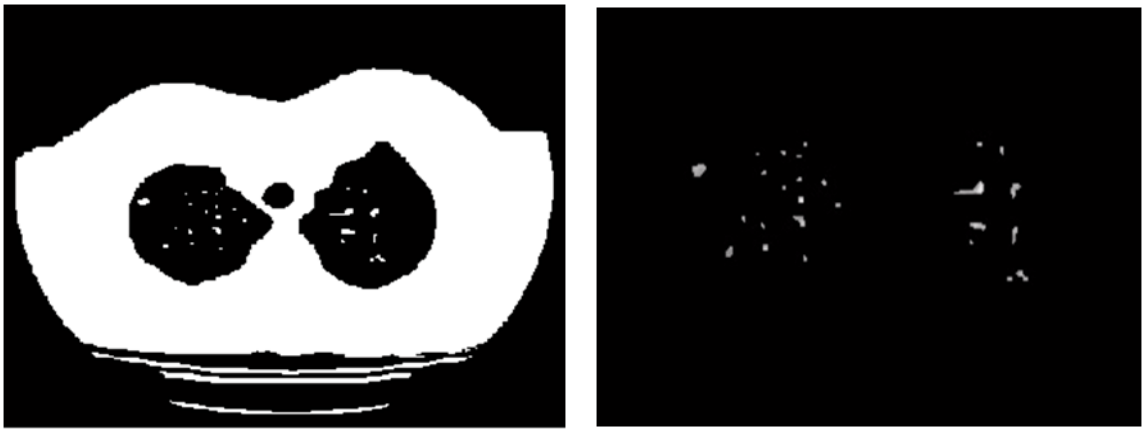


Figure 14: Level Set evolution Segmented output

	Time in seconds	Number of Segmented Cluster	Nature of Segmentation
Multi Threshold	10.4	64	Not Automatic
ARGMM	31.2	60	Fully Automatic
Level Set	14	31	Fully Automatic

Table 1

automatic, but it takes 31 seconds to segment the nodules;however, out of the 31 seconds, 26 seconds are necessary forthe region-growing algorithm. The Level Set Evolution method only takes 14 seconds, but the number of suspected nodules is only 31 compare to 60 in ARGMM, so the chances for not detecting some real cancerous in nature nodules may be increased in Level Set Evolution compared to ARGMM.

Feature extraction of suspected nodules

In section 3 the suspected nodules are segmented using three different methodologies. Approximately 60% of detected nodules in the lung CT scan would not lead to cancer. It is very important not to make the patient panic by categorizing all of the nodules as an indicator for cancer. Hence, finding the discriminative mathematical description for the cancerous and noncancerous nodules is a vital and challenging research area. The structural and Texture pattern of the cancerous and noncancerous nodules need to be mathematically analyzed, and quantitatively specified, so that we can classify the nodules based on their cancerous nature. This Pattern Recognition(PR) process has two steps, Feature Extraction and Classification. In this paper we focus on the feature extraction process.

The two main categories of features, which are relevant to this work, are Structural and Texture features. Structural features can be categorized as Shape and Size features. We applied a criteria based on area and eccentricity ratio value of each segmented nodule in section 3. The nodules thatmeasure less than 10 pixels in area and eccentricity, and with values that deviated from 1, were neglected for feature extraction, as these nodules are very small and very irregular in shape.

1. Structural Features

The Shape/Size features computed for the suspected nodules are the following. The shape/size features always used to compute Binary images, because while finding the shape/size features, the gray value of the pixel is not a matter of concern [17].

1.1 Area

The geometry of the nodules is non-linear; hence, it is not possible to derive a standard mathematical expression to find the area of nodules. The general concept behind the term 'Area' can be a model to find the area of any non-linear shape nodule. After converting the segmented nodule in to binary, simply by counting (adding) total number of 1's present in the image will give the value of Nodule Area in pixels.

$$\textbf{Area of nodule} = \sum_{x=1}^{m} \sum_{y=1}^{n} fb(x,y) \quad \textbf{(9)}$$

where fb(x,y) is the binary form of nodule image, m = number of rows and n= number of columns

1.2 Minor Axis Length & Minor Axis Length

The maximum horizontal diameter of an image is known as Minor Axis Length. This value can be computed by counting the 1's in each row of the binary nodule image, and finding its maximum value. Similarly, the Major Axis Length can be computed in a vertical direction.

$$\textbf{Minor Axis Length} = \max\left\{ \sum_{x=1}^{m} fb(x,:) \right\} \textbf{-------(10)}$$

$$\textbf{Major Axis Length} = \max\left\{ \sum_{y=1}^{n} fb(:,y) \right\} \textbf{--------(11)}$$

1.3 Elongation

Elongation is the ratio of minor to major axis length. This value is close to 1 if the shape is closer to a circle.

$$\textbf{Eccentricity} = \frac{\max\{\sum_{x=1}^{m} fb(x,:)\}}{\max\{\sum_{y=1}^{n} fb(:,y)\}} \textbf{-----(12)}$$

1.4 Equivalent Diameter

Equivalent Diameter is defined as

$$\textbf{ED} = \frac{2}{\sqrt{\pi}}\sqrt{area} \textbf{-----(13)}$$

1.5 Solidity

Solidity is defined as a ratio between the Area to the Convex area, where convex

area is the total number of pixels inside the outline's Convex hull.

$$\text{Solidity} = \frac{\text{Area}}{\text{ConvexArea}} \text{-------(14)}$$

1.6 Extent

Extent is defined as the ratio between Area to the Area of bounding box

$$\text{Extent} = \frac{\text{Area}}{\text{Area of Bounding Box}} \text{-------(15)}$$

1.7 Perimeter

The outline of the nodule image is found using morphological function and the total number of pixels that form the outline, is nothing but the perimeter of that nodule.

The Shape/Size features determined on the suspected nodules segmented using ARGMM method is tabulated in Table 2 (page 109)

2. Statistical Texture Features

Texture is the measure of surface variation. There is a discriminative difference in the textural features of cancerous and noncancerous nodules. Hence, it is very important to mathematically model a texture feature that lead to an efficient and accurate classification of nodules. The basic low level statistical intensity based texture features,such as Mean, Standard Deviation, Max/Min Intensity, Intensity difference, an Mean Intensity Difference are not very discriminative powerful features, as these features are computed directly from the intensity values of the ROI.

2.1 Co-Occurrence matrix feature

The spatial dependency of gray level values can be modeled as a texture feature [18]. The Gray-Level co-occurrence matrix is formed by counting all pairs of pixels separated by distance vector 'd' having gray levels i and j. The Displacement vectorestablishes in which direction the (i,j) occurrence of the pixels need to be counted. The features that can be calculated from a co-occurrence matrix are tabulated in Table 4

Displacement Vector	Angle in deg
[0 D]	0
[-D D]	45
[-D 0]	90
[-D -D]	135
*Where D is the spatial distance	

Table 3

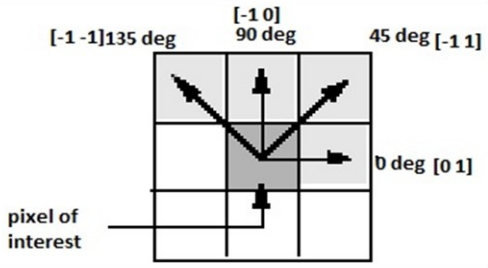


Figure 15

Entropy	$\sum_{i=1}^{m}\sum_{j=1}^{n}$ **(P[i,j]*logP[i,j])**
Energy	$\sum_{i=1}^{m}\sum_{j=1}^{n}$ $P^2[i,j]$
Contrast	$\sum_{i=1}^{m}\sum_{j=1}^{n}$ $(i-j)^2 P[i,j]$
Homogeneity	$\sum_{i=1}^{m}\sum_{j=1}^{n} \frac{P[i,j]}{1+\lvert i-j\rvert}$
Correlation	$\sum_{i=1}^{m}\sum_{j=1}^{n} \frac{\{i*j\}*P[i,j]-(\mu x*\mu y)}{(\sigma x*\sigma y)}$
P[i,j] =occurrence of the gray level pixel pair (i,j) μx , μy = mean in row and column direction σx, σy= Standard deviation in row and column direction	

Table 4

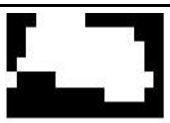
Suspected Nodule	Area	Major Axis Length	Minor Axis Length	Eccen-tricity	Equi Dia	Solidity	Extent	Pe-rimeter
	77.00	13.05	7.99	0.79	9.90	0.96	0.74	33
	43.00	9.29	5.99	0.76	7.40	1.00	0.77	21
	27.00	8.10	5.65	0.72	5.86	0.87	0.55	21
	16.00	7.57	3.12	0.91	4.51	0.89	0.53	15
	16.00	5.77	3.68	0.77	4.51	1.00	0.80	12
	18.00	6.00	4.18	0.72	4.79	0.95	0.72	13
	21.00	6.22	4.59	0.67	5.17	0.91	0.70	15
	40.00	11.99	5.00	0.91	7.14	0.83	0.50	26
	28.00	6.78	5.36	0.61	5.97	1.00	0.78	16
	16.00	5.42	4.09	0.66	4.51	0.89	0.64	13
	56.00	14.63	5.41	0.93	8.44	0.82	0.62	34
	20.00	5.34	4.93	0.38	5.05	0.95	0.80	13
	49.00	10.92	6.94	0.77	7.90	0.84	0.49	28

Table 2

The Co-occurrence matrix Texture features determined on the suspected nodules segmented using ARGMM method, is tabulated in Table 5

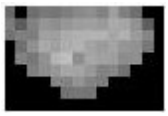
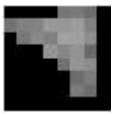
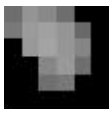
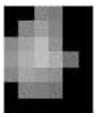
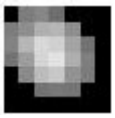
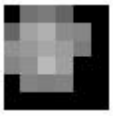
Suspected Nodules	Contrast	Correlation	Energy	Homogeneity
	1.752137	0.787722	0.143984	0.76567
	2.52381	0.598772	0.163517	0.72619
	2.696429	0.620449	0.220026	0.76875
	2.694444	0.551083	0.266975	0.773611
	2.88	0.549931	0.1648	0.736
	3.4	0.546946	0.148889	0.645
	3.828571	0.578749	0.15102	0.679524
	2.366667	0.727925	0.253086	0.754444
	3.142857	0.705562	0.106576	0.605556
	3.633333	0.465872	0.168889	0.697778
	1.685714	0.834278	0.201633	0.787937
	3.433333	0.578866	0.124444	0.662778
	2.409091	0.728048	0.265289	0.777576

Table 5

3. Frequency Domain Feature

Transforming the domain feature may be more discriminative for some cases, so in this work we also proposed to find frequency domain feature of the suspected segmented nodules. Discrete Fourier Transform will segregate the image regions in Low and High frequency regions, but to extract the quantitative values efficiently, it is essential to arrange the pixels using equation (16) before applying Fourier Transform, so that all the pixels corresponding to low frequency accumulated in the center and in various radial directions,to allow that pixels corresponding to high frequency gather at corners.

$$f(x,y) = (-1)^{(x+y)} * g(x,y) \text{ -----(16)}$$

$$F(u,v) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)*\exp[-j2\pi(\frac{mu}{M} + \frac{nv}{N})] \text{ -------(17)}$$

where u=0, 1 ----- M-1
v=0,1 ------N-1

The center value (M/2, N/2) of the frequency transformed image has a zero frequency value, it is also known as the DC constant, which can be a powerful feature during classification. The frequency values at the different radii levels can also be used as a discriminative feature, which are tabulated in below Table 6 (page 112).

4. Complete Local Binary Pattern Feature

Complete Local Binary Pattern (CLBP) evolved from the Local Binary Pattern (LBP)[19][20][21]. The difference between the center pixel and its neighbor pixel can be mathematically modeled as:

$$LBP_{P,R} = \sum_{p=0}^{P-1} Z(g_p-g_c)2^p,\ Z(x) = 1,\ x>=0$$
$$= 0,\ x<0\text{-------(18)}$$

Where:
g_c = gray level value of center pixel
g_p = gray value of the neighbor of gc
P = Total number of neighbor

The function above, applied on all the pixels in an image, and the pixel value is replaced by the decimal equivalent of the binary coded value, is determined by the equation. The LBP matrix computed could be framed as a Histogram using the mathematical function below, which can be used as powerful texture feature.

$$H(k) = \sum_{i=1}^{m}\sum_{j=1}^{n} f(LBP_{P,R}(i,j),\ k),\ k\epsilon[0,K]\text{-------(19)}$$
$$f(x,y) = 1,\ x=y$$
$$= 0,\ \text{otherwise}$$

The number of transitions in the LBP pattern also can be modeled as one texture feature, which can be mathematically defined as:

$$U(LBP_{P,R}) = |Z(g_{p-1} - g_c) - Z(g_0-g_c)| + \sum_{p=1}^{P-1} |Z(g_p-g_c) - Z(g_{p-1} - g_c)| \text{ ---(20)}$$

We can selectively choose the LBP patterns that have a certain U value. $LBP_{P,R}^{U2}$implies that Uniform LBP with U<=2.

In Complete LBP, the difference between the center and the neighbor pixel is separately modeled as Sign and Magnitude.

$$dp=gc - gp \text{ -------(21)}$$
$$completedp = sign(dp) * magnitude(dp) \text{ -------(22)}$$

sign(dp) = 1 when gc>=gp
= 0 when gc<gp
CLBP_S = sign component of LBP
CLBP_M= Magnitude component of LBP

CLBP_S code pattern is similar to the LBP pattern, but CLBP_M need to be normalized properly to create a proper code pattern using the following function

$$CLBP\_M_{P,R} = \sum_{p=0}^{P-1} t(mp,c)2^p,\ t(x,c) = 1,\ x>=c$$
$$= 0,\ x<c\text{-------(23)}$$

Where:
mp is the magnitude of dp
c is the Threshold (can be the mean value of mp of the whole image)

The MLBP and SLBP Histograms of different suspected segmented nodules are shown in figure 16 and figure 17 (page 113). This feature can be made more reliable and discriminative by constructing 3D Histogram with MLBP and SLBP, as shown in figure 18 (page 114-11). These Histogram feature values can be used to classify the cancerous nature of the suspected nodules in the classifier step.

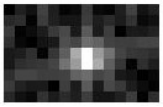
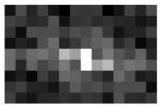
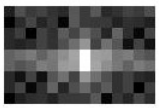
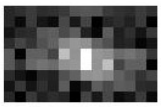
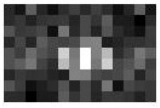
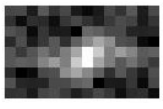
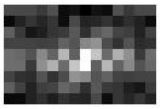
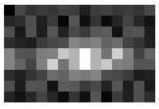
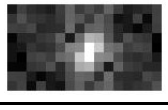
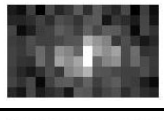
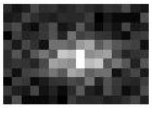
Suspected Nodules Fourier Transform Spectrum	DC Coeff	Center Coeff	Entropy
	0.286275	11.11196518	3.155945102
	0.411765	3.93876198	2.520709684
	0.6	1.436000633	2.412769822
	-0.98431	2.408394409	2.163501478
	0.411765	1.767854286	2.841078322
	-0.97647	1.171434373	2.19524751
	0.796078	0.739540952	2.582404943
	1.552941	3.907692964	2.359159604
	-1.64314	1.309012745	2.336024133
	-0.12157	1.758805698	2.26764773
	0.423529	3.399988552	2.599496047
	-0.56863	0.638217912	1.710340549
	1.133333	2.726618869	2.432006213

Table 6

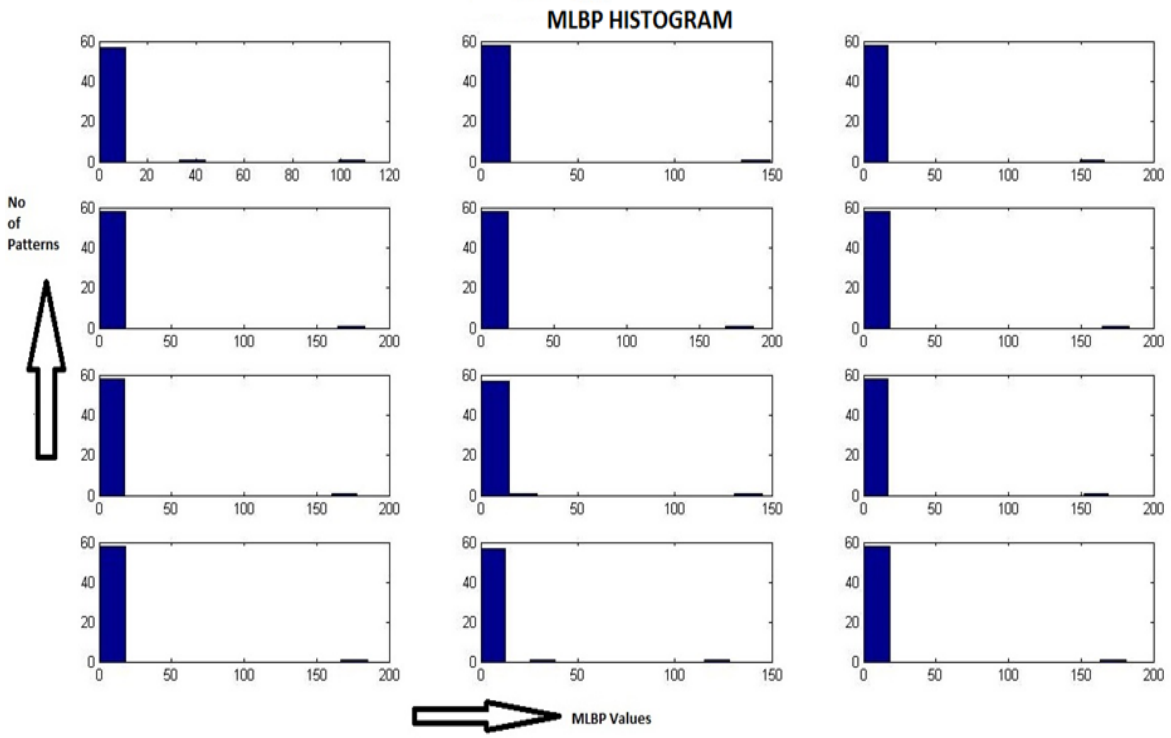


Figure 16

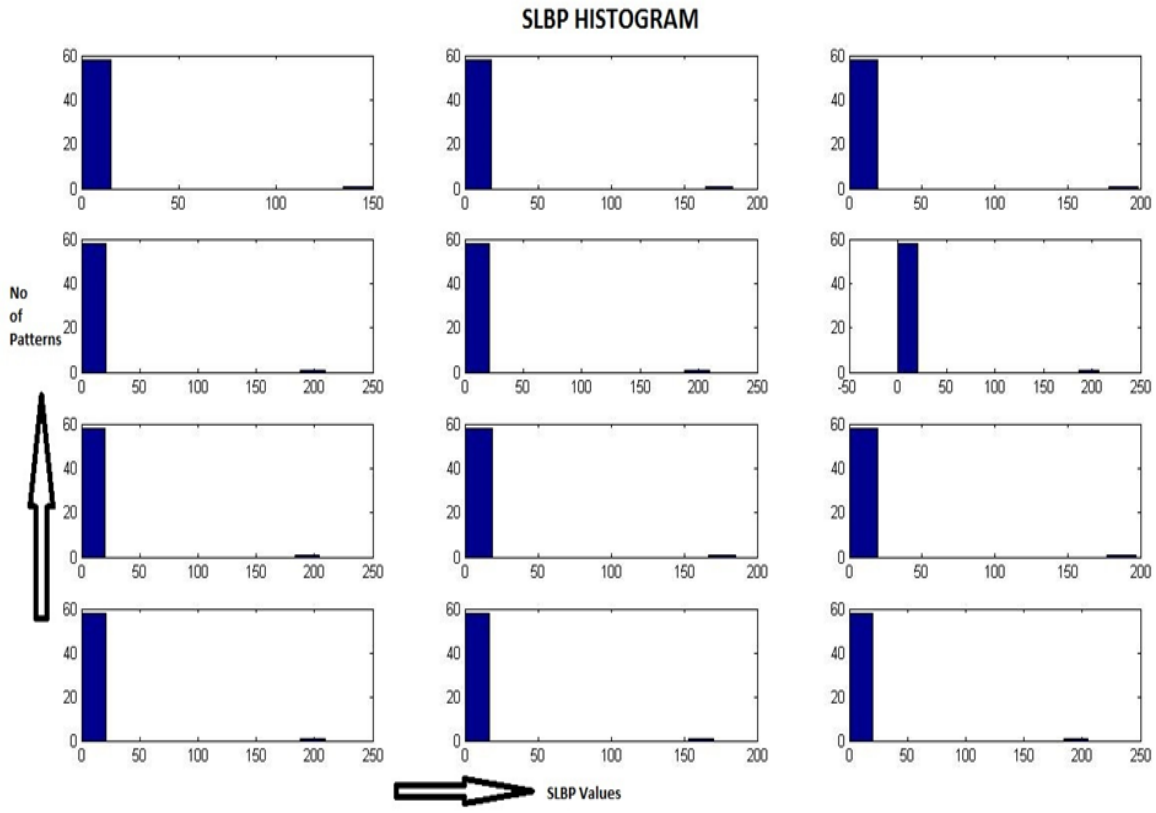


Figure 17

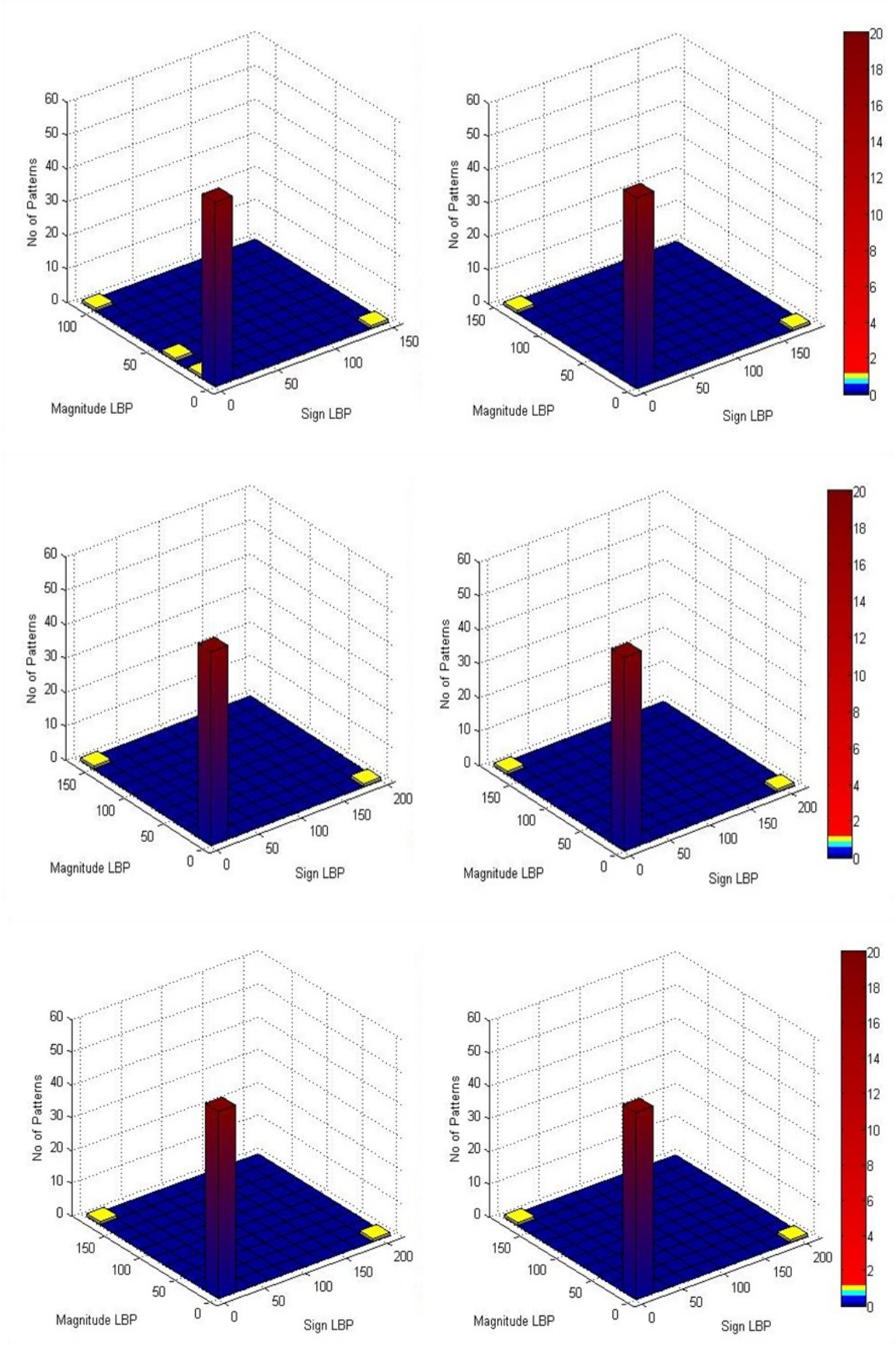


Figure 18a: 3D Histogram between Magnitude LBP and Sign LBP

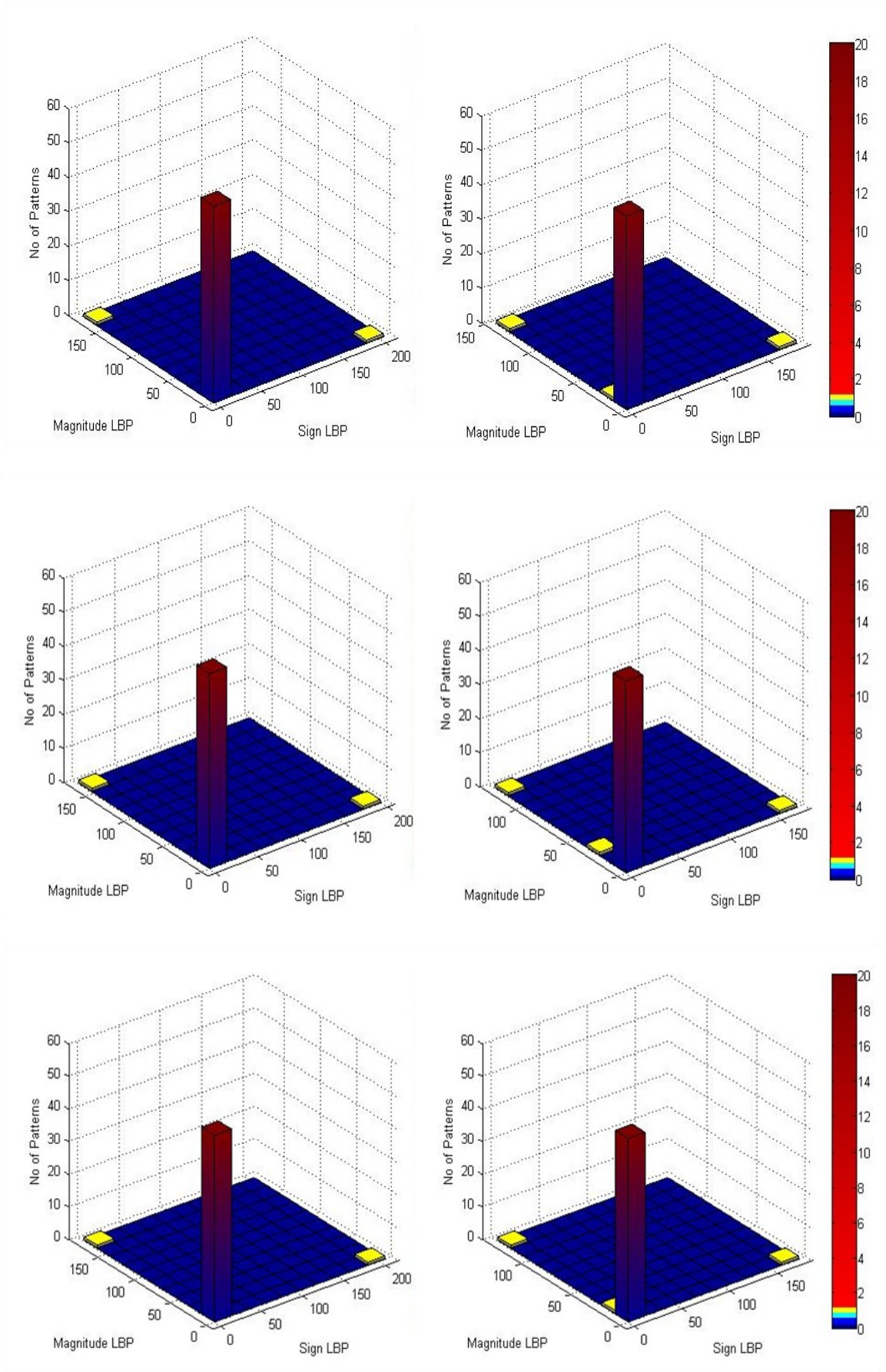


Figure 18b: 3D Histogram between Magnitude LBP and Sign LBP

5. Microscopic Information Pattern

In the Local Configuration Pattern technique [22] the image is resolved in to two features namely Local Structural Information(LSI) and Microscopic Information Pattern(MIP). LSI is nothing more than a Local Binary pattern feature defined in equation 18. MIP contain Image configuration along with pixel-wise interaction relationship detail.

The Local Binary Pattern of figure 19a is 11001111 and figure 19b is 11001111. Both of these patterns are the same, but obviously these two image regions are different, one is smooth and the other one is coarse. To overcome this, the Variance value needs to be considered along with the LBP pattern to make this a powerful feature.

$$\mathbf{VAR} = \frac{1}{P}\sum_{p=1}^{P-1} \mathbf{(g_p-\mu)^2 \text{-------(24)}}$$

This LBP/Variance can't solve the issue with figures 19b and 19c. Both of theseimage sections have the same LBP pattern and also the same variance. Hence, the concept of Microscopic Configuration is used to differentiate these two image patterns quantitatively. MCI estimates the optimal weight for each pattern, with respect to the intensities of neighboring pixels to linearly reconstruct the central pixel intensity.

$$\mathbf{E(w_0,w_1, \text{- - - - - } w_{p-1}) = | g_c - \sum_{i=1}^{P-1} a_i g_i | \text{-------(25)}}$$

wherew_i is weight parameter

$E(w_0,w_1, \text{- - - - - } w_{p-1})$ is reconstruction error

Yimo Guo, Guoying and Matti developed a Least mean square estimation model [ref] to reduce the reconstruction error.

There are N_L pixels in I with the pattern type L. Setting the intensities of these NL pixels as C_L,i(i=0, - - - N_L-1) and forming them as vector below:

$$\mathbf{C_L = (C_{L,0} C_{L,1} \text{ - - - - } C_{L,NL-1})^T \text{ ---(26)}}$$

The intensities of neighboring pixels V (i;0) , - - - - V(i;P-1) where i = 0 ----- NL-1 can be organized as below:

$$\mathbf{V_L} = \begin{pmatrix} V(0;0) & V(0;1) & - - - - - & - & V(0;P-1) \\ V(1;0) & V(1;1) & - - - - - & - & V(1;P-1) \\ & - & & - & - \\ & - & & - & - \\ V(NL-1;0) & V(NL-1;1) & - - - - & - & V(NL-1;P-1) \end{pmatrix}$$

-------(27)

The weight vector can be expressed as $W_L = (w_0,w_1, \text{- - - - - } w_{p-1})^T$

The CL can be expressed as in terms of V_L and W_L as:

$$\mathbf{C_L = V_L W_L \text{ -------(28)}}$$

$$\mathbf{A_L = (V_L^T V_L)^{-1} V_L^T C_L \text{-------(29)}}$$

F_L = FT(AL) where FT is Fourier Transform

The Fourier Transform of the A_L can make this MIP as a rotational invariant, and the magnitude of the Fourier transform of the A_L can be used as reliable feature

The CLBP's Magnitude and Sign deviations, along with the MIP mean values are tabulated in Table 7 (page 117). The Microscopic Information pattern histograms for segmented ARGMM nodules are shown in figure 20 (page 118). These Histogram feature values, along with deviation and mean values, can be used to classify the cancerous nature of the suspected nodules in the classifier steps of this work in the future. In this paper the classification work in not shown.

58	61	53
62	55	51
63	57	59

Figure 19 a

121	175	0
243	55	7
75	188	275

Figure 19 b

243	121	7
275	55	0
175	75	188

Figure 19 c

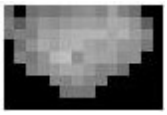
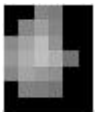
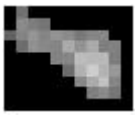
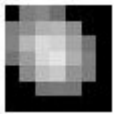
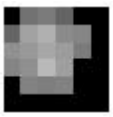
Suspected Nodules	CLBP Magnitude Deviation	CLBP Sign Deviation	MIP mean
	15.14467538	19.44530311	0.673246
	19.356434	23.76281837	0.234568
	21.55624126	25.72900857	0.197531
	23.76499496	27.18358694	0.061728
	24.28954265	27.18041547	0.049383
	23.7722488	26.91652138	0.049383
	23.12221993	26.52030749	0.074074
	18.96318991	24.02044716	0.034065
	21.93915927	25.59934979	0.111111
	24.04268802	27.18041547	0.049383
	16.9954097	22.05125266	0.382716
	23.49796365	26.64936732	0.049383
	17.34385967	22.82880408	0.035023

Table 7

Conclusion

The Segmentation and Feature extraction on Lung CT cross-section images was performed, and the results were tabulated quantitatively in this paper. In this work the suspected nodules were segmented from the Lung CT image cross-sections using threshold based, ARGMM and Levelset evolution methods, in which ARGMM and Levelset are completely automated. The number of suspected nodules segmented from the ARGMM is greater compared to the Levelset method, which helps to analyze even the small nodule clusters with a cancerous nature. On the other hand, the time taken for the ARGMM is grater compared to the Levelset method. After performing the segmentation, the different features are computed on suspected nodules. The Shape and Size features were computed on the binary suspected nodule images. The Texture features were computed from a Gray Level Co-occurrence matrix of the suspected nodule images. To make the texture feature more discriminative, the Complete Local Binary Pattern(CLBP) algorithm was applied on the segmented nodule images, and their pattern histograms were computed separately for the patter magnitude and its sign, which can be efficiently merged and used as the feature during classification. To overcome the difficulties in CLBP, the Microscopic Information Pattern (MIP) features were computed to numerically characterize the minute variation in the segmented nodules. The LBP along with the MIP algorithm can be characterized as a Local Configuration Model, from which the quantitative values can be used as a powerful feature during the classification steps of this work in the future.

References

1. Kostis WJ, Reeves AP, Yankelevitz DF, Henschke CI. Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images. IEEE Trans Med Imaging 2003;22:1259–1274.
2. Gould MK, Donington J, Lynch WR, et al. Evaluation of Individuals With Pulmonary Nodules: When Is It Lung Cancer?: Diagnosis and Management of Lung Cancer, 3rd ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest 2013;143(5 Suppl):e93S-e120S.
3. Chheang S, Brown K. Lung Cancer Staging: Clinical and Radiologic Perspectives. Seminars in Interventional Radiology 2013;30(2):99-113.
4. Xu C, Hao K, Song Y, Yu L, Hou Z, Zhan P. Early diagnosis of solitary pulmonary nodules. Journal of Thoracic Disease 2013;5(6):830-840.

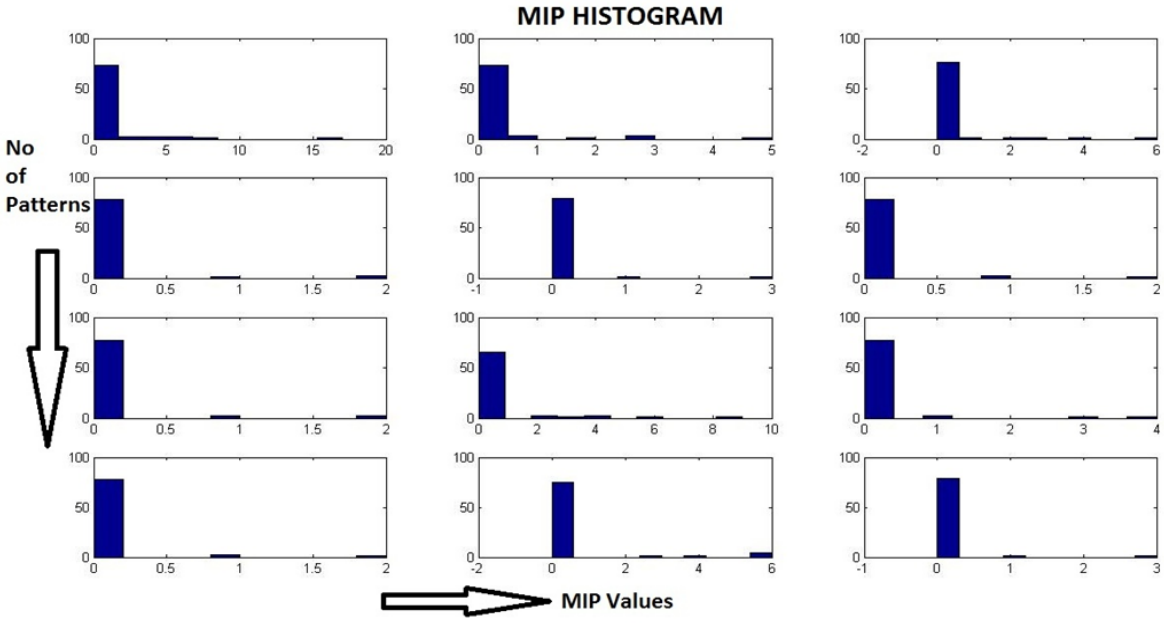


Figure 20

5. Ost DE, Gould MK. Decision Making in Patients with Pulmonary Nodules. American Journal of Respiratory and Critical Care Medicine 2012;185(4):363 -372.
6. Armato SG III, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans. Medical Physics, 38: 915--931, 2011.
7. Wei Y, Shen G, Li J. A Fully Automatic Method for Lung Parenchyma Segmentation and Repairing. Journal of Digital Imaging 2013;26(3):483-495.
8. Senthil Kumar TK, Ganesh EN. Proposed Technique for Accurate Detection/Segmentation of Lung Nodules using Spline Wavelet Techniques. International Journal of Biomedical Science : IJBS 2013;9(1):9-17.
9. Li Q. High performance lung nodule detection schemes in CT using local and global information. Medical Physics 2012;39(8):5157-5168.
10. Qi S, Van Triest HJW, Yue Y, Xu M, Kang Y. Automatic pulmonary fissure detection and lobe segmentation in CT chest images. BioMedical Engineering OnLine 2014;13:59.
11. Khan AN, Al-Jahdali HH, Irion KL, Arabi M, Koteyar SS. Solitary pulmonary nodule: A diagnostic algorithm in the light of current imaging technique. Avicenna Journal of Medicine 2011;1 (2):39-51.
12. Rueda L, Domingo M, Kittler J. Segmentation Using Automatic Seeded Region Growing and Instance-Based Learning. Springer Berlin Heidelberg 2007-01-01. P 192-201
13. Otsu N. A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man, and Cybernetics. (1979) vol. 9, no. 1,pp. 62-66.
14. Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models. International Journal of Computer Vision. (1988) Volume 1, Issue 4 , pp 321-331
15. Osher, Sethian. Book: Numerical Geometry of Images. pp 50-60, ISBN 978-1-4684-9535-5
16. Zhang K, Zhang L, Song H, Zhang D. Re-initialization Free Level Set Evolution via Reaction Diffusion. IEEE Trans. on Image Processing.
17. Mingqiang Y, KidiyoKpalma, Ronsin J. A Survey of Shape Feature Extraction Techniques. Pattern Recognition, IN-TECH. (2008) pp.43-90.
18. Haralick RM, Shanmugan K, Dinstein I. Textural Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics. (1973) Vol. SMC-3, , pp. 610-621
19. Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with Local Binary Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence (2002) 24(7):971-987.
20. Kumar T, Narasimhan G, Umamaheswari R. Texture Pattern Based Lung Nodule Detection (TPLND) Technique in CT Images. International Review on Computers and Software (IRECOS)(2014), 9 (3), pp. 415-426.
21. Zhenhua G, Zhang D. A Completed Modeling of Local Binary Pattern Operator for Texture Classification," Image Processing, IEEE Transactions. (2010) vol.19, no.6, pp.1657,1663.
22. Guo Y, Zhao G, Pietikäinen M.Texture classification using a linear configuration model based descriptor. Proc. the British Machine Vision Conference (BMVC 2011), Dundee, UK, 119.1-119.10