Detection of Lung Cancer Using SVM with Lung Nodule Segmentation

Manoj M. Mhaske[1], Ramesh R. Manza[2], Pallavi K. Pradhan[3]

[1,2,3]*Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India.*

[1]mhaskemanoj@gmail.com,[2]manzaramesh@gmail.com,[3]pradhanpallavi1987@gmail.com

Abstract: *Now a day's lung disease early detection is essential significance to effective treatment where Computed Tomography (CT) screening is considered as perhaps the satisfactory technique for discovering the early symptoms of lung malignant growth. The fundamental objective of the work is to realize nodule or tumor from CT scan images and classify the whole image dataset as benign or malignant. Early detection and prevention is the only way to deal with lung cancer to keep away from loss of life. Sometimes some human errors can take place in the checking of a long series of CT slices of a single patient manually. This automated system can assist to recognize the condition of sickness at early ranges and diagnose effectively and rapidly on a single click which will useful for radiologists and medical doctors to avoid the serious disease stage. Preprocessing, lungs parenchyma masks segmentation, nodule candidate detection and reduction of False Positives (FP) are the fundamental 4 steps which are included in this proposed Computer-Aided Diagnosis (CAD) system for Lung cancer detection. In the proposed strategy all through the preprocessing step, several masks calculated using the thresholding method and morphological operations. Following, suspicious Regions of Interest (ROI) is calculated the usage of a priori information. During feature extraction, numerous features can be calculated in order to avoid suspicious zones the usage of texture features. Finally, we used SVM classifier to classify the numerical image data as cancerous or not. The proposed system achived 95% accuracy.*

Keywords: ***CAD, computed tomography, lung nodule.***

1. INTRODUCTION

Annually, lung cancer patients diagnosed near about 12 million. Although lung cancers is the second main purpose of cancer compared with other cancer types, it is the most inflicting loss of life of most cancers with 19.39% from all cancers based on the World Health Organization (WHO) information [1]. Lung disease malignancies found around one out of five in men and one out of nine in woomen and it is the second most regular malignant growth. Malignant and benign are the two structures where tumors comes into. Benign tumors are not dangerous, in this manner they don't develop and spread to the degree of destructive tumors. Benign tumors are normally not perilous or life threatening. The Malignant or cancerous tumors are Harmful tumors, which growth can develop and spread to different parts or regions of the human body. Travelling of the disease cells from the underlying tumor site to different pieces of the body is called as metastasis. There are two noteworthy kinds of lung malignant growth: A) Non small cell lung cancer (NSCLC) B) Small cell lung cancer (SCLC) [2]. Lung Cancer have two major types:

A) Non-Small Cell Lung Cancer (NSCLC): Most lung malignancies are delegated Non-Small Cell Lung Cancer. About component of these are squamous cell carcinomas (SCC). SCC, in some cases referred to as epidermoid carcinoma, is regularly common in mens and emerges in the covering of the large air paths, or bronchi. Another regular kind of NSCLC is adenocarcinoma, which happens at the exterior edges of the lung. Non-small cell lung cancer are large cell carcinomas, which as a rule create in the littler bronchi. Non-small cell lung most cancers cells growth that starts at the best factor of the lung here and there spreads to the nerves and veins prompting the arm.

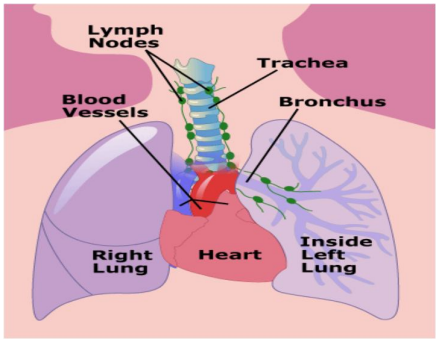


Figure 1. Anatomy of lung

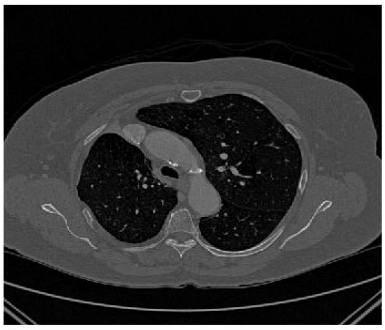


Figure 2. Original lung CT image

Every one of the three subtypes of NSCLC grows in an unexpected way. Treatment is frequently founded at area of specific malignancy and its pace of spread. Squamous cell or epidermoid carcinomas, for the most part, happen in the bronchi in the focal point of the lungs, the cancer cells typically double every 180 days.

B) Small Cell Lung Cancer (SCLC): About one in 4 malignancies involving the lungs are diagnosed as small cell lung cancer (SCLC). There are several sorts of SCLC or oat cell cancer, including a mix of small cell and different cell types. These cancers grow swiftly doubling in cell variety about each 30 days and unfold shortly to lymph nodes and different organs than the non-small cell type [3].

2. Literature Survey

An automated lung cancer detection system development work has been carried out by number of researches. Sayani Nandy and Nikita Pandey in [4] proposed an strategy for the detection of cancer cells from Lung CT scan images. This work provides a method to detect the most cancers cells from the CT scan image. Prof. Samir Kumar B. Was developed a system using Computer-Aided Diagnosis (CAD) for discovering the edges from CT scan images of the lung for the detection of ailments [5]. Thresholding algorithm [6] offers filtering to notice the sputum cell from the raw image for early detection via Fatm Taher et al. M. Tan et.al [7] proposed CADe (Computer-Aided Detection) system in thier work to classify nodules or non-nodules by means of genetic algorithms and Artificial Neural Network; with total of 360 nodules of 3-30 mm in diameter of 134 sufferers enrolled in LIDC society. This CADe system received 87.5% sensitivity with four FP (false positives) per scan. In [8], 420 CT scans of 420 patients, with 3-30 mm in diameter of 379 possibility of malignancy, are randomly chosen from the LIDC database, and the probable malignant nodules are categorized by SVM classifier. This system obtained 97% accuracy for the segmentation stage, 94.4% sensitivity with 7.04 FP per scan for the CADe system, and 93.9 p.c sensitivity with 7.21 FP per case for the classification stage.

Amjed et al. [9] designed an computerized intelligent system for nodule detection and classification of lung most cancers in CT images. In their work, they utilized geometrical facets and morphological image processing techniques. In [10] by way of C. Panyindee and W. Chiracharit, Watershed transform and image mapping were applied to detect the lung nodules existing in PET/CT images. Few methods used Haralick texture features for feature extraction from lung CT image [11, 12] with more than a few classification methods. Balaji et al. used a selective scale based image filtration for texture quantification [11]. Mir Rayat et al. selected prominent facets by using backward search algorithm and Chi-square distance measure was once used to classify the tumor in CT images [12]. Tidke et al. developed a CAD machine for early detection of lung most cancers nodules from the chest computer tomography images. Textural elements extracted from the lung nodules the use of the gray level co-occurrence matrix [13]. Pandy et al. introduced a novel approach of cancerous cell detection from lung CT scan images with the aid of the usage of the Sobel part detection technique [14] in those work. Sudha et al. segmented the lung area in their proposed system with use of thresholding and morphological operations [15]. Ada et al. estimated and detected lung cancer survival the usage of neural community classifiers [16]. In this, they have used histogram equalization for the preprocessing of images. Feature extraction manner and neural community classifier used to take a look at the

country of the patient in its early stage whether or not it is everyday or abnormal. Sankar et al. [17] increased structure for lung cancer cell identification using Gabor filter and smart system.

3. Proposed Methodology

We have proposed an system for Lung Cancer Detection to discover lung cancer disease by means of the use of the SVM Classifier technique. The primary purpose of the proposed system is to automatically classify the lung CT images as normal or abnormal. The input CT images are taken from the TCIA database which is given as input to the pre-processing. After pre-processing, the features are extracted. SVM is used to classify the CT images are normal or abnormal. Abnormal images are recognized through the extracted features such as the images contains lung nodule candidate's problems. The Proposed system shown in Figure 3.

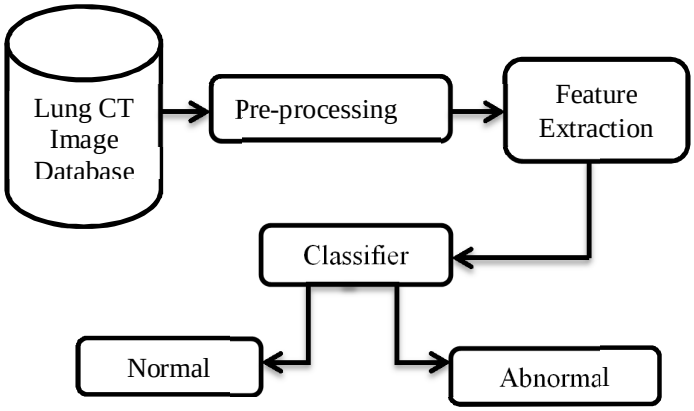


Figure 3. Block diagram of the system

We used the database for the proposed system is bought from The Cancer Imaging Archive (TCIA) which is sponsored through the SPIE, NCI/NIH, AAPM and The University of Chicago. During assessment work we used SPIE-AAPM Lung CT Challenge dataset. The dataset having 70 thoracic CT scans DICOM images and out of which 10 used for a calibration and 60 used for test, spatial coordinates of the nodule locations and the diagnosis for every nodule in the calibration and test datasets. Dimensions of whole CT images dataset are 512 x 512 pixels, with a bit depth of 12 bits [18-19].

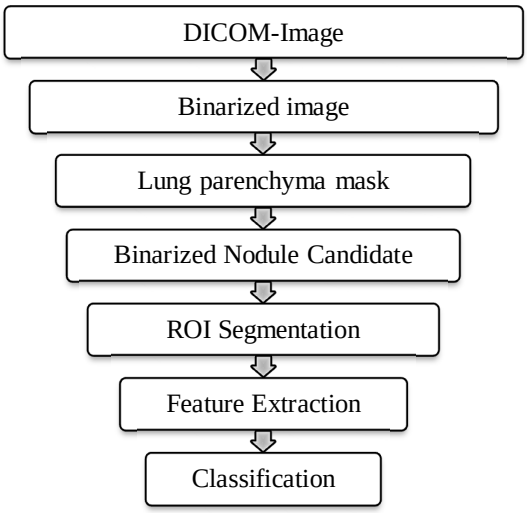


Figure 4. Flow of methods

List of methods are shown in above figure 4 which is used for implementation of our proposed algorithm to diagnose the Lung Cancer. Firstly we have collected the lung CT image database then

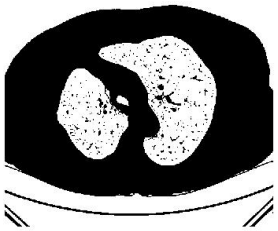
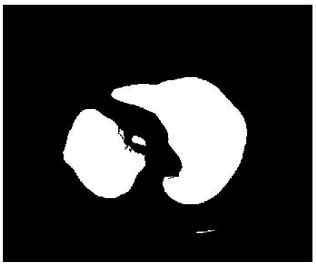
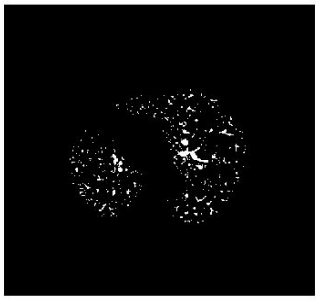
don the binarization to all CT images to enhance the quality of image then separated the lung parenchyma mask and binarized again the nodule candidates and after that finalize the region of interest to segment the image. Finally obtained the texture features for classification using SVM.

4. Experimental Analysis

4.1 Preprocessing

İn preprocessing step firstly we read whole database of lung CT images as input gray level. Then binarized the image and separated lung parenchyma mask. İn next step we obtained binarized lung nodule candidates and finally after FP reduction we got segmented region of intrest (ROI) for furter process shown in table 1. According to provided statistics via SPIE-AAPM lung CT challenge database, Lung-RADS system and [18], FP reduction is initialized. After that, the binary masks of nodule candidates is acquired. Then Employing nodule candidates masks to the unique CT image then got lung ROI candidates shown in table 1.

Table 1. Preprocessing

<table>
<tr>
<td> 
a) Binarization of image</td>
<td> 
b) Lung parenchyma mask separation</td>
</tr>
<tr>
<td> 
c) Obtaining binarized lung nodule candidates</td>
<td> 
d) Segmented ROIs</td>
</tr>
</table>

4.2 Feature Extraction

Features are used to signify relevant data for fixing the computational work in an application. There exist numerous kinds of features. For the texture evaluation, Haralick et al. [20] proposed 14 texture quantities derived from the GLCM matrices. These measures signify the gray degree variants that are related to the smoothness, regularity, heterogeneity and the contrast of an image. From the whole amount of 14 features, we advise to use ten features for feature extraction and extracted samples are shown in below table 2.

Table 2. Feature Extraction

Features	Image1	Image2	Image3	Image4	Image5
Contrast	0.504171	0.304227	0.367909	0.48832	0.440489
Correlation	0.103223	0.158107	0.123121	0.132907	0.170377

Energy	0.883869	0.801547	0.831784	0.887011	0.883288
Mean	0.005905	0.004103	0.004139	0.005831	0.006279
Entropy	1.83305	2.956706	2.44402	1.338888	1.557835
Variance	0.008059	0.008047	0.008053	0.00807	0.008063
Smoothness	0.956459	0.938508	0.939012	0.955927	0.958944
Kurtosis	47.33757	12.3938	27.27931	47.37886	39.14829
Skewness	4.466542	1.171153	2.271823	4.320737	3.396371
IDM	6.573035	0.649064	1.400459	3.750474	3.654014

The following are the definitions of the selected 10 texture features used for the proposed system:

1. **Contrast:** Contrast used to measures the local variations in the GLCM. It calculates intensity difference between image elements and its neighbor image elements for a whole image.

2. **Correlation:** It is used to measure the joint probability occurrence of the required image element pairs.

3. **Energy:** It provides the total of square components within the GLCM. It is conjointly referred to as uniformity or the angular moment.

4. **Mean:** It is used to calculate Average or mean value of array for a random variable vector A made up of N scalar observations.

5. **Entropy:** Entropy is a statistical measure of randomness that can be used to characterize the texture of the input images.

6. **Variance:** It is used to returns the variance of the elements of A along the first array dimension whose size does not equal to 1.

7. **Smoothness:** Smoothness is a measure of relative smoothness of intensity in a region.

8. **Kurtosis:** Kurtosis is the measure of the outliers of a data or distribution.

9. **Skewness:** Skewness is the measure of symmetry or asymmetry of a distribution or dataset around the sample mean.

10. **IDM** (Inverse Difference Moment): It is the local homogeneity and it high when local gray level is uniform and inverse GLCM is high.

4.3 Classification

The SVM classifier is the proposed classifier which is used in this work. The designed system uses the above algorithm that estimates a characteristic to classify the data in two classes [21]. A Support Vector Machine (SVM) is a discriminative classifier officially characterized through an isolating hyperplane. As it were, given named getting ready information (supervised learning), the algorithm outputs an most effective hyperplane which orders new models. In two dimentional space, this hyperplane is a line partitioning a plane in two sections have been in every class lay on either side. On feeding numerous information set of CT scan Images along with the studied parameters to the SVM classifier helps the module to train and classify the distinction between the cancerous and non-cancerous tissue. After the preprocessing and feature extraction from lung CT images, efficiently classification the experimental effects and discussion affords the Sensitivity, Specificity and Accuracy percentages got in the proposed framework [22]. To analyze data and recognize patterns for classification purposes, supervised learning models with associated studying algorithms are used in this paper regarded as support vector machine (SVM) which is a derivative of a computer studying primarily based neural network and appreciably utilized in pattern recognition and regression problems.

Specificity: Specificity estimates the extent of negatives which are effectively distinguished in that capacity. Formula of Specificity is shown below, where TN – True negative and FP – False Positives.

$$\text{Specificity} = TN/TN+FN \quad (1)$$

Sensitivity: İt is also called as genuine positive rate or the review rate in certain fields. It quantifies the extent of genuine positives which are effectively recognized. Formula is shown below, where TP – True Positive and FN –False Negative.

$$\text{Sensitivity} = TP/TP+FN \quad (2)$$

Accuracy: İt is used to represent the ratio of the CT images that are classified correctly.

$$\text{Accuracy} = TP+TN/TP+TN+FP+FN \quad (3)$$

Table 3. Specificity, Sensitivity and Accuracy of proposed system

Image Classifier	True Positive	True Negative	False Positive	False Negative
SVM	33	34	1	2
Result: Specificity-97.14%		Sensitivity-94.28%		Accuracy-95.71%

The final result for this paper is obtained by classification which is carried out by using the SVM classifier and input 70 lung CT scan images given to the SVM classifier where it is classified into two classes. Class 1 is Malignant and class 0 is Non Malignant.

Conclusions

The proposed Computer aided CAD system for segmentation and classification of lung nodules in computer tomography CT scan images can reduce the human errors and gives the quick and correct diagnose of lung cancer. It can pick out lung nodule parenchymal tissue accurately. It has the ability to pick the suspicious zones efficiently and can minimize in a suited way FPs. Quality criterions of the designed system show a proper adequate overall performance and can distinguish between benign and malignant nodules. This automated system can help to radiologists and doctors to avoid the serious disease stage by knowing the condition of disease at early stages also it diagnose correctly and quickly on single click which will few time consuming.

The proposed approach typically focuses on the MATLAB app which has been developed for the detection, segmentation and classification of lung tumors from CT images within a quick span of time based on feature extraction and resolution via the PCA algorithm with suitable aspects selected were classified by using SVM classifier and finally the proposed system obtained 95% accuracy.

Acknowledgement

Database obtained from The Cancer Imaging Archive (TCIA) sponsored by the SPIE, NCI/NIH, AAPM and The University of Chicago for this research work. Also, My guide Dr. R.R. Manza, Dept. of Computer Science and Information Technology, Dr. BAMU Aurangabad.

REFERENCES

[1] World Health Organization, “Cancer, fact sheet no. 297”, 2011.
[2] SEER Training Modules, Module Name. U. S. National Institutes of Health, National Cancer Institute. Day Month Year (of access) <https://training.seer.cancer.gov/>.
[3] Elmar Rendon-Gonzalez, Volodymyr Ponomaryov, “Automatic Lung Nodule Segmentation and Classification in CT Images Based on SVM”, IEEE 978-1-5090-2267, 24 June 2016.
[4] Sayani Nandy, Nikita Pandey, “A Novel Approach of Cancerous Cells Detection from Lungs CT Scan Images’’, International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 8, August 2012 .
[5] Prof. Samir Kumar Bandyopadhyay, “Edge Detection from Ct Images of Lung’’, International Journal of Engineering Science & Advanced Technology, Volume - 2, Issue - 1, 34 – 37.
[6] FatmTaher, NaoufelWerghi and Hussain Al-Ahmad, “Extraction of Sputum Cells using Thresholding Techniques for Lung Cancer Detection”, 2012 International Conference on Innovations in Information Technology.
[7] M. Tan, R. Deklerck, B. Jansen, M. Bister, J. Cornilis, “A novel computeraided lung nodule detection system for CT images”, Medical Physics, October 2011, doi: 10.1118/1.3633941[PubMed Open Access].
[8] M. Firmino, G. Angelo, H. Morais, M. R. Dantas and R. Valentim, “Computer-aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy”, Journal of Negative Results.
[9] Ajmed S. Al-Fehoum, Eslam B.Jaber, Mohammed A. Al-Jarrah, “Automated detection of lung cancer using statistical and morphological image processing techniques”, Journal of Biomedical Graphics and Computing, 4(2), pp. 33 - 42, 2014.
[10] C. Panyindee and W. Chiracharit, “ Detection of Lung Tumors in Chest PET/CT Images by using Watershed Transform and Image Mapping”, The 3rd International Symposium on Biomedical Engineering (ISBME 2008), pp. 9 – 12, 2008.
[11] Balaji Ganeshan, Sandra Abaleke, Rupert C.D. Young, Christopher R Chatwin and Kenneth A. Miles, “Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage”, Cancer Imaging, 10(1), pp. 137-143, 2010.
[12] Mir Rayat Imtiaz Hossain, Imran Ahmed, Md. Hasanul Kabir, “Automatic Lung Cancer Detection using GLCM features”, Asian Conference on Computer Vision ACCV14, Singapore, 2014.
[13] Ms. Swati P. Tidke, Prof. Vrishali A. Chakkarwar, “Classification of Lung Tumor Using SVM”, International Journal of Computational Engineering Research (ijceronline.com) Vol. 2 Issue. 5 September 2012.
[14] Nikita Pandey, Sayani Nandy, “A Novel Approach of Cancerous Cells Detection from Lungs CT scan Images”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 8, August 2012.
[15] Sudha.V, Jayashree.P, “Lung Nodule Detection in CT Images Using Thresholding and Morphological Operations”, International Journal of Emerging Science and Engineering (IJESE) ISSN: 2319–6378, Volume-1, Issue-2, December 2012.
[16] Ada, Rajneet Kaur, “Early Detection and Prediction of Lung Cancer Survival using Neural Network Classifier”, IJAIEM Volume 2, Issue 6, June 2013.
[17] K. Sankar, Dr. M. Prabhakaran, “An Improved Architecture for Lung Cancer Cell Identification Using Gabor Filter and Intelligence System”, International Journal of Engineering and Science (IJES) 2013.
[18] D. Cascio, R. Magro, F. Fauci, M. Iacomi, and G. Raso, “Automatic detection of lung nodules in CT datasets based on stable 3D mass-spring models”, Comput. Biol. Med., vol. 42, no. 11, pp. 1098–1109, 2012.
[19] Public Database Open Access: https://public.cancerimaingachive.net/ ncia/searcMain.jsf;jssionid= 128369143ADF4D948378665461898B03.
[20] R. M. Haralick, K. Shanmugam, I. Dinstein, “Textural Features for Image Classification”, IEEE Trans. Syst.Man. Cybern, vol. 3, no. 6, pp. 610–621, Nov. 1973.
[21] M. Alilou, V. Kovalev, E. Snezhko, and V. Taimouri, “a Comprehensive Framework for Automatic Detection of Pulmonary Nodules in Lung Ct Images”, Image Anal. Stereol, vol. 33, no. 1, p. 13, 2014.
[22] MM Mhaske, RR Manza, YM Rajput, “Detection of Retinal Images as Normal or Abnormal using Texture Feature Analysis to identify the Diabetic Retinopathy”, ISSN No: 2309-4893 International Journal of Advanced Engineering and Global Technology Vol-03, Issue-03,March 2015.
[23] https://shodhganga.inflibnet.ac.in/bitstream/10603/10111/9/09_chapter%203.pdf.
[24] Manoj M Mhaske, Ramesh R Manza, Yogesh M Rajput and B P Gaikwad, “Analysis and Detection of Diabetic Retinopathy using Features Extraction Techniques”, IJCA Proceedings on National conference on Digital Image and Signal Processing DISP 2015(1):28-31, April 2015.