Computer Aided Diagnosis System for Detection of Lung Cancer in CT Scan Images

Disha Sharma and Gagandeep Jindal

***Abstract*—The automated Computer Aided Diagnosing (CAD) system is proposed in this paper for detection of lung cancer form the analysis of computed tomography images. In recent years the image processing mechanisms are used widely in several medical areas for improving earlier detection and treatment stages, in which the time factor is very important to discover the disease in the patient as possible as fast, especially in various cancer types such as the lung cancer, breast cancer. Lung cancer is the second most commonly diagnosed cancer in the United States, and it is the leading cancer related death in the world, with the current fatality rate exceeding that of the next three most common cancers (breast, prostate, and colorectal) combined. In this research, we considered the problem of developing an automated system for detecting the presence of pulmonary nodules in the lung CT .The essence of developing a system like that needed to focus on detecting nodules in their early stages, which are the very small nodules that are likely to be overlooked by the radiologists. This paper involves cancer detection system based on texture features extracted from the slice of DICOM Lung CT images for the identification of cancerous nodules. In developing this system we passed the available lung CT images and its database in basic three stages to achieve more accuracy in our experimental results: Firstly A pre-processing stage involving some image enhancement techniques helps to solve the problem. We preprocess the images (by contrast enhancement, thresholding, filtering, and blob analysis) obtained after scanning the Lung CT Images and secondly separate the suspected nodule areas (SNA) from the image by a segmentation process by using thresholding segmentation mechanism by Otsu thresholding algorithm and region growing techniques. Finally we relied on Texture features which help us to make a comparison between cancerous and non cancerous images. For accurate detection of cancerous nodules, we need to differentiate the cancerous nodules from the noncancerous. We developed an artificial neural network to differentiate them. We trained the neural network by the backpropagation algorithm and tested it with different images from a database of the DICOM CT Lung images of NIH/NCI Lung Image Database Consortium (LIDC) dataset.**

***Index Terms*—Segmentation, Extraction, Computer aided diagnosis, Region Growing, DICOM, ROC.**

I. Introduction

Cancer is one of the most serious health problems in the world field. The mortality rate of lung cancer is the highest among all other types of cancer. Lung cancer is one of the most serious cancers in the world, with the smallest survival rate after the diagnosis, with a gradual increase in the number of deaths every year. Survival from lung cancer is directly related to its growth at its detection time. The earlier the detection is, the higher the chances of successful treatment are. An estimated 85% of lung

Cancer cases in males and 75% in females are caused by cigarette smoking [1] In 2005, approximately 1,372,910 new cancer cases are expected and about 570,280 cancer deaths are expected to occur in the United States. It is estimated that there will be 163,510 deaths from lung cancer, which forms 29% of all cancer deaths. The overall survival rate for all types of cancer is 63%. Although surgery, radiation therapy, and chemotherapy have been used in the treatment of lung cancer, the five year survival rate for all stages combined is only 14%. This has not changed in the past three decades [2]. The purpose of this paper is to develop a CAD system for early detection of lung cancer based on an automatic diagnosis of the lung regions included in chest CT images. The difficulties for detecting lung nodules in radiographs are below:

❖ Nodule sizes are varying widely: Commonly a nodule diameter can take any value between a few millimeters up to several centimeters.

❖ Nodules exhibit a large variation in density – and Hence visibility on a radiograph – some nodules are Only slightly denser than the surrounding lung tissue, While the densest ones are calcified.

❖ As nodules can appear anywhere in the lung field, they can be obscured by ribs, and structures beneath the diaphragm, resulting in a large variation of contrast to the background. To overcome these problems, the author proposed a Computer Aided Diagnosing (CAD) system for Detection of lung nodules. This system consists of five main steps, as shown in Figure 1. In this paper we apply the different image processing techniques such as Erosion, wiener Filter, Dilation, Lung Border Extraction thresholding, region growing, edge detection, ridge detection, morphological operations, fitting of geometrical models or functions and dynamic programming for extraction of lung region. Then segmentation algorithm is applied. [10]

II. Present Work

Ginneken [3] has classify the lung regions extraction approaches into two different categories; either rule-based or pixel classification based category. Most of the proposed approaches belong to rule-based category [4-5], where a sequence of steps, tests and rules are used in the extraction process. Techniques employed are (local) thresholding, region growing, edge detection, and ridge detection, morphological operations, fitting of geometrical models or functions and dynamic programming. On the other hand,

Manuscript received August 16,2011; revised October 10, 2011.

Disha Sharma is M.Tech Student, in CEC, Landran, Punjab. (e-mail: disha.sharma22@gmail.com)

Gagandeep Jindal is Assistant Professor, in CEC, Landran, Punjab.

there is another approach used in lung regions extraction process based on pixel classifications, where each pixel in the CT image is classified into an anatomical class (usually lung or background, but in some cases more classes such as heart, mediastinum, and diaphragm). Classifiers are various types of neural networks, or markov random field modeling, trained with a variety of local features including intensity, location, and texture measures [3]. CADs can be divided into two groups [6]: density-based and model-based approaches. Considering the fact that lung nodules have relatively higher densities than those of lung parenchyma, density-based detection methods employ techniques such as multiple thresholding, region-growing, locally adaptive thresholding in combination with region growing, opening and closing, using the histogram, the top 20% gray values considered as initial cancerous candidate regions, using the histogram the normal tissues are removed, then elliptical-shaped regions, which is in general represent abnormalities, are detected, and fuzzy clustering used to identify nodule candidates in the lungs. False-positive results can then be reduced from the detected nodule candidates by employing a priori knowledge of small lung nodules. For the model-based detection approaches, the relatively compact shape of a small lung nodule is taken into account while establishing the models to identify nodules in the lungs. Techniques such as Morphological filter and the anatomy based generic model have been proposed to identify sphere shaped small nodules in the lung. Nodule candidates are detected using template matching or a modified Hough transform in which edge pixels vote for circles that could cause these edges. After getting the segmentation results, different features should be extracted to be used in the diagnosis phase where sets of rules are formulated to distinguish between true and false cancerous candidates. Different features were extracted in different papers depending on the methods used by the authors in the diagnosis phase. In some approaches uniformity, connectivity, and position features were extracted [7]. In [5] the features such as size, circularity, and mean brightness of region of interests (ROIs) were extracted. Area, thickness, circularity, intensity, variance, localization, and distance from the lung wall are the extracted features in [4]. The underlying idea of developing a CAD system is not to delegate the diagnosis to a machine, but rather that a machine algorithm acts as a support to the radiologist and points out locations of suspicious objects, so that the overall sensitivity is raised.CAD systems meet four main objectives, which are improving the quality and accuracy of diagnosis, increasing therapy success by early detection of cancer, avoiding unnecessary biopsies and reducing radiologist interpretation time. [16,17].

III. Lung CT Image

The lung CT images having low noise when compared to scan image and MRI image. So we can take the CT images for detecting the lungs. The main advantage of the computer tomography image is its better clarity, low noise and distortion. The mean and Variance can be easily calculated. The calculated value is very closer to the original value.

As we can in (Fig. 2), a typical lung CT consists of the following components:

1. The two lungs (left and right)
2. A cross section of a vertebra
3. Bone, only for the sections where a cross section of the ribs was present, for the slices taken at the intercostals spaces, bone part in the CT will not appear.
4. Sternum, similar to bone, sternum is present only in the top CT slices of the thorax.
5. Trachea, a big hole that varies in size along the CT slices of the thorax, and diminishes completely when the trachea branches to the two bronchi.
6. Mediastinium, which is the part in the thorax that contains the heart and is bound by the sternum from the front and spinal cord from the back.
7. Fat and muscle.

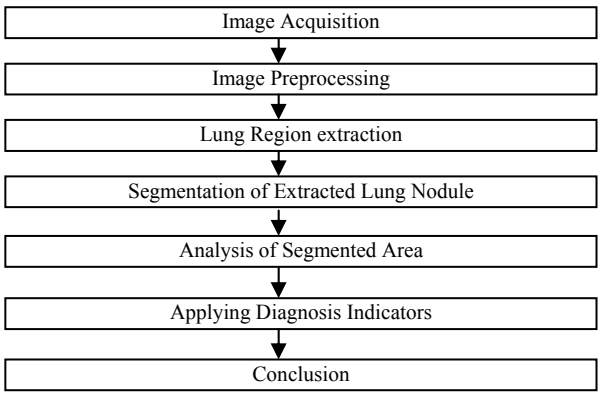

Fig. 1. Block diagram representation of diagnosis procedure.

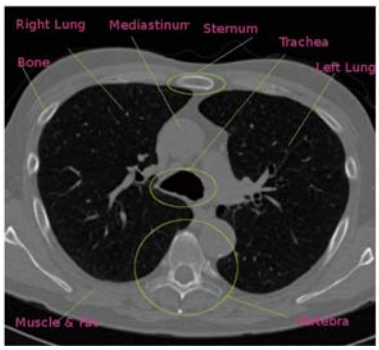

Fig. 2. The Lung CT Image

IV. Preprocessing

Preprocessing is the initial step for detecting the lung cancer. In preprocessing step we have done two steps. They are:1. Denoising; 2. Wiener Filter

A. Denoising

Image denoising algorithm may be mostly used in image processing. Many methods, regardless of implementation, share the same basic idea noise reduction through image blurring. Blurring can be done locally, as in the Gaussian smoothing model or in anisotropic filtering by calculating the variations of an image. White noise is one of the most common problems in image processing. Even a high resolution photo is bound to have some noise in it. For a high resolution photo a simple box blur may be sufficient, because even a tiny features like eyelashes or cloth texture will be represented by a large group of pixels. However, current DirectX 10 class hardware allows us to implement high

quality filters that run at acceptable frame rates. The main idea of any neighborhood filter is to calculate pixel weights depending on how similar their colors are. We describe two such methods: the K Nearest Neighbors and Weiner filters. The input image is a normal RGB image. The RGB image is converted into grey scale image because the RGB format is not supported in Matlab. Then the grey scale image contains noises such as white noise, salt and pepper noise etc. This can be removed by using wiener filter from the extracted lung image.

B. Wiener Filter

The goal of the Wiener filter is to filter out noise that has corrupted a signal. It is based on a statistical approach. Typical filters are designed for a desired frequency response. However, the design of the Wiener filter takes a different approach. One is assumed to have knowledge of the spectral properties of the original signal and the noise, and one seeks the linear time-invariant filter whose output would come as close to the original signal as possible.[20] Wiener filters are characterized by the following:

1. Assumption: signal and (additive) noise are stationary linear stochastic processes with known spectral characteristics or known autocorrelation and cross-correlation
2. Requirement: the filter must be physically realizable/causal (this requirement can be dropped, resulting in a non-causal solution)
3. Performance criterion: minimum mean-square error (MMSE)

The orthogonality principle implies that the Wiener filter in Fourier domain can be expressed as follows:

$$W(f_1, f_2) = \frac{H^*(f_1, f_2) S_{xx}(f_1, f_2)}{\left| H(f_1, f_2) \right|^2 S_{xx}(f_1, f_2) + S_{\eta\eta}(f_1, f_2)} \quad (1)$$

where $S_{xx}(f_1, f_2), S_{\eta\eta}(f_1, f_2)$ respectively are power spectra of the original image and the additive noise and $H(f_1, f_2)$ is the blurring filter. It is easy to see that the Wiener filter has two separate parts, an inverse filtering part and a noise smoothing part. It not only performs the deconvolution by inverse filtering (high pass filtering) but also removes the noise with a compression operation (low pass filtering).

Implementation: To implement the Wiener filter in practice we have to estimate the power spectra of the original image and the additive noise. For white additive noise the power spectrum is equal to the variance of the noise. To estimate the power spectrum of the original image many methods can be used. A direct estimate is the period gram estimate of the power spectrum computed from the observation:

$$S_{yy}^{per} = \frac{1}{N^2}[Y(k,l)Y(k,l)^*] \quad (2)$$

where *Y(k,l)* is the DFT of the observation. The advantage of the estimate is that it can be implemented very easily without worrying about the singularity of the inverse filtering. Another estimate which leads to a cascade implementation of the inverse filtering and the noise smoothing is

$$S_{xx} = \frac{S_{yy} - S_{\eta\eta}}{|H|^2} \quad (3)$$

which is a straightforward result of the fact: The power spectrum can be estimated directly from the observation using the periodogram estimate. This estimate results in a cascade implementation of inverse filtering and noise smoothing:

$$W = \frac{1}{H} \frac{S_{yy}^{per} - S_{\eta\eta}}{S_{yy}^{per}} \quad (4)$$

V. Extracting Lung Region

It is often important to separate regions or objects of interest from other parts of the image. Methods for performing segmentations vary widely depending on the specific application, Imaging modality, and other factors. For example In some applications it may be useful to classify image pixels into anatomical regions, such as bones, muscles, and blood vessels, while in others into pathological regions, such as cancer, tissue deformities and multiple sclerosis lesions. The purpose of the segmentation of the lung region in the CT image is to achieve a better orientation in the image. It starts by applying the image slicing algorithm to each DICOM CT image of the raw data. The resulting binary images are then analyzed to choose among them the best image that may help in extracting the lung regions from the raw CT-image data with a certain degree of accuracy and sharpness. To refine the chosen image, other techniques were used for different purposes in a sequence of steps. Erosion, median filter and dilation steps aim to eliminate irrelevant details that may add extra difficulties to the lung border extraction process. The outlining step aims to extract the structure's borders. The lung border extraction step aims at separating lung structure from all other uninteresting structures. [11].

VI. Segmenting Extracted Lung Nodule

The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image

Segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics [8]. The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image (edge detection). Each of the pixels in a region is similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic(s).We has done this using edge detection and steps for it are:

1. Edge is a set of connected pixels that lie on the boundary between two regions.
2. Edges are detected by Sobel Methods.
3. We choose Sobel method because of its Accuracy.

Sobel Edge Detection Method

1. Convolve the image g(r, c) to get smooth Image.
2. Bring the image in frequency domain either by using Harr wavelet or by using convolation.
3. Give the upper and lower threshold values to Find the edges and the others are discarded.

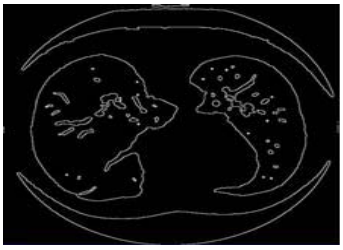

Fig. 3. Edge Detection Output image

VII. Feature Extraction

The Image features Extraction stage is very important in our working in image processing techniques which using algorithms and techniques to detect and isolate various desired portions or shapes (features) of an image. After the segmentation is performed on lung region, the features can be obtained from it and the diagnosis rule can be designed to exactly detect the cancer nodules in the lungs. This diagnosis rules can eliminate the false detection of cancer nodules resulted in segmentation and provides better diagnosis. In the literature we found among the features used in the diagnostic indicators.

- Area of the interest
- Calcification
- Shape and
- Size of nodule
- Contrast Enhancement

Similarly, we experimentally found the above texture features suitable to achieve accurate diagnosis. As a matter of fact, the first feature (the area of the candidate region or object) is used to:

- Eliminate isolated pixels (seen as noise in the segmented image).
- Eliminate very small candidate object (Area is less than a thresholding value).

The use of this feature usually eliminates a good number of extra candidate regions that do not have a chance to form a nodule; furthermore its use tends to reduce the computation time needed in the next diagnostic steps. The 2nd feature is calcification in which we Diffuse, central, laminated or popcorn calcifications are benign patterns of calcification.

These types of calcification are seen in granulomatous disease and hematomas.

All other patterns of calcification should not be regarded as a sign of benignity. The exception to the rule above is when patients are known to have a primary tumor. For instance the diffuse calcification pattern can be seen in patients with osteosarcoma or chondrosarcoma. Similarly the central and popcorn pattern can be seen in patients with GI-tumors and patients who previously had chemotherapy. The 3rd feature is Shape Japanese screening studies showed that a polygonal shape and a three-dimensional ratio > 1.78 was a sign of benignity.

A polygonal shape means that the lesion has multiple facets. A peripheral sub pleural location was also a sign of benignity in this study. The three-dimensional ratio is measured by obtaining the maximal transverse dimension and dividing it by the maximal vertical dimension. A large three-dimensional ratio indicates that the lesion is relatively flat, which is a benign sign. The 4th feature is A solitary pulmonary nodule (SPN) is defined as a single intraparenchymal lesion less than 3 cm in size and not associated with atelectasis or lymphadenopathy.A nodule greater than 3 cm in diameter is called a mass.

This distinction is made, because lesions greater than 3 cm are usually malignant, while smaller lesions can be either benign or malignant. Swensen et al studied the relationship between the size of a SPN and the chance of malignancy in a cohort at high risk for lung cancer. Their findings are listed in the table on the left. They concluded that benign nodule detection rate is high, especially if lesions are small.

Of the over 2000 nodules that were less than 4 mm in size, none was malignant. The 5th feature is Contrast Enhancement is taken. In this less than 15 HU has a very high predictive value for benignity (99%).

After a baseline scan, 4 consecutive scans at 1 minute interval are performed.

This applies only for nodules with the following selection criteria:

1. Nodule > 5mm
2. Relatively spherical
3. Homogeneous, no necrosis, fat or calcification
4. No motion or beam hardening artifacts

VIII. Applying Diagnosis Indicator

For accurate detection of cancerous nodules, we need to differentiate the cancerous nodules from the noncancerous. We developed an artificial neural network to differentiate them. We trained the neural network by the backpropagation algorithm. Relative (or receiver) operation characteristic (ROC) analysis is an analytical procedure for measuring the accuracy of a system. ROC curves show a relationship between the true-positive probability and the false-positive probability. Applying this filter has the effect of decreasing the number of False positives that richly exist in the initial candidate objects. This decision may reduce the computation time needed for the diagnostic rules. By implementing all the above rules, the maximum of regions which does not considered as cancerous nodules are eliminated. The remaining candidate regions are considered as cancerous regions. This CAD system helps in neglecting all the false positive cancer regions and helps in detecting the cancer regions more accurately. [12, 13].

IX. Results And Discussion

For experimentation of the proposed technique, the CT images are obtained from a NIH/NCI Lung Image Database Consortium (LIDC) dataset that provides the chance to do the suggested research. This experimentation data consists of 1000 lung images. Those 1000 lung images are passed to the proposed this system. The diagnosis rules are then generated from those images and these rules are passed to the classifier for the learning process. After learning, a lung image is passed to the proposed system. Then the proposed system will process through its processing steps and finally it will detect whether the supplied lung image is with cancer or not. On one hand, user have developed an automatic CAD system for early detection of lung cancer using Lung CT images in

which a high level of sensitivity has been achieved, with a reasonable amount of false positives per image, (90% sensitivity with 0.05 false positives per image)[14,15] .This prevents the system from hindering the radiologist's diagnosis. On the other hand, the proposed CAD system is capable of detecting lung nodules with diameter ≥ 3 mm, which means that the system is capable of detecting lung nodules when they are in their initial stages. Thus facilitating early diagnosis will improve the patients' survival rate. Discriminate between cancerous and non-cancerous candidate nodules. On one hand, we have developed an automatic system for early identification of lung cancer using chest CT images in which a high level of sensitivity has been achieved, with a reasonable amount of false positives per image, (90% sensitivity with 0.05 false positives per image). This prevents the system from hindering the radiologist's diagnosis. On the other hand, the proposed CAD system is capable of detecting lung nodules with diameter ≥ 3 mm, which means that the system is capable of detecting lung nodules when they are in their initial stages.[18,19] Thus facilitating early diagnosis will improve the patients' survival rate.

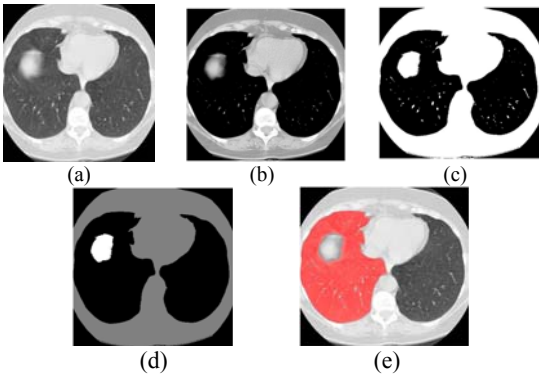

Fig. 4. Segmentation steps: (a) Original, (b) Preprocessing (c) Thresholding (d) Segmenting Lung Region, (e) Showing Cancerous Nodule

X. Conclusion

In this paper, we achieved our purpose in developing an automatic CAD system for early detection of lung cancer by analyzing LUNG CT images using several steps. The approach starts by extracting the lung regions from the CT image using several image processing techniques, including binary image slicing, erosion, resizing image, wiener filter. We introduced the using of binary image slicing technique instead of the thresholding technique that is used in the first step in the extraction process to convert the CT image into a binary image. Binary image slicing technique is both faster and data- and user-independent compared to the thresholding technique. After the extraction step, the extracted lung regions are segmented using Region growing segmentation algorithm. Then, the initial lung candidate nodules resulting from the Region growing segmentation are analyzed to extract a set of texture features to be used in the diagnostic rules. Finally extracted texture features helps to make a comparison between normal and abnormal images. An Artificial Neural Network (ANN) is developed using Back propagation algorithm to differentiate the cancerous nodules from other suspected nodule areas in the X-ray images. This consists of classifying these suspicious regions into positive and negative regions. A positive region is a region that the radiologist feels should go to follow-up for additional information. In The accuracy of 85% approximate the accuracy indicated by surgeons and radiologists for locating cancerous nodules during reading clinical CT images in 2.5–7.0 mm.

References

[1] Singapore Cancer Society, "http://www.singaporecancersociety.org.sg" .
[2] American Cancer Society, "Cancer Statistics, 2005", CA: A Cancer Journal for Clinicians, 55: 10-30, 2005,"http://caonline.amcancersoc.org/cgi/content/full/55/1/10".
[3] B.V. Ginneken, B. M. Romeny and M. A. Viergever, "Computer-aided diagnosis in chest radiography: a survey", IEEE, transactions on medical imaging, vol. 20, no. 12, (2001).
[4] K. Kanazawa, Y. Kawata, N. Niki, H. Satoh, H. Ohmatsu, R. Kakinuma, M. Kaneko, N. Moriyama and K. Eguchi, "Computer-aided diagnosis for pulmonary nodules based on helicalCT images", Compute. Med. Image Graph, vol. 22, no.2(1998),pp. 157-167.
[5] D. Lin and C. Yan, "Lung nodules identification rules extraction with neural fuzzy network", IEEE, Neural Information Processing, vol. 4,(2002).
[6] B. Zhao, G. Gamsu, M. S. Ginsberg, L. Jiang and L. H. Schwartz, "Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm", journal of applied clinical medical physics, vol. 4, (2003).
[7] A. El-Baz, A. A. Farag, PH.D., R. Falk, M.D. and R. L. Rocco,M.D.,"detection, visualization, and identification of lung abnormalities in chest spiral CT scans: phase I", Information Conference on Biomedical Engineering, Egypt (2002).
[8] Linda G. Shapiro and G.C. Stockman., Computer Vision: Theory and Applications. 2001: Prentice Hall
[9] The DICOM Standards Committee. DICOM homepage http: //medical.nema.org/, September 2004.
[10] B.Magesh, P.Vijaylakshmi,M.Abhiram ,"Computer aided Diagnosis System for identification and classification of Lessions in Lungs", International Journal of Computer Trends and Technology- May to June Issue 2011.
[11] Rachid Sammouda,Jamal Abu Hassan1,Mohamed Sammouda,Abdulridha Al-Zuhairy, Hatem abou ElAbbas,"Computer Aided Diagnosis System for Early Detection of Lung Cancer Using Chest Computer Tomography Images",GVIP 05 Conference, 19-21,December 2005, CICC, Cairo, Egypt.
[12] Yamomoto. S, Jiang. H, Matsumoto. M, Tateno.Y, Iinuma. T, Matsumoto. T, "Image processing for computer-aided diagnosis of lung cancer by CT (LSCT)", Proceedings 3rd IEEE Workshop on Applications of Computer Vision, WACV '96, pp: 236 – 241, 1996.
[13] Yeny Yim, Helen Hong and Yeong Gil Shin, "Hybrid lung segmentation in chest CT images for computer-aided diagnosis", Proceedings of 7th International Workshop on Enterprise networking and Computing in Healthcare Industry, HEALTHCOM 2005. pp: 378 – 383, 2005.
[14] Penedo. M. G, Carreira. M. J, Mosquera. A and Cabello. D,"Computer-aided diagnosis: a neuralnetwork- based approach to lung nodule detection", IEEE Transactions on Medical Imaging, vol: 17, pp: 872 – 880, 1998
[15] Kanazawa. K, Kubo. M and Niki. N, "Computer aided diagnosis system for lung cancer based on helical CT images", Proceedings of the 13th International Conference on Pattern Recognition, pp: 381 - 385 vol.3, 1996.
[16] Yamamoto. T, Ukai. Y, Kubo. M, Niki. N, Satou. H and Eguchi. K, "Computer aided diagnosis system with functions to assist comparative reading for lung cancer based on helical CT image", Image Processing, 2000 International Conference on Proceedings, pp: 180 - 183 vol.1, 2000.
[17] Cheran. S. C, Gargano. G, "Computer aided diagnosis for lung CT using artificial life models", Seventh International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2005, 2005.
[18] R. Wiemker, P. Rogalla, T. Blaffert, D. Sifri, O. Hay, Y. Srinivas and R. Truyen "Computer-aided detection (CAD) and volumetry of pulmonary nodules on high- resolution CT data", 2003.
[19] S. G. Armato, M. L. Giger and H. MacMahon, "Automated detection of lung nodules in CT scans: Preliminary results", Med. Phys., Vol.28,2001
[20] Weiner Filtering, " http://www.owlnet.rice.edu".