



# INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume 4, Issue 6)

Available online at: [www.ijariit.com](http://www.ijariit.com)

## Detection of lung cancer using computerized tomography scan: A review

Sakshi Sharma

[lky519382@gmail.com](mailto:lky519382@gmail.com)

Chandigarh University, Ajitgarh, Punjab

### ABSTRACT

*Lung cancer is very effective and causes deadly diseases among human. It is detected in the last stage, not in the initial stage, which can give the worst results. It also defects other parts of the body. Lung cancer has two types that are non-small lung cancer and small lung cancer. Smoking can be one of the main reason for lung cancer. This can be diagnosis by the different techniques many techniques were used to detect lung cancer, but CT is better for accurate results using imaging techniques. CT scan is helpful in predicting the performance of classification. Classification plays a major role in image analysis. Firstly, CT scan image as an input pass to the system through the image processing and then segmentation is performed. The main components of the structure are CT scan images, segmentation, nodule detection, feature extraction. Training data and testing data is used to determine the accuracy of both the algorithms and check the better accuracy. The main objective is to evaluate the computed tomography technique through SVM and then compare it with Naive Bayes to increase the accuracy, but there are some limitations to reach 100%.*

**Keywords**— *Detection on lung cancer, CT scan images, SVM, Naive Bayes*

### 1. INTRODUCTION

Lung cancer is a major defect that creates a cause in the lungs. Oxygen is absorbed by chambers present in our lungs during the inhaling process and carbon dioxide is released during exhaling. In the United States, lung cancer has become a major issue among humans. Each year it transports more humans than colon cancers, predation, ovary, and breast combined. It is difficult to trace because the causes are usually revealed during the last phase. The computed tomography images were reliably processed to detect lung cancer so that it could efficiently reveal all suspicious and nonspecific nodules. However, variations in the potential of computed tomography images and the various misjudgments of physicians and the radiologist may make it difficult to monitor the cancer cell. [1]

The main objective is to analyze the different computer-assisted techniques available for the lung cancer diagnosis. Then analyze the advantages, limitations, and disadvantages of all techniques to discover the best current technique and then propose improvements in the best current model to obtain a new model greatly improved. The lung cancer detection techniques were classified in the order of their detection accuracy. Then each of them was analyzed in each step and its drawbacks and limitations were pointed out. [2]

The main diagnosis, that is, computer-assisted, has recently become the most favorable tool to help doctors and radiologists detect cancer with greater precision. Many systems have been developed and the diagnosis of lung cancer is being investigated. However, some systems do not have a reasonable accuracy and, therefore, must be improved to have a more accurate detection that has a range of almost 100%. With abnormal cell growth, its different types are classified according to the type of cell initially affected, harming humans when damaged cells divide uncontrollably and generally form tumors.

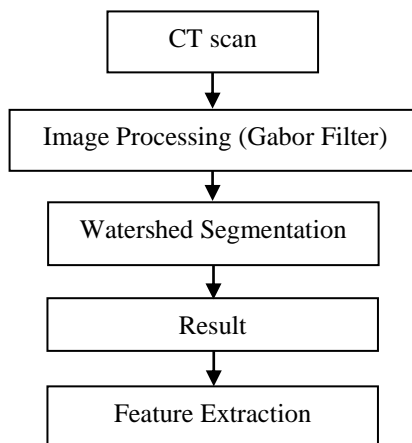
Tumors grow and interfere with human systems that release hormones that alter bodily functions. Lung cancer begins in the cells lining the bronchi, bronchioles, or alveoli. It acts in areas of precancerous changes and causes the cells of the human body to grow abnormally. Cell growth may seem abnormal under a microscope, but it does not form a tumor at this stage that shows no symptoms, which makes cancer undetectable. Cancer can allow the abnormal growth of cells to form a tumor that can be detected in imaging tests. In later stages, cancer cells spread rapidly to other parts of the human body. Most cases are not detected until they cause symptoms.

There are many common symptoms of lung cancer, such as aspergillus infection, chest pain, weight loss and loss of appetite, asthma, and a feeling of weakness. It is known that there are two types of tumors; benign and malignant tumors. Benign tumors are less than 3 mm and are curable cancer cases. Malignant tumors are larger than 3 mm and are uncontrollable. [9]

### 1.1 Related work using image processing

The Gabor filter is used in the preprocessing of images to improve the image and for segmentation. The controlled watershed method is used to detect the cancer symptom. This model supports the extraction of the attribute as area, perimeter, and eccentricity of cancer nodules. It will give the differentiation of the models given through its result as 90.1%, which is better than others. However, the best idea today. (Figure 1), has some limited scenario that is given below:

- Some characteristics of the cancer nodules have been extracted.
- Noise elimination, image smoothing without preprocessing can cause the increase in symptom detection that has run significantly.
- The evaluator has not performed any restriction operation, such as start-up or malignancy for the removal of a cancer infection.



**Fig. 1: Current model**

Changes have been made to the best solution today and a new proposed model was implemented, as shown in Fig. No. 2 given below. Apart from the Gabor purifier, the Median filter and the Gaussian filter have been well implemented in the preprocessing stage. After the pre-processing of the processed snapshot, the use of watershed segmentation is segmented. This gives the cancer image a marked symptom. In addition to functions such as the area, the perimeter and the wording of the eccentricity, characteristics such as the centroid, the diameter, and the added pixel were extracted well in the characteristic extraction stage of the detected cancerous symptom. After the detection of the carcinogenic nodule, the best model and also its extraction of characteristics and calculation of the efficiency ends. But cancer classification has not been implemented as benign or malignant. Therefore, the next phase of classification of the cancerous nodule has been made using SVM. The extracted properties are used as training functions and the coaching model is processed. The cancer-causing node is then classified using a predictor trainer model.

#### 1.1.1 Image Pre-processing

First, within the pre-processing of snapshots, the median purifier is used well in the pixel-scale image of the pixel of the CT images. Some noises are implanted in computed tomography during the process of image acquisition that helps during the incorrect discovery of the nodules. At some point, it can be diagnosed as cancerous nodules. Therefore, these eliminated noises must be removed to detect cancer alternatively. The media filter removes the noise from the ct scan image. Later, the media purifier, the Gaussian filter is well implemented. These filters help the segmentation of cancerous nodules even if you touch other false nodules.

#### 1.1.2 Segmentation

This process finds things or limits that support the area of essential interest in the pixel of the image. Divide the pixels of the image in the region to identify full details. In the detection of cancer of the diseased lung, he segmented the cancerous symptom in the pixel of the CT image. In this proposed model, the implementation of the watershed segmentation line is fine. The final characteristic is that it can distract and analyze well the things that touch the image. This process helps the proper segmentation of cancer nodes, even if it meets other faulty symptoms.

#### 1.1.3 Feature Extraction

In the extraction of characteristics, characteristics such as area, perimeter, diameter, eccentricity, and average intensity. These features extractions are later used as training functions for the developed classifier.

#### 1.1.4 Classification

Classification classifies the malignant or benign as nodules detected. As a classifier, the support vector machine (SVM) is used. SVM is a supervised learning technique. It is a supervised method. It helps to define the classifier of data between classes [9].

This process is defined as  $D(x) = wTx + b$ , where  $x_i$  is input training,  $w$  This is a dimensional vector, the term bias is  $b$ . Here,  $i=1.....m$ .

$$D(x) = wTx + b \geq 1 \text{ for } y_i=1$$

$$D(x) = wTx + b \leq -1 \text{ for } y_i=-1$$

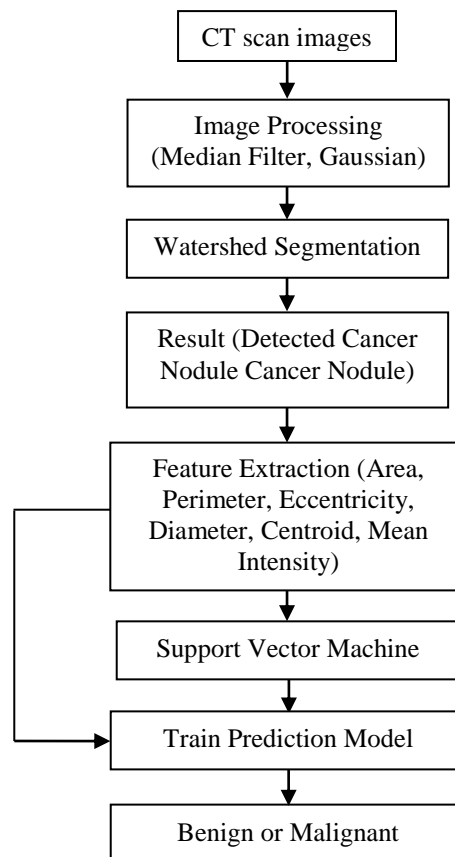


Fig. 2: SVM model

### 1.2 Related work using Naive Bayes

This classifier is a direct and better algorithm for the classification task. Even if we are working on a data set with millions of records with some attributes, it is recommended to try the Naive Bayes approach.

The Naive Bayes classifier gives excellent output when we use it for the analysis of texts. Like the processing of natural language.

Baye's theorem is titled after the Reverend Thomas Bayes. It is based on conditional probability. The conditional probability is the probability that something will happen, given that something else has already happened. We can calculate the probability of an event using conditional probability.

$$P(h|d) = (P(d|h) * P(h)) / P(d)$$

- $P(H)$  is the probability that hypothesis  $H$  is true. This is known as the previous probability.
- $P(E)$  is the probability of the evidence (independently of the hypothesis).
- $P(E | H)$  is the probability of the evidence given that the hypothesis is true.
- $P(H | E)$  is the probability of the hypothesis given that the evidence is there.

### 1.3 Comparison between SVM and Naïve Bayes

Naive Bayes belongs to the class of generative models for classification. The posterior probability is modeled from the conditional class densities. So the output is a probability of belonging to a class.

On the other hand, SVM is based on a discriminant function given by  $y = w.x + b$ . Here, the weights  $w$  and the  $b$ -parameter  $b$  are estimated from the training data. Try to find a hyper-plane that maximizes the margin and there is an optimization function in this sense. It is more likely that intelligent SVMs with respect to performance that uses the core of the radial basis function work better since they can handle nonlinearities in the data.

Naive Bayes performs better when features are independent of each other, which often does not happen in reality. That said, it still works well even when the features are not independent. The major difference between the models we are building from the features point of view is that Naive Bayes is treated for the independent features, while SVM observes the communication between them at some extent, as long as you use a non-kernel device linear (Gaussian, rbf, poly, etc.).

Therefore, if we have given a problem and also have interaction, we mostly used SVM that is better to capture them and therefore, classification is better.

In some of the cases, the consensus for researchers and practitioners of LD is used, so SVM is better than Naive Bayes. It is a bit difficult to compare the two methods from a theoretical point of view. One is probabilistic in nature and another one is geometric. However, it is fairly easy to find a function where one is dependent between variables that have not been fixed by Naive Bayes (and  $(a, b) = ab$ ), and we know that it is not a universal approximator.

However, the SVM with non-linear Kernel is (like the neural networks of 2/3 layers), it means theory matches the practice from that point of view. But in the end, it all goes down to the performance of our problem: basically, we want to decide the simplest method that will give us better results for our problem and have a better performance. Spam detection is famous for the Naive Bayes solution and facial recognition in images using a similar method improved with reinforcement, etc.

## **2. CONCLUSION**

The best current model does not have satisfactory results in terms of accuracy and does not classify the degree of cancerous nodules. However, a new one is proposed. This system helps the detection of cancerous nodules of the lung computed tomography by segmenting watersheds to detect cancer and the SVM is used for the classification of nodules such as benign or malignant. This model was successful in the detection of cancer with a high precision that is 92%, better than the current and the most elegant and with 86.6% accuracy.

In general, we analysed an improvement in the proposed system compared to the modern classic. Therefore, the proposal does not classify in several phases such phases are I, II, III, IV of cancer.

Thus, the additional improvement of this scope can be applied by classification in several phases. In addition, accuracy can be increased with proper preprocessing with the elimination of false values.

## **3. REFERENCES**

- [1] Suren Makajua, P.W.C. Prasad\*a, Abeer Alsadoona, A. K. Singhb, A. Elchouemic, "Lung Cancer Detection using CT Scan Images" 6th International Conference on Smart Computing and Communications, ICSCC 2017, 7-8 December 2017, Kurukshetra, India.
- [2] Abdelwadood M. Mesleh, "Lung Cancer Detection Using Multi-Layer Neural Networks with Independent Component Analysis: A Comparative Study of Training Algorithms," Volume 10, Number 4, December 2017 ISSN 1995-6673.
- [3] Jin, X., Zhang, Y., & Jin, Q. (2016) "Pulmonary Nodule Detection Based on CT Images Using Convolution Neural Network." 2016 9Th International Symposium on Computational Intelligence and Design (ISCID). DOI: 10.1109/ISCID.2016.1053.
- [4] Mr. Sandeep A. Dwivedi<sup>1</sup>, Mr. R. P. Borse<sup>1</sup>, Mr. Anil M. Yametkar<sup>2</sup>, "Lung Cancer Detection and Classification by using Machine Learning & Multinomial Bayesian," IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) e-ISSN: 2278-2834,p- ISSN: 2278-8735. Volume 9, Issue 1, Ver. III (Jan. 2014), PP 69-75
- [5] Anand M. Basware, Kalpana D. Joshi and Prakash S. Nalwade, "A Comparative Study Of K-Means and Weighted K-Means for Clustering," International Journal of Engineering Research & Technology, Volume 1, Issue 10, December-2012.
- [6] Modeled S. AL-TARAWNEH, "Lung Cancer Detection Using Image Processing Techniques," Leonardo Electronic Journal of Practices and Technologies Issue 20, January-June 2012, p. 147-158.
- [7] Khobragade, S., Tiwari, A., Patil, C., & Narke, V. (2016) "Automatic detection of major lung diseases using Chest Radiographs and classification by the feed-forward artificial neural network." IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES): 1-5.
- [8] Anjali Sevani, Hardik Modi (2018)" Implementation of Image Processing Techniques for Identifying Different Stages of Lung Cancer," International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 8 (2018) pp. 6493-6499.
- [9] Sruthi Ignatius, Robin Joseph, Jisha John, Dr. Anil Prahladan (2015) "Computer Aided Lung Cancer Detection and Tumor Staging in CT image using Image Processing," International Journal of Computer Applications (0975 – 8887) Volume 128 – No.7
- [10] Anam Tariq, M. Usman Akra, and M. Younus Javed, "Lung Nodule Detection in CT image using Neuro-Fuzzy classifier", IEEE Transactions (2013).