

# AUTOMATIC DETECTION OF LUNG CANCER IN CT IMAGES

G. Vijaya<sup>1</sup>, A. Suhasini<sup>2</sup>, R. Priya<sup>3</sup>

<sup>1</sup>Research Scholar, Dept. of Computer Science and Engineering, Annamalai University, Chidambaram

<sup>2</sup>Associate Prof., Dept. of Computer Science and Engineering, Annamalai University, Chidambaram

<sup>3</sup>M.E. Student, Dept. of Computer Science and Engineering, Annamalai University, Chidambaram

## Abstract

Lung cancer is the most critical reason for death. To enhance cancer detection the radiologists using distinctive scans and X-ray's. Consequently, we use CT scan images for inspecting the interiors of the body. An automatic cancer detection system proposed to distinguish cancerous tumor from the CT scan images. The cancer detection scheme consists of four stages. They are preprocessing, segmentation, feature extraction and classification. These four levels are used in image processing to enhance the tumor identification precision. The final outcome of this paper is to find cancer tumor as benign (or) malignant.

**Keywords**— Computed Tomography, Data mining classification, Feature extraction, Lung cancer.

\*\*\*

## 1. INTRODUCTION

Lung cancer is the most cause of cancer death in the worldwide. Early detection and treatment of lung cancer is exceptionally vital. Uncontrolled cell growth causes cancer. Lung cancer occurs for out-of-control cell growth in one or both lungs. Cigarette smoking is the most critical reason for lung cancer; other factors such as environment pollution, mainly air, and excessive alcohol may also be contributing to lung cancer. These atypical masses are called tumors. Tumors are either non-cancerous (benign) or cancerous (malignant).

Ankit Agrawal[1] developed a lung cancer outcome calculator; it surveys the cancer level based on the patient report. The paper [2], serves the second opinion to the specialists in improving the fastness and accuracy of cancer detection. The association rule mining technique is used to find hot spots in lung cancer datasets. The datasets are having different nominal and numeric attributes [3]. Virtual dual energy (VDE) radiography images are used to detect cancer tumor, using this VDE technology, cut the overlapped ribs and clavicles between the tumors [4]. Chest x-ray images are used to diagnose lung diseases. Hence, Gaussian based matching, local binary pattern and gradient oriented features are used to diagnose lung cancer [5]. Breath hold CT image sequences are used to estimate the lung air volume and its variations in an image sequence [6].

Lung cancer risk prediction system for predicting the risk levels of the patient. It uses k-means clustering algorithm for identifying relevant and non-relevant data. This risk prediction system should helpful in the detection of a person's predisposition for lung cancer [7]. The texture features are extracted using gray level co-occurrence matrix. For classification, SVM classifier is used [8]. Canny edge detection and flood-fill algorithms are applied to find the tumors and rule based technique is applied to classify the cancer nodules [9]. Classification of

the lung cancer by artificial neural network is depicted in [10]. Positron emission tomography (PET) images are taken and active shape model has been used to segment the tumor in [11]. Multilevel thresholding approach and data fusion techniques are used to segment medical color image segmentation [12]. The electronic nose detection system takes the small print of cancer affected patients and it checks the variations in the smell print [13]. Data mining is useful in lung cancer classification and Ant colony optimization algorithm helps in increasing or decreasing the disease prediction [14]. Cancer gene expressions are used, to detect cancer; decision rule and ensemble learning algorithms are used for classification in [15].

Our work in this paper presents an automatic cancer detection system to find the lung cancer tumors using the lung CT (computed tomography) images. The lung cancer detection scheme consists of four stages. They are preprocessing, segmentation, feature extraction, classification. Preprocessing is the initial stage in order to remove the unwanted noise present in the original image using median filter. The segmentation is the second stage. This stage is used to identify the cancer tumor from lung CT images using edge detection and boundary tracing. A third stage is feature extraction, this stage is used to find the size of the tumor based on the area, perimeter, and irregularity index. The final stage is classification stage, this stage is to separate the tumor as benign (or) malignant. To classify the lung cancer, by using the data mining, classification techniques like SMO (Sequential Minimal Optimization), J48 decision tree, Naive Bayes. Once the classification is performed, we have to compare the experimental results of the above classification techniques, and determine which one gives accurate and correct answers.

Data mining is the non-trivial extraction of implicit, previously unknown and potentially useful information about data.

Data mining, classification techniques is used in this project, sequential minimal optimization is often slow to converge to a solution-especially when the data is not linearly separable in the space spanned by the nonlinear mapping. Naive Bayes implements the probabilistic Naive Bayesian classifier. This can improve performance if the normality assumption is grossly incorrect. The naive Bayes algorithm is based on the probability. An advantage of Naive Bayes is that it only takes a small amount of training data to estimate the parameters necessary for classification. J48 algorithm is based on decision tree. The J48 decision tree algorithm identifies the attributes that must be used to separate the tree further based on the notion of information. The logit boost algorithm is an implementation of additive logistic regression which performs classification using a regression scheme as the base learner, and can handle multi-class problems.

Types of tumors

- Benign
- Malignant

### Benign

If the tumor is benign, then the size of the tumor is less than 3mm. This is starting level of cancer tumor. Under this category is easily curable.

### Malignant

If the tumor is malignant, then the size of the tumor is greater than 3mm. This is an uncontrollable level of cancer tumor. Under this category is not curable.

## 2 MATERIALS AND METHODS

### 2.1 Preprocessing Stage

Preprocessing is the initial stage is to remove unwanted noise present in the original image using a median filtering technique. The goal of median filtering is to filter out noise that has corrupted image. It is based on a statistical approach. Typical filters are designed for a desired frequency response. Median filtering is a nonlinear operation often used in image processing to reduce “salt and pepper” noise.

A median filter is more effective than convolution when the goal is to simultaneously reduce noise and preserve edges. The median filter is a nonlinear digital filtering technique, often used to remove noise. Such noise reduction is a typical preprocessing step to improve edge detection on an image. The Fig-1 show entire process of this paper.

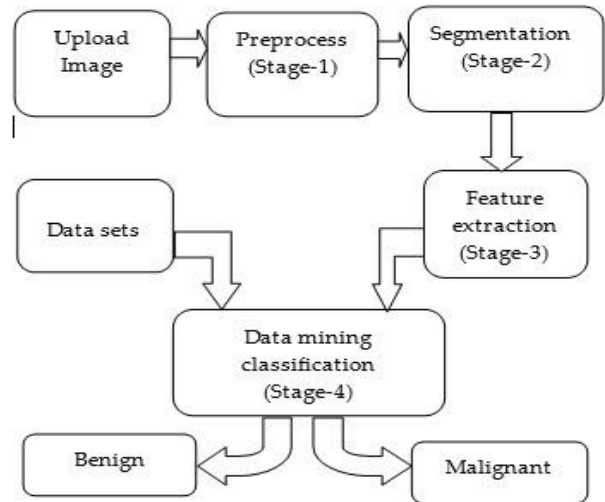


Fig -1: Block diagram of the proposed technique

### 2.2. Segmentation Stage

Segmentation stage is to separate the objects and borders (lines, curves) in an image. The original image is converted as a binary image that has only two possible values for each pixel. Typically the two colors used for a binary image are black and white, though any two colors can be utilized. Boundary tracing, selecting the starting point and the direction are critical for certain objects, especially the object that has some holes inside. Here, non-body pixels are traced.

Sobel method helps to discover the edges in an image, it does so by seeing the image gradient. Image gradient is the change in the intensity of the image. The intensity of the image will be of maximum value where there is a separation of two dissimilar regions thus an edge must exist there. The gradient will be greater where the intensity value is very large. The Sobel operator uses this greatest value to detect edges in an image.

In this proposed method our region of interest is tumor extraction. The proposed method consists of following steps.

1. The CT scan image is converted to grayscale image and we use a median filter to remove any noise if present.
2. Then the grayscale image is converted into a binary image. A binary image constructed such that all pixels with the corresponding gray level greater than the selected threshold are 1 and all other pixels are zero.
3. Boundary tracing is applied to trace the outline of non-body portions in the lungs of the binary image.
4. Morphological operation is used to delete all the unwanted pixel based on the pixel area.
5. Smoothing is often used to produce a less pixelated image.
6. The Sobel method finds edges using the Sobel approximation to the derivative. It returns edges at those

points where the gradient of I is maximized.

7. In conclusion, the filling operations automatically determines which pixels are in holes and then changes the value of those pixels from 0 and 1 and fills the holes in that binary image.

### 2.3. Feature Extraction Stage

Feature extraction is used to estimate the size of the tumor, to calculate the size of the tumor we need geometrical features like area, perimeter, irregularity index etc.

$$\text{Irregularity index} = 4\pi \frac{\text{Area}}{\text{Perimeter}^2} \quad (1)$$

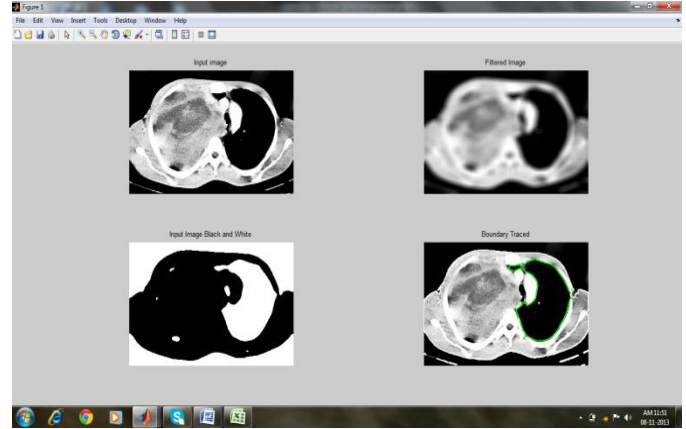
The number of pixels having the values '1' in the image array gives the area of the segmented tumor image. The number of boundary pixels in the tumor image is estimated as the perimeter of the tumor image.

### 2.4 Classification Stage

To classify the lung cancer, by using the data mining, classification techniques like Sequential minimal optimization (SMO), J48 decision tree, Naive Bayes, Logit boost etc., Once the classification is performed, we have to compare the experimental results of the above classification techniques, and find out which one gives efficient and correct results. For our convenience, we will use some data attributes like age at diagnosis, gender, marital status, smoking, panparag, tobacco, area, business, exercise, symptoms, treatment, tumor size, cancer stage and we include the feature extraction attribute like the size of the nodule as classification attributes. The final outcome of this project is to detect the cancerous nodule as benign (or) malignant.

## 3. RESULTS AND DISCUSSION

An automatic detection of lung cancer from CT images is performed based on image processing techniques; the experimental results show step by step process of cancer detection scheme. The Fig-2 shows the input image, filtered image, binary image and boundary traced image. The original image has to be filtered using the median filtering technique to remove noise. The filtered image is converted into a binary image. Then only we trace the boundaries of the non-body pixels.



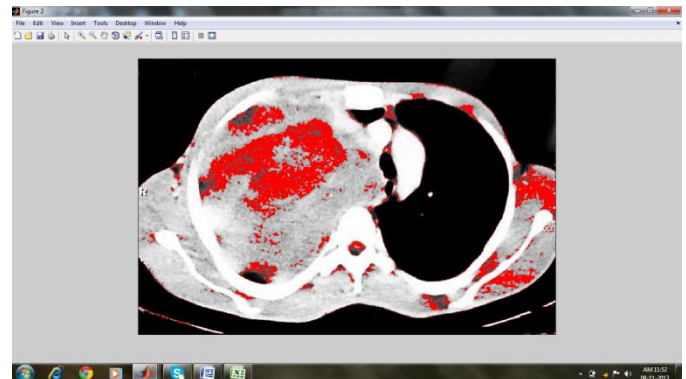
**Fig 2** Input Image, Filtered Image, Binary Image and Boundary Traced Image

Fig-3 shows the edge detected image. The edge has to be detected by Sobel edge detection method. After applying the Sobel method the edges are clearly visible.



**Fig 3** Edge Detected Image

Fig-4 shows tumor detected image. From this image we can easily identify the cancer affected area with the contrast color.



**Fig 4** Tumor Detected Image

Fig-5 shows the SMO classification output. Fig-6 shows the decision tree based on tumor size; the decision tree shows how many patients under benign and malignant category.

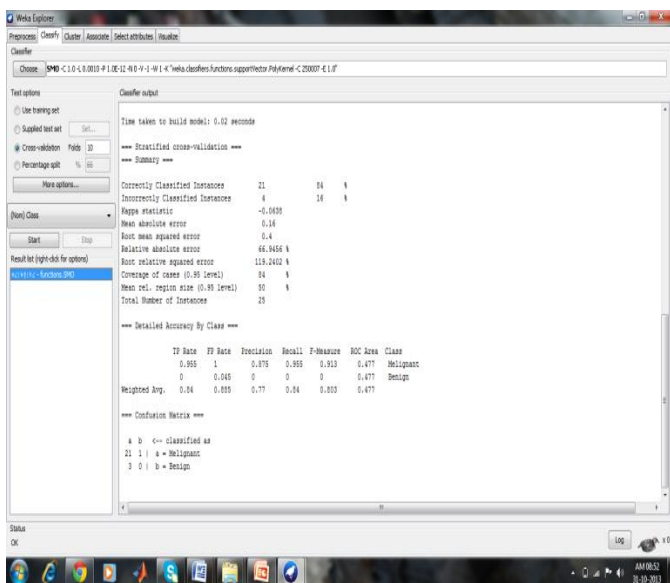


Fig 5 SMO Classification Output

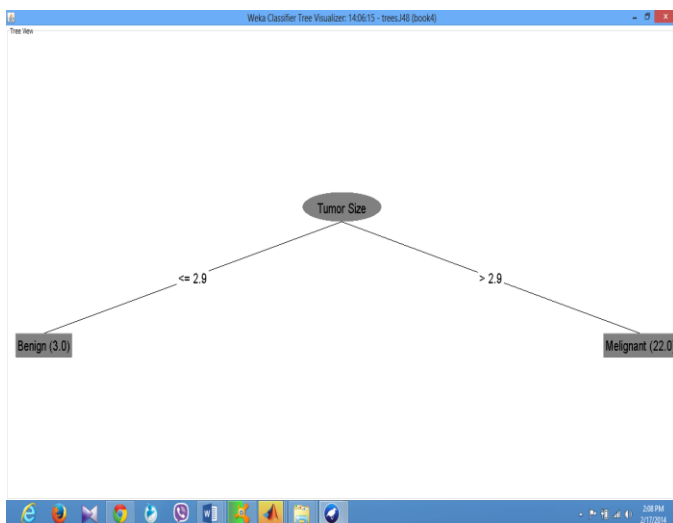


Fig 6 Decision Tree Based on Tumor Size

Table 1 depicts the performance measurement of various classification methods and its corresponding accuracy in percentage. Here Accuracy means the correctly classified instances. That is, malignant tumor correctly classified as malignant and benign tumor correctly classified as benign.

Table 1: Performance Measurement

Classification types	Accuracy (%)
SMO	84%
J48 Decision tree	96%
Logit boost	98%
Naïve Bayes	86%

#### 4. CONCLUSIONS

This proposed work addresses the image processing techniques to recognize the cancerous nodule from the lung CT images. In this paper, we develop an automatic detection of lung cancer in CT images using several image processing techniques. The accuracy of the tumor detected is checked using data mining classification techniques. Finally, we compare the classification techniques and find which one gives more correctness.

Future work incorporates, creating neural network models for the foreseeing framework. We likewise want to do comparable dissection for various cancers.

#### REFERENCES

- [1] Ankit Agrawal, Sanchit Misra, Ramanathan Narayanan, Lalith Polepeddi, Alok Choudhary, "A Lung Cancer Outcome Calculator Using Ensemble Data Mining on SEER data", ACM International Conference on BI-OKDD, pp. 1-9, 2011.
- [2] Joao Rodrigo Ferreira da Silva Sousa, Aristofanes Correa Silva, Anselmo Cardoso de Paiva, Rodolfo Acatauassu Nunes "Methodology for automatic detection of lung nodules in computerized tomography images", Computer Methods and Programs in Biomedicine, Vol. 98, pp. 1-14, 2010.
- [3] Ankit Agrawal and Alok Choudhary, "Identifying HotSpots in Lung Cancer Data Using Association Rule Mining", 11th International Conference on Data Mining Workshops, pp. 995-1002, 2011.
- [4] Sheng Chen and Kenji Suzuki, "Computerized Detection of Lung Nodules by Means of 'Virtual Dual Energy' (VDE) Radiography", IEEE Transactions on Biomedical Engineering, Vol. 60, No.2, pp. 369-378, 2013.
- [5] Tao Xu, Irene Cheng, Richard Long and Mrinal Mandal, "Novel coarse-to-fine dual scale technique for tuberculosis cavity detection using chest radiographs", EURO SIP Journal on Image and Video Processing, Vol. 3, pp. 1-18, 2013.
- [6] Ali Sadeghi Naini, Ting-Yim Lee, Rajni V. Patel and Abbas Samani, "Estimation of Lung's Air Volume and Its Variations Throughout Respiratory CT Image Sequences", IEEE Transactions on Biomedical Engineering, Vol. 58, No.1, pp. 152-158, 2011.

- [7] Kawsar Ahmed, Abdullah-Al-Emran, Tasnuba Jesmin, Roushney Fatima Mukti, Md Zamilur Rahman, Farzana Ahmed, "Early Detection of Lung Cancer Risk Using Data Mining", *Asian Pacific Journal of Cancer Prevention*, Vol. 14, pp. 595-598, 2013.
- [8] Ms. Swati P. Tidke, Prof. Vrishali A., Chakkarwar, "Classification of Lung Tumor Using SVM", *International Journal of Computational Engineering Research*, Vol. 2, Issue 5, pp. 1254-1257, 2012.
- [9] B. Magesh, P. Vijayalakshmi, M. Abirami, "Computer Aided Diagnosis System for the Identification and Classification of Lesions in the Lungs", *International Journal of Computer Trends and Technology*, pp. 110-114, 2011.
- [10] S. A. Patil and M. B. Kuchanur, "Lung Cancer Classification Using Image Processing", *International Journal of Engineering and Innovative Technology*, Vol. 2, Issue 3, pp. 37-42, 2012.
- [11] S. A. Patil and V R Udipi, "Chest X-ray features extraction for lung cancer classification", *Journal of Scientific and Industrial Research*, Vol. 69, pp. 271-277, 2010.
- [12] Stephen R Bowen, Matthew J Nyflot, Michael Gensheimer, Kristi R G Hendrickson, Paul E Kinahan, George A Sandison and Shilpen A Patel, "Challenges and opportunities in patient-specific, motion-managed and PET/CT-guided radiation therapy of lung cancer: review and perspective", *Clinical and Translational Medicine*, Vol. 1, Issue 18, pp. 1-16, 2012.
- [13] Rafika Harrabi and Ezzedine Ben Braiek, "Color image segmentation using multi-level thresholding approach and data fusion techniques: application in the breast cancer cells images", *EURASIP Journal on Image and Video Processing*, Vol.11, pp. 1-11, 2012.
- [14] Vanessa H. Tran, Hiang Ping Chan, Michelle Thurston, Paul Jackson, Craig Lewis, Deborah Yates, Graham Bell and Paul S. Thomas, "Breath Analysis of Lung Cancer Patients Using an Electronic Nose Detection System", *IEEE Sensors Journal*, Vol. 10, No.9, pp. 1514-1518, 2010.
- [15] Noah Lee, Andrew F. Laine, Guillermo Marquez, Jeffrey M. Levsky and John K. Gohagan, "Potential of Computer-Aided Diagnosis to Improve CT Lung Cancer Screening", *IEEE Reviews in Biomedical Engineering*, Vol. 2, pp. 136-146, 2009.
- [16] Parag Deoskar, Dr. Divakar Singh, Dr. Anju Singh, "Mining Lung Cancer Data and Other Diseases Data Using Data Mining Techniques: A Survey", *International Journal of Computer Engineering and Technology*, Vol. 4, Issue 2, pp. 508-516, 2013.
- [17] Abid Hasan, "Evaluation of Decision Tree Classifiers and Boosting Algorithm for Classifying High Dimensional Cancer Datasets", *International Journal of Modeling and Optimization*, Vol.2, No. 2, pp. 92-96, 2012.
- [18] Hualong Yu, Jun Ni, Yuanyuan Dan, Sen Xu, "Mining and Integrating Reliable Decision Rules for Imbalanced Cancer Gene Expression Data Sets", Vol. 17, No.6, pp. 666-673, 2012.
- [19] V.Krishnaiah, Dr. G. Narsimha, Dr.N. Subhash Chandra, "Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques" Vol.4, Issue 1, pp. 39-45, 2013.
- [20] P. Ayyadurai, P. Kiruthiga, S.Valarmathi, S. Amritha, "A Study of Lung Cancer Analysis and Identification Using Sobel Edge Detection Method and WEKA Tool", Vol. 3, Issue 3, pp. 2706-2712, 2013.