

Detect the Stage Wise Lung Nodule for CT Images Using SVM

Ganesh Jadhav¹, Prof. Anita Mahajan²

Department of Computer Engineering, Dr. D. Y. Patil School of Engineering and Technology,
Lohegaon, Pune, India¹

Department of Computer Engineering, Dr. D. Y. Patil School of Engineering and Technology,
Lohegaon, Pune, India²

ABSTRACT: Lung cancer is a disorder that occurs due to the uncontrolled cell growth in tissues of the lung. It is extremely difficult to detect it in its early phases as its signs appear just in the advanced phases. The goal is to automate the classification process for the early detection of Lung Cancer. It provides classification algorithm i.e. Support Vector Machine and optimization is used. Evaluation would be done on the basis of correctly classified trial data. By using computerized tomography (CT) images, a watershed scheme used to section lung tumours and computed tumour-related image attributes. The measures for detection of lung cancer start with process of accepting CT Images. These CT images are further processed for utilizing instruction and testing methods attributes which are classified using (SVM) Support Vector Machine. This classification helps in evaluating the results of the input CT image. Here we implement the system which can find lung nodule and display its stage by computing its size.

KEYWORDS: computed tomography CT, SVM Support Vector Model, Lung cancer, watershed scheme.

I. INTRODUCTION

Medical images are essential part of medical analysis and treatment. These images are different from typical photographic images mainly because they show internal human body as opposed to an image of surfaces. They include both projection x-ray images and cross-sectional images, such as those acquired through computerized tomography (CT). Medical image processing is a branch of image processing that handles such images. It is driven both from the particular CT images and by the medical applications which make them useful. Medical images include a wealth of hidden information that can be used by physicians in making reasoned decisions in regards to a patient. However, extracting this useful hidden information is a critical initial step to their use. Although early cancer detection and treatment can improve the survival rate of carcinoma patients, lung cancer constants speeds once surgical operation of the malignant tumors can differ from half hour to hour within the previous studies. The process to be used depends upon the type and complexity of image to be mined. For example, it is simpler to mine 2-dimensional x-rays as compared to 3- dimensional CT scans. However, some processes are essential to the task of medical image mining and processing, regardless of the complexity of image.

II. REVIEW OF LITERATURE

In [2], Krishnaiah, Dr.G.Narsimha, Dr.N.SubhashChandra the authors said that Cancer is the most important cause of death for both men and women. The early detection of cancer can be useful in treating the disease completely. So that the requirement the techniques to detect the occurrence of cancer nodule in early stage is improving. A disease that is normally Misdiagnosed is lung cancer. Earlier identification of Lung Cancer saves tremendous lives, failing which may lead to other serious issues causing unexpected critical ending.

Advantages: In this, the performance of the novel system is based on algorithms and classifiers.

Limitation: This system is time consuming it can not detect stage of the cancerous nodule.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

In [3] the ZakariaSulimanZubi, RemaAsheibaniSaadAuthors mentioned that Medical images mining is a bright area of computational intelligence applied to automatically analyzing patient's records aiming in the discovery of new knowledge possibly helpful for medical decision making. Detection of lung cancer in early stage is the key of its treatment. In Generally, a measure for early stage lung cancer analysis mainly includes those utilizing X-ray chest pictures, CT, MRI, etc.

Advantages: System is completely decentralized, fault-resilient, scalable, and reliably.

Limitations: This system detects cancerous nodule but it doesn't calculate its Size as well as stage. Because complexity and infeasible algorithm its not suitable to implement.

In [4], author ZakariaSulimanZubi and emaAsheibaniSaadstated that Lung cancer is a disease of uncontrolled cell development in tissues of the lung, Lung cancer is one of the most common and deadly diseases in the world Detection of lung cancer in early stages the key of its treatment. In Generally, a measure for early phase lung cancer analysis mainly includes those utilizing X-ray chest pictures.

Advantages: Attractive features of system include its simplicity, provable correctness, and provable functionality

Limitation: Due to high features included this system required more computation and hardware cost in terms of storage. It does not work on CT images perform only on chest MRI.

In [5], G. Liu, Z. Lin, and Y. Yu propose a new supervised strategy which combines discretization and feature selection to select the most relevant attributes which can be used for classification function. The classification technique to be used is Associative Classifiers. The features used are texture features extracted from MRI Images.

Advantage: This system allows designers to achieve higher performance for a given lung images.

Limitation: This system have more sophistication in maintaining the features of database for each Image. If any updating is done in database then system need to update the whole database table additionally.

In [6] PaolaCampadelli, Elena Casiraghi, and Diana Artioli, present a system processing electronic postero-anterior (PA) chest radiographs, that starts by creating an accurate segmentation of the lung field place. The segmented lung area includes eventhose parts of the lungs hidden behind the heart, the backbone, as well as the diaphragm, which are usually excluded from the processes presented in the literature.

Advantages: System is completely decentralized, fault-resilient, scalable, and reliably.

Limitations: It requires more training set storage cost as well as fault tolerance is challenging in real time implementation which result in less efficiency and poor quality.

III.SYSTEM OVERVIEW

This paper represents A Computer assisted system to detect the stage wise lung nodule for medical images. In this system we are using some techniques which are essential to the task of medical image processing, Lung Field Segmentation, Data Processing, Feature Extraction, Classification using SVMs. The techniques used in this paper states to classify CT lung images as normal and abnormal. Different learning experiments were performed on two different data sets, created through feature selection and SVMs trained with different parameters; the results are compared and reported.

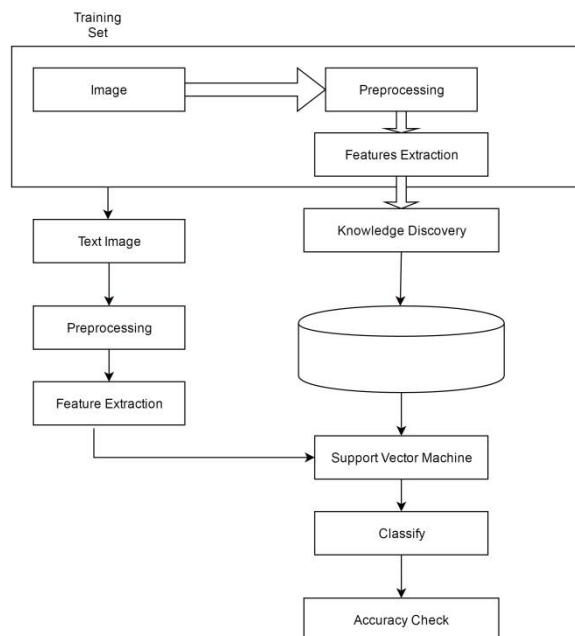


Fig.1 System Architecture

In our system we first handle the training set of CT image. Pre-processing task is completed on those images which consist of image grey scaling and Image binarisation by using thresholding technique. After that system implements the process of segmentation of the image by using watershed algorithm. Here system get the exact cancerous nodule part on which attribute extraction is performed and maintain it as knowledge discovery in database. For testing we give the test image to the system which the above process is performed. Here system utilizes SVM i.e. support vector machine for the classification of the images. After classifying the final outputted results shown in the form of size of cancerous nodule and its own stage using its greater accuracy.

IV.ALGORITHM

1. GLCM

The basic GLCM algorithm is as follow: 1. Count all pairs of pixels in which the first pixel has a value i, and its matching pair displaced from the first pixel by d has a value of j. 2. This count is entered in the ith row and jth column of the matrix Pd[i,j] 3. Note that Pd[i,j] is not symmetric, since the number of pairs of pixels having gray levels[i,j]does not necessarily equal the number of pixel pairs having gray levels [j,i]. 4.The elements of Pd[i,j]can be normalized by dividing each entry by the total number of pixel pairs. 5. Normalized GLCM N[i,j], defined by:

$$N[i,j] = \frac{P[i,j]}{\sum_i \sum_j P[i,j]}$$

2. Watershed

Segmentation using the watershed transform works better if you can identify, or "mark," foreground objects and background locations. Marker-controlled watershed segmentation follows this basic procedure:

1. Compute a segmentation function. This is an image whose dark regions are the objects you are trying to segment.
2. Compute foreground markers. These are connected blobs of pixels within each of the objects.
3. Compute background markers. These are pixels that are not part of any object.
4. Modify the segmentation function so that it only has minima at the foreground and background marker locations.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

5. Compute the watershed transform of the modified segmentation function.

3. SVM

Support vector machines are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification. The basic SVM takes a set of input data and for each given input, predicts, which of two classes forms the input, making it a non-probabilistic binary linear classifier. From given set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. In the proposed method we are using linear classifier. Best hyper plane is the one that represents the largest separation or margin between the two classes. So we choose the hyper plane so that the distance from it to the nearest data point on each side is maximized. If such a hyper plane exists, it is known as the maximum margin hyper plane and the linear classifier it defines is known as a maximum classifier, which is shown in below fig.

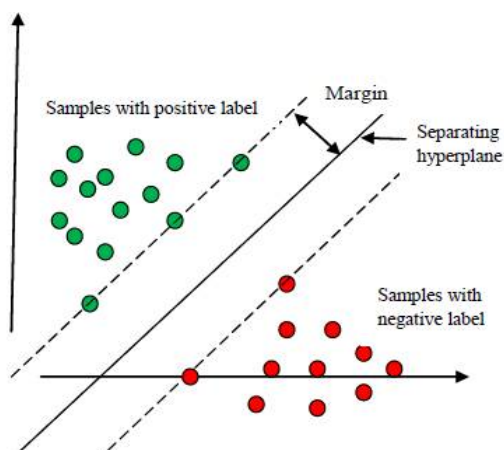


Fig.2 Maximum margin classifier

V. MATHEMATICAL MODEL

Input

Let S is the Whole System Consist of

$$S = \{I, P, O\}$$

Where,

I = input.

$$I = \{U, Q\}$$

U = User

$$U = \{u_1, u_2, \dots, u_n\}$$

Q = Query

$$Q = \{q_1, q_2, \dots, q_n\}$$

P = Process

$$P = \{WA, GLCOMA, SVM\}$$

WA = Watershed Algorithm.

GLCOMA = Grey level co-occurrence matrix algorithm.

SVM = Support Vector Machine.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

$$MSE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [X(i, j) - X_c(i, j)]^2 \dots (1)$$

Where, X (i, j) = original image.
X_c (i, j) = compressed image.

$$PSNR = 10 \log_{10} ((255^2) / MSE)$$

Output: The predicted result will be the output of the system

Advantages:

- Utilization of time management.
- Speedy process.
- Issue of maintaining privacy can be solved.
- High security.
- Data is highly safe.

Applications:

In hospitals and pathological laboratories.

V. SYSTEM ANALYSIS

When the images were given to the SVM classifier for testing, we obtained the following outputs.

Table 3: SVM outputs and accuracy parameters

Qtr	Experimental Results	SVM Output	Parameters
1st Qtr	Benign	Class 0	TP
2nd Qtr	Malignant	Class 1	FP
3rd Qtr	Malignant	Class 1	TN
4th Qtr	Benign	Class 0	FN

Where,

TP- predicts benign as benign.

FP- predicts benign as malignant.

TN- predicts malignant as malignant.

FN- predicts malignant as benign.

Performance measures:

The following parameters are calculated on the basis of the results obtained for all the images in database.

1) Accuracy (AC) = (TP+TN) / (TP+TN+FP+FN) = (25 + 18) / (45) = 0.9556 Accuracy in % = 95.56%

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

- 2) Sensitivity (SE) = $TP / (TP+FN) = 25 / (25+ 1) = 0.9615$
 3) Sensitivity in % = 96.15% 3) Specificity (SP) = $TN / (TN+FP) = 18 / (18 + 1) = 0.9473$ Specificity in % = 94.73%

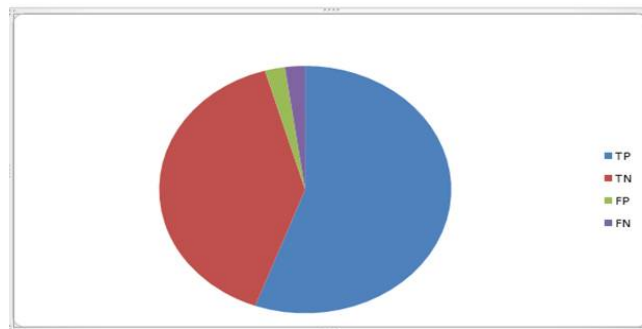


Fig.2 Graphical representation in Bar Graph

Thus, from the results we found that our proposed system has achieved 95.56% accuracy, specificity of 94.73% and better sensitivity i.e. 96.15% that means our system has higher accuracy for the classification of benign type of tumor as compared to malignant tumor.

Early detection of cancer can be useful in treating the disease completely. So that the requirement of techniques to detect the occurrence of cancer nodule in early stage is increasing. For experiment of the proposed technique, the CT images are obtained from a Lung Image Database Consortium (LIDC) dataset that provides the opportunity to do the recommended research. This experiment data consists of 130 lung images. Those 130 lung images are handed to the proposed this system. proposed system will process through its processing steps and finally it will detect whether the provided lung image is cancer or maybe not and if it is having cancer then its size as well as stage also displayed.

Table 1

Accuracy	Sensitivity
75	83.8
75.8	84.2
73	80.7
76	84.2
72.5	80.1

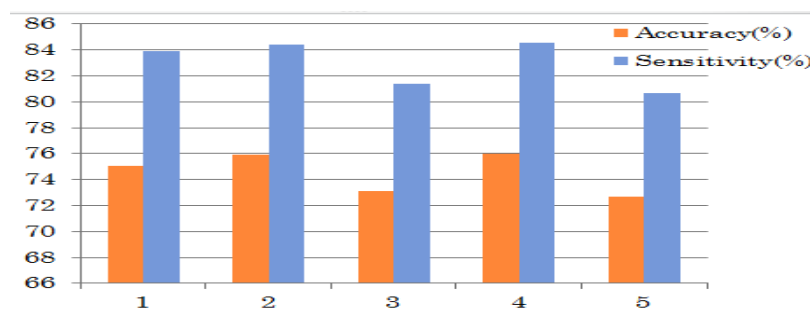


Fig.3 Graphical representation in Bar Graph

VI. CONCLUSION

Different phases of image processing were implemented on Lung Nodules. From these different image processing techniques, will provide the efficient de-noising. Segmentation completed by a watershed algorithm, gives different

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

area of image. In this paper, we presented a SVM classifier system for classifying lung nodules for the specified database. Subset Evaluator is used to extract the different features of image and which takes less time for creating the result. These results are passed through SVM classifier which provides 95% accuracy.

ACKNOWLEDGEMENT

I would prefer to give thanks the researchers likewise publishers for creating their resources available. Additionally, I appreciate to commentator for their vital recommendations moreover I'm conjointly grateful to reviewer for their valuable suggestions and also thank the college authorities for providing the required infrastructure and support.

REFERENCES

- [1]NastaranEmaminejad, Wei Qian, Yubao Guan, Maxine Tan, YuchenQiu, Hong Liu, and Bin Zheng. "Fusion of Quantitative Image and Genomic Biomarkers to Improve Prognosis Assessment of Early Stage Lung Cancer Patients" DOI 10.1109/TBME.2015.2477688, IEEE Transactions on Biomedical Engineering
- [2] Krishnaiah, Dr.G.Narsimha, Dr.N.Subhash Chandra. "Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques" V. Krishnaiah et al. / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (1) , 2013, 39 – 45
- [3] Maxine Tan, Rudi Deklerck and Bart Jansen, Michel Bister and Jan Cornelis "A novel computer-aided lung nodule detection system for CT images" IEEE Transactions On Biomedical Engineering, Vol. 58, No. 6, June 2011
- [4]ZakariaSulimanZubiandemaAsheibaniSaad "Using Some Data Mining Techniques for Early Diagnosis of Lung Cancer" Recent Researches in Artificial Intelligence, Knowledge Engineering and Data Bases ISBN: 978-960-474-273-8.
- [5] G. Liu, Z. Lin, and Y. Yu "An Approach for Discretization and Feature Selection Of Continuous-Valued Attributes in Medical Images for Classification Learning" International Journal of Computer and Electrical Engineering, Vol. 1, No. 2, June 2009 1793-8163.
- [6] Paola Campadelli, Elena Casiraghi, and Diana Artioli "A Fully Automated Method for Lung Nodule Detection FromPostero-Anterior Chest Radiographs" IEEE Transactions on Medical Imaging, Vol. 25, No. 12, December 2006.
- [7] S. Singhal., et al., "Prognostic implications of cell cycle, apoptosis, and angiogenesis biomarkers in non-small cell lung cancer: a review", Clin. Cancer Res., vol. 11, pp. 3974-3986, 2005.
- [8] G. Bepler, et al., "RRM1 and PTEN as prognostic parameters for overall and disease-free survival in patients with non-small-cell lung cancer," J. Clin.Oncol.,vol 22, no. 10, pp. 1878-1885, May. 2004.
- [9] E. P. Diamandis, "Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations," Mol Cell Proteomics, vol. 3, no. 4, pp. 367-378, Apr. 2004.
- [10] C. Poleri, et al., "Risk of recurrence in patients with surgically resected stage I non-small cell lung carcinoma: histo-pathologic and immune-histo-chemical analysis," Chest vol. 123, no. 6, pp. 1858-1867, Jun. 2003.