Lung Cancer Detection Using Image Processing Techniques

1Aman Kumar, 2Anchit Agarwal, 3Asheesh Kumar Sharma, 4Javed Miya

1,2,3 Student, Dept. of Information Technology, Galgotia’s College of Engg. And Technology, Greater Noida, Uttar Pradesh, India
anchita45@gmail.com

4 Assistant Professor, Dept. of Information Technology, Galgotia’s College of Engineering And Technology, Greater Noida, Uttar Pradesh, India
javed.miya@galgotiacollege.edu

ABSTRACT
Lung cancer is one of the most growing harmful diseases as we can see various patients suffering from this all around the countries and the biggest problem for this is early detection as this disease only comes to know when it is at very last stage. This paper focus is on early detection of lung cancer at its initial stage so that the person survival rate or life expectancy would increases for some year. In this for detection of lung cancer we use the Computed Tomography images as it is better than other medical images which are used in medical imaging. The cancer are of four stages in which the first and second stage can be handled easily and the person can be handled very easily. But if the tumor in third stage it becomes somehow difficult to handle and recover and it becomes very dangerous at it’s final stage. There are various techniques that have come’s early as in this the doctor detect the tumor by themselves by use of biopsy report and other techniques. These techniques are time taken and money also and are very prominent to human errors as when the patient required immediate treatment they cannot get and lead to the loose of his/her life. So in this case the image processing techniques are very useful for detection of lung cancer as through this we can detect the cancer as fast as possible and chances of error is reduced and patient get accurate treatment as required to him. So in this we are using MATLAB through each technique made. In our project process like image pre-processing, segmentation, feature extraction has been examined in details. We are planning to make more accurate system to give the exact outcomes.

1. INTRODUCTION
Lung cancer mostly occurs in those people who does smoking and at first it beings in the lung. Cancer is the disease in which the cell of the body grow out of control. The cancer is mainly of 2 category cancer non- small cell and small cell lung cancer depending upon how they appear under microscope. In lung cancer, the cell overgrowing is seen in lungs which are the very most important parts of our body system or organ for doing various tasks inside the body like gas-exchange and breathing. In a recent studies it is found that,[1] in males more than 85% of lung cancer is found and more than 75% in females or women the cancer is detected whose main cause of occurrence is smoking. The lung cancer can be detected who is smoking currently or the person has left the smoking from...
According to the American cancer society it is calculated that in 2017 nearly about the 222,500 lung cancer disease cases are occurred in the United States and from that 155,800 have been died because of not getting proper treatment[2]. Lung cancer is difficult to detect at the early stage for decreasing death rates as mostly it comes to know at the final where the chances of survival is very low. The death ratio’s can be decreased by early diagnosis and detection of a cancer and then giving the proper treatment facilities to the patient as early as possible. The imaging technique used for detection for lung cancer is CT imaging which is most accurate and reliable for diagnosis and detection of the disease as it will reveal all suspect and unsuspect of the Cancer’s clot [3]. The detection of a lung cancer is very difficult at the early stages as it is only detected at the last stage so to detect at early stage we use various computer-aided drafting system that had made to detect the lung cancer at the beginning stages so that the death rates would be reduced. Detection and Diagnosis for a lung cancer at a beginning stage would increases the chances of the life expectancy upto 15% to 50%[4]. They occurs because of the deformity of the cells in a body. It is mainly divided into 2 types: it can harmless or malignant. The cancer which can be immediately remove or can be stop during its growing period to others or various organs of a body, is a kind of harmless cancer. The other part of the cancer is that which grow’s rapidly and spread vigorously to all the others part of a body is called the harmful cancer. The MI (medical imaging) is to be very much used method to look inside the whole anatomy as it is very comfortable and reliable and safe techniques for those persons who are affected through a lung cancer diseases. It play’s the most crucial role for the clot observations along with the therapy of Lung carcinoma at the early stage. It is most accurate, efficient and reliable method for the diagnosing the lung cancer as compared to other methods.

In medical Imaging world there are distinct type of imaging that should be used, for a lung cancer observation and examination the CT pictures are being most well-liked as they give better transperancy, image enhancement, less destroation and low noise in the CT images which made this technique more useful for detection of lung cancer. So at last we compare our images to the another pictures in terms of calculating the variance, median, area of images. The cancer observations or discovery process consists of 4 parts: Preprocessing of image and Image improvement, image partitioning, analysis of features, image classification. In all these Processes segmentation part of lung images is determined as the most tough task as this include many image preprocessing parts and another various steps which becomes difficult for the implementation[5].

2. TYPES AND TECHNOLOGY USED

In the recent year there are various tools and technology used by various researchers for the observation of cancer using many kinds of techniques related to image processing so that how can we easily observe a cancer in the initial period so that a death ratio would be decreased. The main aim of researchers have to developed such type of a system which are able to predict and detect the cancer at a very early stages instead of getting knowledge about the lung cancer at the very last stage where the chances of the survival becomes very less. The also try to improve the accuracy and reliability of the such systems or methods or tools and technology which are predicted early and for detection system by feature extraction, segmentation, preprocessing and classification techniques for extracted database[6].

Sankaret al. [7] For lung segmentation the best method is marker watershed segmentation method is carried out for a partioning of lung images. This methods provides us more accurate results.

Patle and Sirohi, Roy [8] proposed the active contour model and fuzzy interference system which is used to detect lung cancer clot. For image improvement and image variation the grey transformation is used. Before the segmentation process
the binarization of images is done or performed and the image results are partition using the model that is active contour model. Classification of cancer images is carried out by using the inference method. Lung cancer nodules are extracted on the attributes such as mean, correlation, perimeter, area, eccentricity and convex area are taken out and on that basis we used to give training to the machine. Overall, accuracy of the system would be approx 93.11%. The drawback of the above procedure that the proposed method would not distinguish for cancer that it is a benign or malignant.

G. Kaur and H. Singh[9] proposed the automatic procedure for cancer detection and size observation for each frame of Positron emission tomography pictures of lung. It states that information is represented by the use of frequency domain characteristics(discrete cosine transform, wavelets) and spatial features (geometric moments) and then on that basis the comparison has been made. The methods and machines used is: Nearest neighbor and K-means clustering and are some machines which are used for performance analysis of the above attributes and on the basis of that provides the results. This procedure facilitates the physicians and doctors in accurate analysis of lung cancer detection so that proper surgery and treatment should be provided to the patient which is suffering from this disease so that his life expectancy rate increases in order to that the death rates decreases.

3. METHODOLOGY

In medical image segmentation, the accuracy is foremost important, as it deals with human lives. This methodology mostly focuses to obtain a more correct and accurate, precise results. In this proposed method, we are enhancing the differences of the input images by pre-processing the image through the various pre-processing method. In this first the picture is changed to a black and white picture at the very first step it is done. Now with the help of Gabor filter the images are enhances the image contrast. Gabor filters orientation and frequency representations is same like of those as the person visualization systems, and this had been establish as the agreeable for the pattern illustration and differentiation. Image processing technique, a segmentation part is very most important part of the processing of image, in this method as we get the enhanced images and then these images are segmented into two or more than two small-segments or small parts that will make this technique best as the image becomes precise and large data gets converted into small data and this makes the very simpler illustrations for the extraction of features from the images. The partitioning part of images is used to find the lines, curves, objects and boundaries in the images or binaural systems. So at last the overall aim of the segmentation of the image is that it would provide very simple, clear and concise image so that it only provide’s us the user specific data on which we have to work on further steps and can easily make analysis on the images. There are various examples in which segmentation of image plays vital role are in traffic controls systems, face recognition, recognition task and video surveillance etc.

There are mostly three main processes that we are using throughout our survey paper; Pre-processing of images, extraction of features of images after its pre processing and then the final result process and the output is produced. So for conducting each of these process we use MATLAB throughout our project as it is very fast and it has in built many libraries which are used in the image processing. In detection of cancer in lungs we carried out our detection procedures in various steps as shown below in the fig(1).
**Image Acquisition**

This is the very initial phase that is to collect the images of computer-assisted tomography scan of the victim who is suffering from the this disease from hospitals or extract from the database. As we know that the pulmonary computer-assisted tomography scan pictures have the less undesirable effects produced due to noise and distortion as we differentiate to others images like radiography images and magnetic resonance images hence they are considered as best in image processing and also for developing the techniques or systems. The main aim of using CT pictures is that, it provides better clarity and quality of images and less distortion and noise in the images so that it makes it more reliable to use CT images. The images were collected in raw form and these images have lots of noise (means undesirable effects which are produced due to it). So for enhancing the quality, clearness, brightness, exactness of the image we have to reduce the undesirable effects from the background of an image which is because of noise in it. Hence, various methods has been proposed for reducing the noise in the image to enhance the quality of a photos as required for the further upcoming steps so the various techniques are like smoothing, enhancement etc.

**Image Preprocessing**

The main aim of image preprocessing deals with the enrichment of the data in the picture or image which tries for the removal or suppressing of undesirable noise and distortion and improves the image for the further extraction of the features from the image and further evaluation. This phase starts with the enhancement of the image as first the image is acquired now form the acquired image the image is undergone various preprocessing steps. The main aim of image preprocessing deals with the removal of unwanted areas from the image due to which the image content can be divided into two parts that is valuable and un-valuable content of image and this is done by enhancing the image features like curves, edges of the image and texture of the image[10]. The preprocessing phase deals with the removal and filtration of the un-desirable area in an image and it is also concerned with the removal of blueness from the image. The preprocessing involved in conversion, image resizes, noise removing and enhancement of image quality and at last produces the image. So for noise filtration various researchers used various filtering techniques based on the types of the noises present in the image[11]. Some filtration techniques which are used for filtration of different types of noise are such as [10]:

1. **Gaussian Noise**: This is the noise which we cannot usually see in the image very easily.

2. **Poscision Noise**: In this the noise is present due to the result of image detectors and disks recorded as the mean and variance of it is equal.

There are many more noise which we will talk further.

**Image Enhancement:**

Image enhancement is a technique whose basic aim to improve the quality of the given image. This provides the results as the other image which demonstrates the features in a manner that are better than the original image in terms of there appearance and other factors. It is the process of manipulating the digital images simple to change its brightness, contrast, color etc which make the image analysis very easy and comfortable for the future work on the image. So to make the image quality better we would like to enhance the image from noising, distortions and interference. There are 3 main methods which are used for image enhancement purpose i.e : Gabor filters (which show the best results) as compared to these two others method Fast And Fourier Transform, Auto improvement algorithm(that provides the bad results for further segmentation process). So in this we
use the gabor filter to enhance the images as it has the best and accurate results as compared to others as we can see it in further table that is studied from various research paper.

Image enhancement technique is broadly classified into two main categories[12]:

1. **Spatial domain** – This is the method which combines the pixel of the image and this method directly operates on the image pixels rather than other part. This is also known as spatial filtering. As is works on the pixels become for the zooming purpose and the edge detection purpose. Mathematical expression for this method is given as:

   \[ G(U,V) = T[f(u,v)](1) \]

2. **Frequency domain** – Now when we talk about this method this deals with the change in pixel valves of a image. In this the image is represented by mathematical function and any change in frequency in a image means the change the geometry of image means its shape, structure and other parts. The main benefit or advantage of using the frequency domain is that we can handle the difficult task very easily through it as compared to the spatial domain method. Now to make it simple we convert the space domain (pixel) into frequency domain (mathematical function) so that we can easily perform as required task and the again convert it easily into the space domain and the tool which we can use for this is like Fourier Transforms and Fourier Series.

**Segmentation Process**

Image partitioning or segmentation is a procedure of partitioning the images into small segments or multiple small images. The main aim of image segmentation is to change the way of representation of an image to the image that is more meaningful and is easy for carrying out analysis. It also helps to finds the boundaries and entity in the images such as lines and curves. It is a technique of partitioning the image into several images based on the image features like its area, perimeter, intensity, colour etc. Image segmentation is useful for improvement of pictorial information for human interpretation. The applications of image segmentation include recognition task, automatic task handling, broadcast survillence, compression of image. Segmentation is based on texture, grayscale, motion and dept etc. The algorithms used in the image segmentation process are: Thresholding, Marker controlled watershed, k clustering, K-NN algorithm etc. Segmentation is the process of grouping pixels together on the basis of there similar attributes.

**Thresholding Method**

The image thresholding method is the simplest method in the image partitioning process. It is a effective and very easy and simple method for separations in an image of foreground from background based of there sameness. It is based on the level of clip which basically converts the back and white images to the 0-1 image format. Common Image thresholding image includes the algorithm such as histogram and multi-level thresholding. Thresholding is when you classify the pixel values in an image. In OpenCV thresholding is done on grayscale images, which are image which have pixel values ranging from 0–255.

**Extraction Of Features**

The feature extraction from the image is the most important stage of pre processing that uses various algorithms and techniques to detect the nodules or clot or desired portion and shapes in the images. In image processing the feature extraction starts from an intial set of measured data and builds the derived values. The main goal of feature extraction is to get the how much area of the lung is affected [12]. For extraction of features from the image we have to focus on some features [12].

**4. PROPOSED SYSTEM**

The proposed will work in the following steps[13] -:

**STEP-1:** In very first step we collect the images of lung cancer from different cancer hospital/cancer database or over the internet [13].
STEP-2: Now access the particular images one by one into the matlab tool by giving commands to it [13].
STEP-3: Now the process of enhancement of images is carried out [13].
STEP-4: Segmentation process [13].
STEP-5: Feature Extraction [13].

Enhancement Processes of Images

Image enhancement is a technique that are aim for enhancing the image quality so that it would be interpreted clearly. This provides the results as the other image which demonstrates the features in a manner that are better than the original image in terms of there appearance and other factors.

Gabor’s Filter:

The Gabor filter is used for improvement of image quality and texture analysis and it is also called as linear filter. The images are presented which are based on function which is known as Gabor’s function which constitute an good local and multi-scale decomposition in terms of logon’s that are simultaneously localized in space and frequency domains.[14]. By using Gabor Filter the texture features of the image is extracted and it also provides excellent results for ultrasound and kidney images also. This filter is also used for edge detection system. For cancer detection in lungs in this 2D images are used and the response of impulse is recorded using the Gaussian filter function [13] modulus with sinusoidal function.

This is applied to the 2-Dimensional radiography images which are used for the enhancement of images of lung cancer. As in this we can see in the table below the diagram that the results of gabor filter is better than the other two methods which are discussed below that is Fast and Fourier method and auto enhancement method which all are used for the enhancement of images.

As per the recent paper we have viewed that the gabor filter technique is very efficient as seen in above table which is studied according to the previous paper as because the gabor filter is chosen because in there paper we studied that by using the gabor filter the new image that we get is good at it’s clearness and brightness and in terms of contrast as compared to above methods i.e Fast and fourier transform and Auto enhancement technique.

Image segmentation process

Image segmentation divides the larger images into smaller images which becomes easy to process and and identify lung cancer nodules. Segmentation is the process of making the black and white images into small segments which becomes easy for analysis. This process includes different types of operations like edge detection, watershed transform etc. The segmentation of image is done to make the image more meaningful and easier to analyze. The image segmentation is done to extract a tumor part from the cancer image and this is our ROI which we wanted and other is our unwanted area.

There are various methods for segmentation of images i.e Thresholding approach and other is Watershed segmentation approach and we are going to use this approach as because it gives the better results.

Marker-Controlled Watershed Segmentation Method

This is a technique which is overcome of the watershed algorithm as there is the disadvantage of the over segmentation in the watershed algorithm. The marker watershed algorithm is used to find the existence of the object at the specific location in the image. We use this method or approach because it has more accuracy and quality than the other approach like thresholding approach. It is of two types internal marker and external marker. It follows various steps:

1. First it identify the image color and then it convert it into black and white image.
2. Now the mathematical segmentation function is used to calculate the gradient magnitude.
3. Now in the picture it points out the object form the foreground object.
4. Now in the image it find out the marked point in the image
present in background of an image. It now find the watershed.

5. Finally the result from the image is obtained.

**5. RESULT ANALYSIS /DISCUSSION**

In the results part we come to know the abnormality and normality of the images. In this we are extracting the features from the images by using binarization and masking methods.

**BANARIZATION METHOD**

It is the approach in which the results are produced such as if in image the back pixels are more in the number than the white pixels than the image of lung is considered as normal. So on the basis of this approach we start counting a dark ppi to distinguish between the ordinary and unordinary images so that we can get average which is later used as a intensity value, now if a dark ppi are more in there count then the intensity value (threshold value) that we have find or kept in our study then it tells that the image should be considered as normal image otherwise the image is considered as abnormal image in that where the count of dark ppi is less than the intensity value(threshold value) . The intensity value which we take for our study as 17179.42[18]. Figure 2 shows the pixeliazation and binarization process and method.

<table>
<thead>
<tr>
<th>Patient No.</th>
<th>Black Px</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>22504</td>
<td>Image Normal</td>
</tr>
<tr>
<td>II</td>
<td>27661</td>
<td>Image Normal</td>
</tr>
<tr>
<td>III</td>
<td>19673</td>
<td>Image Normal</td>
</tr>
<tr>
<td>IV</td>
<td>11399</td>
<td>Cancerous Image</td>
</tr>
<tr>
<td>V</td>
<td>12865</td>
<td>Cancerous Image</td>
</tr>
</tbody>
</table>

Fig 3 Detection Results[18]

**6. IMPLEMENTATION**
So at last we can conclude that the lung cancer is a dangerous disease and it is necessary to detect it at a very early stage otherwise it increases in our body and at last the survival rate of the person becomes less. As this disease detection is a very tough job. So the above tools and methods that we have seen are used to find the tumor which is a difficult task but as the all things have limitations these tools. As our proposed methods have various methods like binarization, gabor filter and feature extractions which help us to find the tumor from the computerized tomography scan images. So at last we can say that the system achieve its desired results which we are making or using this. The future scope of this method can be used to find or locate the brain tumor, breast cancer etc.

7. CONCLUSION AND FUTURE SCOPE
We have seen for early disease detection and treatment stages the image improvement techniques are developing and while developing these techniques the time factor is taken into consideration and the accuracy of results so that these techniques can deliver there best for further processing and detection of cancer in the images. So the image quality and accuracy are two important factors. The techniques which are present used are X-ray, CT and at last MRI images.

As the if we talk about the treatments process it's include chemotherapy, surgery, radiation therapy and others therapy which are very time taking process in detection of lung cancer and lots of prone to the human error and also very costly and painful so to reduce this process to find the tumor the image processing techniques is taken into consideration which are efficient and less time taken and the best part of this is that is less prone to human errors and give accurate results that how much area of lung is affected through cancer and helps doctors to take proper precautions so that it would no grow more and stop there and give best treatment to there patient. CT scan images are used as these pictures are having low distortions as comparison to the other images like X-ray and MRI images. Now these acquired images are gone to various steps which are discussed in our proposed system and at each stage we get the best results. These stages helps physicians expert to diagnose the lung images and help to identifies a present phase of a tumor that at which stage it is Ist, 2nd etc. on the basis of some of its experiences.

Further this paper can be improved by introducing various tools and techniques like neural networks or using svm machines to detect the presence of lung cancer in humans which reduces more computation time than the time that is taking at present and complexity involved in this method.

REFERENCES


Practices and Technologies, Volume 11


