A Study of Image Segmentation and Classification of Lung Tumors

Vaibhavi Patel¹, Prof. Bijal Talati²
¹PG Student, SVIT, VASAD, Gujarat, India
²H.O.D., SVIT, VASAD, Gujarat, India

ABSTRACT

Lung Tumors is an important reason of death worldwide as it states to the uncontrolled progress of abnormal cells in the lung. If not treated, this growth can extend earlier the lung by process of metastasis into close-by tissue and different parts of the body. The image processing methods used commonly in various medical areas for enlightening prior detection and handling stages, in which the time span or elapse is very essential to classify the disease in the patient as possible as fast, especially in many tumors. The method uses first detection of lung mass tissue uses segmentation techniques. Geometrical features Extraction technique used for calculating statistical features. At last, classification used to classifies the tumor. In this paper includes various segmentation, feature types and classification methods with their merits and demerits.

Keywords: Image Processing, Tumor Detection, Tumor Classification, Segmentation, Computer Tomography (CT)

I. INTRODUCTION

Lung tumor is furthermore seen as carcinoma of the lung. Lung cancer in particular, is one of the major causes for cancer related deaths worldwide. [1] Cancer patient has the smallest survival rate after the finding the tumor, it continuous increases the in the sum of deaths each year. Being from lung Tumor is straightforwardly depending upon identification of tumor and at its discovery time. The early detection of the tumor will increase the survival rate of the cancer patient.[3] As indicated by an overview led by The Hindu in 2010 Cancer executed 5,56,400 individuals the nation over. The 30-69 age group accounted 71% (3,95,400) of the death. In 2010, growth alone represented 8% of the 2.5 million aggregate male passing and 12% of the 1.6 million aggregate female passing in this age gathering (30 to 69 years). At almost 23%, oral malignancy caused the most figure of passing among men. It was trailed by stomach malignancy (12.6%) and lung tumor (11.4%). Because of women, cervical malignancy was the driving reason (around 17%), trailed by breastcancer (10.2%). Consequently, from this review, it is apparent that the lung malignancy is a genuine explanation behind death what's more; its rate is expanding every year. [12]

![Figure 1. Death rate per year due to Lung Tumor](image)

The bigger section (85–90%) of occurrences of lung cancer are due to the tobacco smoke. Around 15–25% of cases occur in people due to the second hand smoke. Another reason of occurrence lung cancer is hereditary components and introduction to asbestos and different types of air contamination. [1]

Lung tumor can be detect by taking mid-section radio outlines and enlisted tomography (CT) analyzes. [1] CT lung image is mainly used for the detecting the lung
tumor nodules. After taking CT image in the diagnosis process consist following stage,

![Diagram of Lung Tumor Detection and classification system]

- **Image preprocessing** - in these process it enhance the picture quality and removes the unwanted noise and background information.
- **Image Segmentation** - Image segmentation is the procedure of isolating the lung lobes from CT image and it groups the suspicious region.
- **Feature Extraction** - After Segmentation, positive region will be extraction using the feature such as the perimeter, shape, size, area, color, etc.
- **Classification** - After identifying the cancerous nodule classifies the tumor types. [1][8][10]

### II. TUMORS TYPES

A tumor does not mean cancer. Lung tumor is the abandoned cells development of irregular cells that start on one or together lungs; typically it is the cells that line the airways. If not treated, this advancement of tumor can extent past the lung by practice of metastasis into near to tissue and diverse parts of the body. Most tumors that start in the lung, identified as basic lung malignancies, are carcinoma that got from epithelial cells. [1] A lung nodule is round lesion that is mainly two types cancerous or non-cancerous. [8]

![Figure 2. Tumors Types]

A tumor does not mean cancer. Tumors can be not cancerous is called benign or threatening cancerous called is malignant. Benign tumor cannot harmful to the human health and it cannot spare to the other part of body. [10]

![Figure 3. Benign Tumor]

Not at all like benign tumors, harmful ones develop quickly, they are eager, they search out new an area, and they spread. The unusual cells that frame a threatening tumor increase at a quicker rate. [10]
Main goal of segmentation is dividing and simplify the image into a meaningful structure, which is easy to understand. For the tumor detection segmentation of CT images in 2D, slice by slice has numerous valuable applications for the medical area, for example, representation and volume estimation of region of interest, location of anomalies. [3] Mainly two values of intensity is used in segmentation - discontinuity and similarity. Discontinuity defines as segment the image based on unexpected changes in intensity, for example, edges the images. Similarity classification depends on dividing the image into district regions that are comparable as per a predefined measure. In the medical application is segment the tumor part. It will identify the affected lung area with the help of these we differentiate the cancerous tumor and non-cancerous tumor. [2] Following table describes segmentation techniques with its merits and demerits, which is taken from survey.

### Table 1. Image Segmentation Techniques with pros and cons

<table>
<thead>
<tr>
<th>Segmentation Techniques</th>
<th>Description</th>
<th>pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gabor Filter[3,13]</td>
<td>Gabor function contain 1st and more than one-level decomposition in terms of logons areas. A Gabor filter is a linear with impulse response with the harmonic purpose grew by a Gaussian.</td>
<td>Segment text area and graphic area in simple background. Efficient Used in document Analysis.</td>
<td>Time consuming method. Gabor filter has Computational complexity which prevented its use in practice.</td>
</tr>
<tr>
<td>Watershed Segmentation[2,3]</td>
<td>Watershed segmented and Extracts seeds that show the vicinity of Questions alternately. Foundation toward particular picture. Areas need aid. Afterward situated will make territorial minima inside the topological surface.</td>
<td>Image split According to our demanded resolution. Handle sharp edges, topographical changes, 3D effects</td>
<td>Sensitive to noise Not suitable for edgeless image Note suitable for image whose boundary are very smooth</td>
</tr>
<tr>
<td>K-Means Clustering[1]</td>
<td>k- Means is unsupervised algorithm. Image clustering means groping the pixel according to some condition. Then k cluster center randomly chosen. The distance between the each pixel to a cluster centers are Calculated. The distance calculated.. Single pixel is compared to all cluster centers using the distance formula. The pixel is moved to parted to cluster which has shortest distance among all.</td>
<td>Scale Well Efficient Simple, understandable Items assign automatically in cluster.</td>
<td>Choosing the Wrong K All item forced into a cluster Sensitive with outlier</td>
</tr>
</tbody>
</table>
IV. FEATURE EXTRACTION

Image feature is the one pieces of information. In the image feature extraction is the process of transfer arbitrary data like image to the relevant numeric data. This numeric data used in the classification process. Mainly feature has two types - Shape and texture. Detailed view of feature is in the following table. [7,8]

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Advantage</th>
<th>Limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shape</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area, Perimeter, Major and Manor Axis</td>
<td>Easy to implement, Less Complex, Less Time Consuming</td>
<td>Works with Binary Image only, feature value change when change image dimension</td>
</tr>
<tr>
<td>GLCM Feature</td>
<td>Computation Time is Low, Low memory Consumption</td>
<td>Works with Gray scale images, feature vector is low so classification accuracy will less.</td>
</tr>
<tr>
<td>Haralick Features</td>
<td>Computational accuracy of feature vectors is high, Classification accuracy is high</td>
<td>Due to 13 features the computation of feature vectors is complex and time consuming.</td>
</tr>
</tbody>
</table>

V. CLASSIFICATION TECHNIQUES

Classification defined as the task categorizes the any of given objects within a given category called class. [8] Following table describes classification techniques with its pros and cons, which taken from survey.

<table>
<thead>
<tr>
<th>Classification techniques</th>
<th>Description</th>
<th>pros</th>
<th>cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive Bayes[3,7]</td>
<td>They are the statistical classifiers. It is based on probability theory. As outcome, it provides class membership with certain probabilities. Naive Bayesian classifier is based on Bayesian theory.</td>
<td>Database is less than great bias low alteration classifier like NB will work well. It is Easy to implement. Need less training data. simple yet powerful model They return not only the prediction but also the degree of certainty. No complicated optimization required. Small memory footprint</td>
<td>Database is less than procreative class will work well. It cannot learn interaction between features. No occurrences of a class label. If certain attribute value together then it results zero probability. Large dataset can’t use it.</td>
</tr>
<tr>
<td>Neural Network[1,6,18]</td>
<td>Inspired from the natural neural network of human nervous</td>
<td>High degree of non-linearity possible. Testing process is fast.</td>
<td>Complex to tune parameters. More time to build model.</td>
</tr>
<tr>
<td>System</td>
<td>Description</td>
<td>Advantages</td>
<td>Disadvantages</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Neural Networks</td>
<td>Consists of units (neurons), decided in layers, which convert an input vector into some output. Generally networks are defined as Feed forward, Back propagation, radial basis function, recurrent neural network, etc.</td>
<td>In complex domains, it provides good result. If domain is continues then it provides better result.</td>
<td></td>
</tr>
<tr>
<td>Decision Tree [11]</td>
<td>A structure mainly contains nodes and branches. Topmost nodes are root nodes. Other inner nodes are test nodes on an attribute and each leaf node holds a class label.</td>
<td>Split the data more correctly. Handle missing values. It does not need any domain knowledge. Easy to understand. Learning and classification steps of a decision tree are simple and fast.</td>
<td>Rules are too much simple. Needed more training dataset. Not much useful in practical approach.</td>
</tr>
<tr>
<td>SVM (Support Vector Machine) [5,8,11,14,17]</td>
<td>A Support Vector Machine is a binary classifier. It uses kernel function to transform low dimensional training samples to higher. Use quadratic programming to find the best classifier boundary hyper plane. It uses the kernel trick, so you can build in expert knowledge about the problem via manufacturing the kernel.</td>
<td>High accuracy. Easy to generate rules. Easy to understand. Most Effective methods in classification. Nice theoretical guarantee regarding over fitting. It will work well even data is not linearly divisible in the base feature space. SVM maximizes margin, so the model is more robust. SVM supports kernels, so you can design model for even nonlinear relations.</td>
<td>Hard to interpret. Take high time to expect the new instance. If several key parameters are present then only it achieves the finest classification result. Memory intensive</td>
</tr>
</tbody>
</table>

### VI. CONCLUSION

This paper provides various types of tumor segmentation, Feature types and classification techniques would give satisfactory results and help medical professionals with lung cancer diagnosis and ultimately save a lot of lives. Detail review includes description of different segmentation and classification methods containing its pros and cons.

Furthermore, we intended to survey of optimization algorithms and select suitable one and will be implementing through appropriate software. To increase accuracy of implemented algorithm, it will be optimize with suitable optimization technique.

### VII. REFERENCES

[1]. P. B. Sangamithraa, S.Govindaraju. Lung Tumour Detection and Classification using EK-Mean Clustering. In Full-text paper was peer-reviewed and accepted to be presented at the IEEE WiSPNET (2016).


[14]. WANG Xiao-peng, ZHANG Wen, and CUI Ying. Tumor segmentation in lung CT images based on support vector machine and improved level set.