# LUNG CANCER DETECTION TECHNIQUES: A REVIEW

**NAMITA AWASTHI[1], VIMAL KUMAR AWASTHI[2], SUDHRITI SEN GUPTA[3]**

[1]*P.G, Student, Department of Computer Science & Engineering, KIT Kanpur, AKTU, U.P India*
[2]*Assistant Prof., Department of Computer Science & Engineering, KIT Kanpur, AKTU, U.P India*
[3]*Assistant Prof., Amity Institute of Information Technology (AIIT), AUUP, India*

---

**ABSTRACT:** Modern 3D medical imaging offers the potential for major advances in science and medicine as higher fidelity images are produced. Due to advances in computer-aided diagnosis and continuous advancements in the field of computerized medical image visualization, there is a need to develop one of the most important fields of scientific imaging. There are many types of cancer, of which lung cancer is one of the most common cancers. Machine learning techniques are widely used for lung cancer screening. This article compares different machine learning techniques for lung cancer detection.

**Keywords:** Lung Cancer, Segmentation, Feature Extraction, SVM. K-mean, DWT

## I. INTRODUCTION

Image processing is a technique used to enhance raw images or images captured from different cameras from different sources. With the help of image processing, important data can be recovered efficiently. In recent years, various methods have been developed in image processing techniques for efficient extraction of complex information. Image processing technology has been used by various operations in recent years. This approach is widely used in the military, clinical, and research fields. Some associations also use an image processing approach to simplify manual workload and take positive actions [14]. Image processing is applied in many applications, in particular to improve the optical description of images. For the preparation of the images, various calculations are also implemented. Digital image processing includes both visual image processing and analog image processing. Digital image processing uses different methods. Visual and digital image processing can be performed using images. This technique uses various areas such as infographics for image generation. This technique also provides assistance in manipulating and modifying images. The image or image is analyzed using processor hallucination or computer vision [1].

In lung cancer, the abnormal cells multiply and develop like a tumor. The lymphatic fluid that surrounds the lung tissue carries cancer cells from the lungs to the blood. Lymph circulates through the lymphatic vessels. This lymphatic fluid drains into enlarged lymph nodes in the lungs and in the mid-chest region. Lung tumor growth always takes place towards the mid-chest area due to the constant flow of lymphatic fluid towards the center of the chest. When a cancer cell leaves its area of origin, it metastasizes. This cancer cell now travels to a lymph node or to another part of the body with the help of the bloodstream. Primary lung tumor is a type of cancer that originates in the lung [3].

### 1.1. Phases of Lung Cancer Detection

The different stages of lung cancer screening are described below:-

**A. Image Enhancement**: In this procedure, the image is processed and smoothed. This process improves image quality and also removes image noise. Therefore, this process offers a superior key to digital imaging. The image enhancement procedure is related to image pre-processing techniques. The development of distinction, image sharpness is some of the goals of the image enhancement process. Careful examination of image enhancement techniques can be classified into two main sections:

- Spatial domain method: this technique is implemented directly in the image pixel.
- Frequency domain method: this method is performed on the Fourier transform of an image. This method is described as a low-level exemption approach [4].

Image quality is the foundation of image enhancement techniques. The optimization of the treatment results is carried out without arithmetic criteria. The image enhancement phase uses three approaches, such as the Gabor filter, automatic enhancement, and fast Fourier transform. Here is a brief description of these approaches:

a. **Gabor filter:** This filter was developed by Dennis Gabor for TC descriptions. The Gabor filter is a linear filter used to examine texture. This filter examines the presence of a particular type of frequency content in the image according to particular instructions in a localized area. A two-dimensional Gabor filter is a kind of Gaussian kernel function fitted by a plane sine wave in the spatial field. This filter is considered a fundamental instrument in computer visibility and image processing in particular for the study of texture due to its optimal localization properties in the dimensional and frequency domain [4].

b. **Fast Fourier transform**: This approach works with the Fourier transform of a specific image. The field of incidence is a space where each image classification at the image location F symbolizes a quantity where the intensity norms of the image "I" differ by a precise detachment connected to F. The fast Fourier transforms approach it is used in image enhancement.

c. **Auto-enhancement:** This approach is based on oblique monitoring and mathematical operation. In this operation the mean and deviation are calculated.

**B. Image Segmentation:** The image segmentation procedure maps a digital image into different sections as sets of pixels also recognized as super pixels. The main objective of this process is to modify an image demonstration in a simpler investigative way. Image cropping is used to identify the location of objects, boundaries, and edges of images. In this process, a label is assigned to each pixel in an image, and therefore pixels with the identical label share defined characteristics. There are several techniques available for image segmentation. These techniques are shown below:

- **Thresholding Approach:** This approach is considered a very important technique for image segmentation. The sectioned image obtained from the threshold has several advantages, such as less storage space, fast distribution speed, and ease of operation compared to a grayscale image that typically comprises 256 steps. In the presented work, a grayscale image is used for the thresholding process. In this process, the RGB image is converted to a binary image. The resulting image is in black and white form. This black and white image contains only two tones. The black color in the image represents level 0 while the white color in the image represents level 1. The threshold will be between 0 and 1 because the image contains only two levels. After reaching the threshold value, the image will be sectioned based on it.
- **Marker-Controlled Watershed Segmentation:** The method of segmentation of watersheds by markers indicates the appearance of objects or bottom in an explicit position of the image [5]. The tag-controlled watershed segmentation technique consists of two methods: the outer type is connected to the bottom, and the inner type is connected to the item of importance. Image segmentation by watershed modification works accurately if we can discover or 'tag' foreground features and background positions to find 'watersheds' and 'watershed ridge lines' in an image by adjusting it as a flat where bright pixels are tall and dark pixels are thick. According to the evaluation of experimental inclination during the entire cut-off period, the marker-controlled watershed segmentation approach shows precision and eminence over the threshold approach.

**C. Feature Extraction:** In the field of image processing, functionality plays an extremely important role. Different image preprocessing approaches like binarization, thresholding, normalization, masking approach, etc. they are implemented in the sampled image before obtaining functionalities. Then, to obtain the characteristics, feature elimination approaches are performed for the classification and detection of the descriptions. These are the techniques involved in extracting features:

a. **Binarization**: In this approach if the amount of black pixels in a digital x-ray picture is above than that of white pixels in a description then it is concluded that the x ray details are normal lung picture otherwise it is considered cancer affected.

b. **Masking approach**: This technique depends on the fact that the ample are emerged in the form of white linked regions within ROI (lungs), because they augment the proportion of cancer presence. The appearance of solid blue color describes the usual case while presence of RGB ample demonstrates the tumor occurrence.

**D. Classification:** various classifiers are used for performing the classification on the basis of retrieved characteristics. A brief description of several classifiers is given below:

i. **Support Vector Machine (SVM) Classifier:** SVM stands for support vector machine and it is a classification algorithm that is based on optimization theory. As it maximizes the margin it is also known as a binary classifier. All the data points of an individual class are separated by the best hyper plane, this can be identified through the classification provided by SVM. In the SVM the largest the best hyper plane is described by the largest margin between the two classes. There are no interior data points when there is maximum width between the slabs parallel to the hyper plane which are also known as margin. The maximum margin in hyper plane is separated by the SVM algorithm [6].

ii. **Naïve Bayes Classifier**: The main aim of this classifier is the implementation of a strategy where future objects are assigned to a group in the presence of a pattern of objects for every class. The applied variable vectors are demonstrated with the help of future entities. These kinds of issues are recognized as the issues of supervised categorization. The researches have projected different techniques for the development of strategies for these issues. One of the most significant techniques of naïve bayes classifier is Pone. This is also identified as idiot's Bayes, simple Bayes, and independence Bayes. This technique can be developed in a simple manner and does not need essential restriction inference processes. Therefore this approach is considered appropriate for large data samples.

iii. **Decision Tree Classifier:** These classifiers are considered as non-parametric supervised learning techniques and used for categorization and deterioration. The main aim of this approach is the development of a model for the accurate prediction of an intended variable in accordance with several key variables. In this approach, each core nodule communicates with one of the key variables. Each side demonstrates the value of the target variable. In this classifier, every interior (non-leaf) nodule is labeled with the help of a key trait. The rounded sections created from a nodule named with a trait are labeled with every probable characteristic value. Each sheet of this classifier is labeled with a class or a prospect allocation larger than the groups [7].

iv. **K-Nearest neighbor:** This classifier depends on the learning by similarity. The n dimensional arithmetic qualities are utilized for the description of training sets. A point is represented by each training sample in the n dimensional region. The superior element of the training samples is amassed in an n-dimensional sample space besides these lines. The k-nearest neighbor classifier seems for the sample space for the k training samples that are nearby to the unidentified model in the case of unrecognized training sample. Euclidean distance is the term used for description of "closeness".

## II. LITERATURE REVIEW

**Amir Roointa [8]** in 2019 examined the development of an inclusive molecular description of tumor mass. It was clarified that disease biomarkers play a fundamental role in the early detection and analysis of tumors. The presented work summarized the rapid development of biosensor kits for the discovery of lung tumor biomarkers.

**Jing Songa[9],** in 2019, proposed a new microscopic hyperspectral imaging approach for the identification of lung tumors affected by ALK. In this approach, a home microscopic hyperspectral imaging scheme was used to capture images of five classes of lung tissue. To minimize the perceptible clamor of the tape and the sound element of authentic information, a pre-processed algorithm has been introduced. After that, a combination of support vector machine, massive study, and aggregate processing was used for the projection of a segmentation algorithm. The variation of the ethereal graphs, the virtual amount in the middle of the cytoplasm and the cell nucleus was used to differentiate the fluctuating spectral curves and the relative ratio between the cytoplasm and the cell nucleus for affected and unaffected ALK lung cancers. The results tested showed that the ALK affected pool contained a comparative amount of cytoplasm of 77.3%, while the ALK positive pool contained a comparative amount of cytoplasm of 40.6%. The results of the research linked to the quantitative examination and the ethereal curves demonstrated that the treatment of the lung tumor affected by ALK implemented with drugs at low concentration would develop towards the lung tumor not affected by ALK.

**Guobin Zhang [10]** presented in 2019 a serious evaluation of the CADe scheme for automatic recognition of lung cancer using CT descriptions to summarize existing developments [10]. In the initial phase, a brief description of the CADe system was provided. After that, a full summary of the five main mechanisms of the CADe scheme has been provided. These mechanisms included information gathering, pretreatment, lung image segmentation, nodule recognition, and reduction of false positives. A brief summary of the superior nodule

Detection methods and classifiers based on understanding, false positive value, and other limited data were also provided. After several studies, the CADe scheme was evaluated as essential for the timely recognition of malignant lung tumors.

**Moritz Schwyzer [11]** estimated in 2018 the usefulness of machine learning for the recognition of lung tumors in FDG-PET images in the very low quantity PET setting. In the absence of lung tumor, the artificial neural network report on

patients with selective lung cancer was examined. The sensitivity rate of 95.9% and 91.5% was achieved by the artificial neural system for the detection of lung cancer. The artificial neural network achieved 98.1% and 94.2% accuracy, at medium dose and ultra-low dose PET of 3.3%, as a result. The tested results demonstrated that the machine learning approach aids in the fully automatic recognition of lung tumors at extremely low and efficient radiation doses of 0.11 mSv. It has also been suggested that further advancements in this technique could improve the accuracy of lung tumor detection approaches.

**Suren Makajua [12]** in 2017 stated that CT images could be used for the recognition of lung tumors. The main objective of this study was the evaluation of different automated technologies, the investigation of the best existing method, the recognition of its limitations and drawbacks and the projection of a decisive system with several advances. To this end, lung tumor recognition approaches have been ranked based on the accuracy of their lung cancer analysis. At each step, these lung cancer recognition methods were reviewed and their limitations and drawbacks were taken into account. Different techniques for detecting lung cancer have been identified to have shown different precision. Some techniques have shown a lower accuracy rate, while other techniques have shown a good accuracy rate for lung cancer detection, but no technique has shown 100% accurate lung cancer detection.

**Madhura J [13]** presented in 2017 a review of noise reduction approaches for the diagnosis of lung cancer [13]. Lung cancer has been claimed to be a serious disease caused by abnormal growth of cells in lung tissue. Among all the other types of tumors, the lung tumor has been identified as the most common cancer. As a result, this cancer has become the cause of death for several cancer patients. The early recognition of lung cancer was very important in protecting several lives. The review study presented presented a brief overview of the lung tumor. This review work also described the different types of noise present in images, techniques for obtaining visible images, and noise removal methods. This article also provides a brief review of existing denoising methods.

**K. Punithavathy[14]** developed in 2015 a methodology for the automatic detection of lung cancer from PET / CT images [14]. Image pre-processing methods such as Adaptive Contrast Limited Histogram Equalization (CLAHE) and Wiener filtering have been used to eliminate artifacts due to variations in contrast and noise. The lung regions of interest (ROI) were extracted from the images using morphological operators. Haralick statistical texture features were preferred because they extract more texture information from cancer regions than visual assessment. Fuzzy mean grouping (FCM) was used to classify regions as normal or abnormal. The proposed method was performed using PET / CT images of lung cancer patients and was implemented using MATLAB. The performance of the proposed methodology was evaluated using the Receiver Operating Characteristic (ROC) curve. The proposed method allows a better classification and better detection of cancer with an overall precision of 92.67%.

**Anjali Kulkarni [15]** developed in 2014 an image enhancement technique for the earliest detection and processing steps; the time factor is taken into account to discover abnormality problems in the target images [15]. Images captured by CT are processed. The region of interest, that is, the tumor, is precisely identified from the original image. The Gabor filter and basin segmentation give the best results for the pretreatment stage. From the extracted region of interest, three characteristics are extracted, namely area, perimeter, and eccentricity. These three characteristics help identify the stages of lung cancer. The results indicate that the tumors are of different sizes. By measuring the tumor area, the stage of lung cancer can be accurately detected using the proposed method. The results show good potential for the early detection of lung cancer.

**Anam Tariq [16]** in 2013 he proposed a computerized system for the detection of pulmonary nodules on CT images. The automated system consists of two stages, namely lung segmentation and enhancement, feature extraction and classification. The segmentation process will separate the lung tissue from the rest of the image, and only the lung tissue examined is considered candidate regions for detecting malignant nodules in the lung portion. A characteristic vector is calculated for the possible abnormal regions and the regions are classified using neuro fuzzy classifier. It is a fully automatic system that does not require manual intervention and the experimental results show the validity of our system.

## III. REVIEW METHODOLOGY

This research work is related to the detection of lung cancer from the CT image using image processing techniques. The proposed methodology has four phases for the localization and characterization of lung cancer. These are the different phases of lung cancer detection:

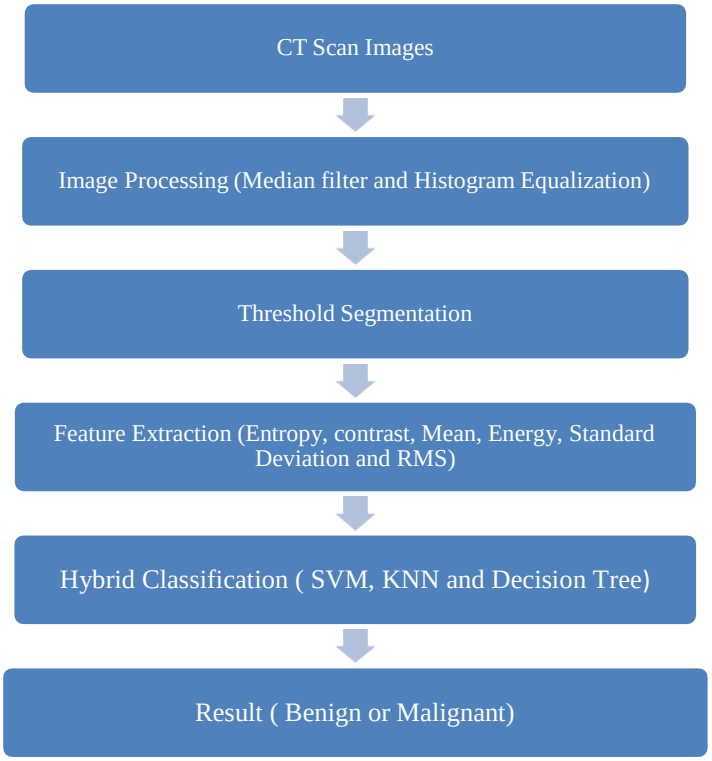


**Fig 3.1:** Proposed Model

1. **Pre-processing:-** Preprocessing is the first phase during which the CT image is taken as input. A noise removal technique will be applied that will remove noise from the input image.
2. **Segmentation: -** In the second phase, the region-based segmentation approach will be applied which will segment the similar and different regions of the CT image. The image segmentation procedure maps a digital image into different sections as sets of pixels also recognized as super pixels. The main objective of this process is to modify an image demonstration in a simpler investigative way. Image cropping is used to identify the location of objects, boundaries, and edges of images.
3. **Feature Extraction: -** Characteristics extraction is the third phase, in which the GLCM algorithm will be applied to extract characteristics from the CT image. In the field of image processing, functionality plays an extremely important role. Different image preprocessing approaches like binarization, thresholding, normalization, masking approach, etc. they are implemented in the sampled image before obtaining functionalities.
4. **Classification: -** In the last phase, the hybrid classification approach will be applied, which allows the cancerous part to be categorized and located. The hybrid classification algorithm is the combination of SVM, KNN, and decision tree. SVM stands for Support Vector Machine and is a classification algorithm based on optimization theory. Since it maximizes the margin, it is also known as a binary classifier. All data points of an individual class are separated by the best hyperplane, this can be identified by the classification provided by SVM. In SVM, the largest and best hyperplane is described by the largest margin between the two classes. There are no interior data points when there is a maximum width between slabs parallel to the hyperplane, also known as margins. The maximum margin in the hyperplane is separated by the SVM algorithm. This classifier depends on the learning of similarities. The n-dimensional arithmetic qualities are used for the description of training sets. One point is represented by each training sample in the n-dimensional region. The top element of the training samples is accumulated in a sample space with n dimensions in addition to these rows.

## IV. CONCLUSION

For lung cancer detection, image processing is used. There are three steps in detecting the cancerous nodule.To detect the presence of cancerous nodules, CT scans are used. Further the pre-processing composed of two processes.Image enhancement and image segmentation are these two processes. For human viewer interpretability of information in the image is improved by image enhancing step. There are many enhancement algorithms such as Gabor filter, fast fourier transform, log gabor filter and auto enhancement. In pre-processing second step is Image segmentation. The purpose of image segmentation is to divide the image into a significant region and identify the relevant object or information from the digital image. The output from the segmentation process is goes to feature extraction stage. Features such as area, perimeter and irregularity are found out in feature extraction.

## V. REFERENCES

[1] DeepikaChoudhar, Anjali Lather, Sandeep Yadav, "Overview of Image Processing", 2014, International Journal for Research in Applied Science & Engineering Technology

[2] Ranu Gorai, "A Survey Of Digital Image Processing", 2016, International Journal of Research in Engineering, Technology and Science, Volume VI, Special Issue

[3] Mokhled S. AL Tarawneh, "Lung Cancer Detection using Image Processing Techniques", 2012, Leonardo Electronic Journal of Practices and Technologies

[4] Shradhafule, "Lung Cancer Detection Using Image Processing Techniques", 2017, International Research Journal of Engineering and Technology (IRJET)

[5] BariqiAbdillah, AlhadiBustamam, and DevviSarwinda, "Image processing based detection of lung cancer on CT scan images", 2016, The Asian Mathematical Conference

[6] Keziah T, Haseena, "Lung Cancer Detection Using SVM Classifier and MFPCM Segmentation", 2018, International Research Journal of Engineering and Technology (IRJET)

[7] P. Bhuvneshveri, A. Briantha Therese, " Detection of Cancer in Lung with KNN Classification using Genetic Algorithm", 2014, 2nd International Conference on Nanomaterials and Technologies

[8] Amir Roointan, , "Early detection of lung cancer biomarkers through biosensor technology: a review",2019, PBA 12266

[9] Jing Songa, Menghan Hua, JianshengWanga, Mei Zhoua,b, Li Suna, Song Qiua, Qingli Lia, Zhen Sun, YitingWanga, "ALK positive lung cancer identification and targeted drugs evaluation using microscopic hyperspectral imaging technique",2019, Infrared Physics & Technology

[10] Guobin Zhang, Shan Jiang, Zhiyong Yang, Li Gong, Xiaodong Ma, Zeyang Zhou, Chao Bao, Qi Liu, "Automatic nodule detection for lung cancer in CT images: A review", 2018, CBM 3128

[11] Moritz Schwyzer, Daniela A. Ferraro, Urs J. Muehlematter, Alessandra Curioni-Fontecedro, Martin W. Huellner, Gustav K. von Schulthess, Philipp A. Kaufmann, Irene A. Burger, Michael Messerli, "Automated Detection of Lung Cancer at Ultralow dose PET/CT by Deep Neural Networks - Initial results", 2018, LUNG 5827

[12] Suren Makajua, P.W.C. Prasad, AbeerAlsadoona, A. K. Singhb, A. Elchouemi, " Lung Cancer Detection using CT Scan Images", 2017, 6th International Conference on Smart Computing and Communications, ICSCC

[13] Madhura J, Dr .Ramesh Babu D R, "A Survey on Noise Reduction Techniques for Lung Cancer Detection", 2017, International Conference on Innovative Mechanisms for Industry Applications

[14] K.Punithavathy, M.M.Ramya, SumathiPoobal, "Analysis of Statistical Texture Features for Automatic Lung Cancer Detection in PET/CT Images", International Conference on Robotics, Automation, Control and Embedded Systems – RACE 2015 18-20 February 2015

[15] Anjali Kulkarni, AnaghaPanditrao, "Classification of Lung Cancer Stages on CT Scan Images Using Image Processing", 2014 IEEE International Conference on Advanced Connnunication Control and Computing Technologies (lCACCCT)

[16] Anam Tariq, M. Usman Akram and M. YounusJaved, "Lung Nodule Detection in CT Images using Neuro Fuzzy Classifier", 2013 Fourth International Workshop on Computational Intelligence in Medical Imaging (CIMI), Pages: 49 – 53

## BIOGRAPHIES

NAMITA AWASTHI
P.G Student,
Department of Computer Science & Engineering,
KIT Kanpur AKTU, U.P India.

VIMAL KUMAR AWASTHI
Assistant Prof., Department of Computer Science & Engineering,
KIT Kanpur AKTU, U.P India.

SUDHRITI SEN GUPTA
Assistant Prof., Amity Institute of Information Technology (AIIT), AUUP, India