RESEARCH ARTICLE OPEN ACCESS

A Comparative Study On preprocessing,Segmentation And Classification Techniques For Cancer Detection

Aarohi Trivedi*, Prof.Himani Trivedi**
**(Department of Computer Engineering LDRP Institute Of Technology And Research,*
*** (Lecturer, Department of Computer Engineering LDRP Institute Of Technology And Research,Gandhinagar*
Corresponding author; Aarohi Trivedi

ABSTRACT
Cancer is one of the diseases which is the reason of death all over the world. Among them lung cancer is most common type of cancer. Early detection of cancer is very important task for the survival of human being. Image processing is most commonly used technique in medical imaging. There are different steps involved in detection of cancer: Preprocessing, segmentation, feature extraction, classification. This paper presents a detailed survey of preprocessing, segmentation and Classification techniques that are useful for detection of cancer.
Keywords **-** Computer Aided Diagnosis(CAD),CT scan, Lung cancer

---

Date Of Submission:12-10-2018 Date Of Acceptance: 27-10-2018

I. INTRODUCTION

Cancer is the uncontrolled growth of abnormal cells. These cells may form a mass called a tumor. It can be malignant means cancerous tumor or benign means non-cancerous tumor. TABLE 1 shows the difference between these two types. There are many types of cancer. The most common type of cancer is lung cancer which has been found in both men and women. Lung cancer can occur due to many reasons one of the reason is cigarette smoking. About 90% of lung cancer can be caused due to tobacco usage. Also it can be caused by a combination of genetic factors and exposure to radon gas, asbestos, second-hand smoke, or other forms of air pollution. Current assessment given by recent data presented via health association reveals that due to lung cancer about 7.6 million demises all around the world. In addition, fatality coming out of tumor is headed for growing, and turns out to be around 17 million globally in 2030 [1].Though cancer is the common disease which is the reason of human deaths, to decrease the death rate early detection is necessary.

Medical imaging is a field in which there are different modalities like X-Ray, CT-scan, MRI are most common and useful technologies in the diagnosis of cancer. For detection of lung cancer, CT imaging technique is very popular and gives good results. It is having fast acquisition time for the image
with higher resolutions and then removing the unwanted things which is occurred by abdominal movements. Figure 1 shows the CT image of lungs. Through CT scan imaging modality radiologists can detect cancer but they cannot detect on early basis. Sometimes small nodules can be missed by human eyes. With the rapid advancements in the field of medical imaging radiologists can take a second opinion from the computerized analysis of different modalities.

Basically there are different steps associated with early detection of cancer. 1)Preprocessing 2)Region of Interest (ROI) identification 3)Tumor segmentation 4)Feature extraction 5)Classification. This paper gives detailed survey of preprocessing and segmentation methods.

Figure 1: Original CT image

II. LITERATURE REVIEW

Many researchers have proposed and implemented model regarding early detection of lung cancer. Shoji Kido, Yasusi Hirano,Noriaki Hashimoto [2] proposed a CAD model in which CNN is used as a classifier and Region based CNN(R-CNN) for detection of different objects

like tumor part using bounding boxes. The accuracy for classification is 95.2%,for diffuse lung diseases system has accuracy of 84.7%.The main advantage of the model is it does not require image feature extractor which is hard to implement instead it uses image itself. But the limitation of this model is to train CNN large dataset is required.

LilikAnifah,Haryanto,RinaHarimurti,Zaimah Permatasari [3] have used median filter to remove noise from the image and histogram equalization for enhancement. They have used ROI based segmentation technique to extract tumor part and classification is based on Backpropogation neural network. They have extracted features like contrast, correlation, homogeneity, variance, energy using GLCM method. The accuracy of the model is 80% which is still less.

Deep Prakash Kaucha, Abeer Alsadoon, A. Elchouemi, Sasikumaran Sreedharan [4] proposed a model whose accuracy is 95.16%,sensitivity is 98.21% and specificity is 78.69%. They have applied preprocessing to remove noise. After that they have first extracted different regions from lung image using ROI and then applied Discrete Wavelet Transform (DWT) as segmentation.

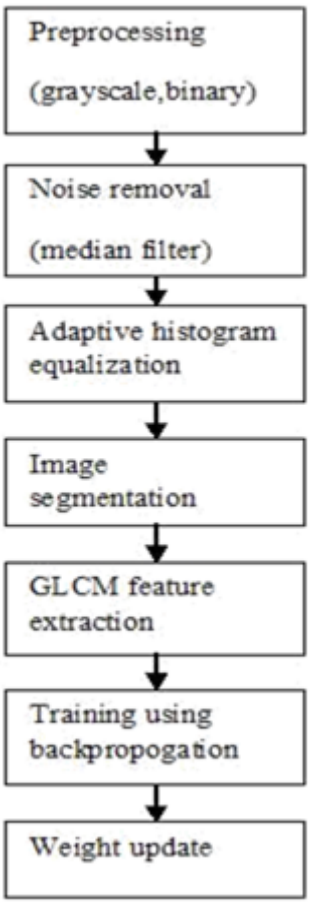

Figure 2: Proposed system [3]

DWT is having more memory and power consumption. Feature extraction is based on GLCM .SVM based classification to classify cancerous and non-cancerous tumors. Suren Makaju, A. K. Singh [5] proposed a model having accuracy of detecting cancer is 92% and classifier accuracy is 86.6%.They have used median filter and Gaussian filter to remove noise. Gaussian filter gives somewhat blurred output because it smoothens the image. Sometimes after applying Gaussian filter pixels having important details also get missed out due to smoothening. They have used watershed algorithm for segmentation which can separate touching objects so that it can help proper segmentation of cancer nodules if it is touching to other false nodules.For classification they have used SVM.The model doesn't justify stages of the cancer.

K.Gopi [6] proposed a model having an accuracy of 92.46%.They have separated lung lobes from CT image in preprocessing stage using binarization, thresholding, segmentation. But in some cases tumor touches the lung walls so separating lung lobes from original CT image can also remove some details of tumor pixels. They have used EK means clustering algorithm for tumor segmentation. But clustering is sensitive to initialization condition of cluster number and centre. They have extracted features like entropy, correlation, convexity using GLCM and used SVM as a classifier.

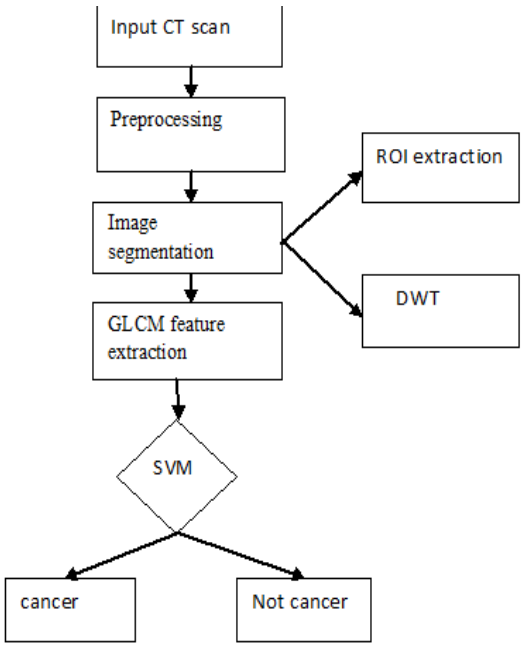

Figure 3: Proposed system [4]

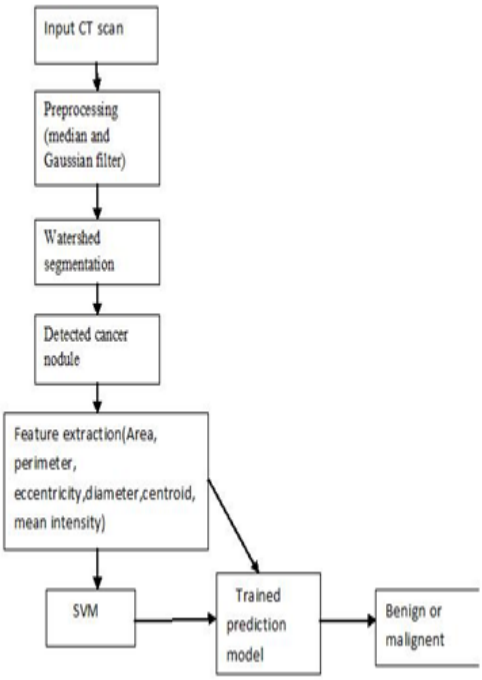

Figure 4: Proposed system [5]

III. IMAGE PREPROCESSING

Image preprocessing methods are useful for intensity improvement, noise removing, enhancement of an image. For enhancement of image histogram equalization, high boost filters are used.CT images are having some kind of noise. To remove noise from CT images three filters are found: Median filter, Gaussian filter, Wiener filter.

Comparative study of noise removing filters

a. Median Filter

Median Filter is the non-linear filter which replaces the value of every pixel by the median of intensity values in neighbourhood.Mainly median filter is useful in salt and pepper noise.

b. Gaussian Filter

Gaussian filter modifies the input image by convolution with a Gaussian function. It can be useful to remove speckle and Gaussian noise. Gaussian smoothing operator is a 2-D operator that is used to blur images and remove details and noise [7].

c. Wiener Filter

Wiener filter (a type of linear filter) to an image adaptively, tailoring itself to the local image variance If the variance is large, wiener performs little smoothing. If it is is small, wiener performs more smoothing and preserving edges and other high frequency parts of an image [8].

Figure 6: Original Image

Figure 7: Median Filtered

Figure 8: Gaussian Filtered

Figure 9: Wiener Filtered

TABLE 2
COMPARISION RESULTS OF DIFFERENT FILTERS

	Entropy	SNR	PSNR
Median filter	5.9371	10.7520	25.3110
Gaussian Filter	6.1870	17.8195	24.6942
Wiener Filter	6.7878	28.9782	35.8529

After comparing three filters we found that among three filters Gaussian filter blurs the resultant image so sometimes it may possible that it can miss some important details also.In median filtered image it can have sharp edges. If we apply median and wiener filter on the same image we can have an image which is sharply preserved.

IV. IMAGE SEGMENTATION

Segmentation plays an important role to find out particular interested area. It partitions the image into different regions to extract useful information. Figure 6 shows the taxonomy of segmentation techniques.

Edge Based segmentation

Edge based segmentation techniques are first find the edges by using different-different operators. Since an object can be represented by its edges.

1) **Sobel operator:**

The Sobel method of edge detection for image segmentation finds edges using the Sobel approximation to the derivative. It precedes the edges at those points where the gradient is highest

[9].It performs 2-D spatial gradient measurement on an image and so emphasizes regions of high spatial frequency that correspond to edges. The convolution mask of Sobel operator are as shown table 3, which are used to obtain the gradient magnitude of the image from the original [10].

TABLE 1 SOBEL OPERATORS

-1	0	1
-2	0	2
-1	0	1

1	2	1
0	0	0
-1	-2	-1

2) **Prewitt operator:**

The prewitt operator is an approximate way to estimate the magnitude and orientation of the edge [10]. Prewitt operator provides us two masks one for detecting edges in horizontal direction and another for detecting edges in an vertical direction shown in table.

TABLE 2 PREWITT OPERATORS

-1	0	1
-1	0	1
-1	0	1

1	1	1
0	0	0
-1	-1	-1

3) **Canny edge detector:**

The first step of canny detector is to smooth the image using Gaussian low-pass filter which smoothens also edges. Second step is to calculate gradients. Based on gradient magnitude and direction edges are sharpened. Calculation of the gradient is done by the Sobel kernel which is applied to each pixel in the image in both direction. Strength of the edge is determined by gradient magnitude while direction of the gradient is used to determine for each pixel the direction with the largest change in intensity [11].Only the pixel with maximum edge strength will be preserved, otherwise it will be suppressed through a process on non maximum suppression for edge thinning. When sharpening is finished reduction in edge numbers is performed using double thresholding to separate strong, weak and suppressed edges [11].

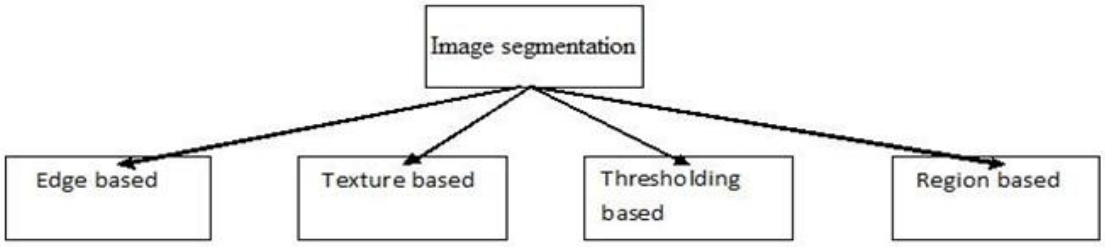

Figure 10: Image segmentation techniques

4) **Discrete Wavelet Transform(DWT):**

DWT is a technique used to extract features of each image from image, which extracts maximum highlighting pixels present in images to progress results.When we apply DWT on an image, it divides the image into frequency components. They are called approximation coefficients (LowLow or LL), horizontal (LowHigh or LH), vertical (HighLow or HL) and detail coefficients (HighHigh or HH). The low frequency

components are approximate coefficients holding almost the original image i.e. Tumor area and high frequency components are detailed coefficients holding additional information about the image [12].Low frequency corresponds to just the structural or predominant information of the image while high frequencies corresponds to the edges of the image. These detailed coefficients can be used to analyze the image.

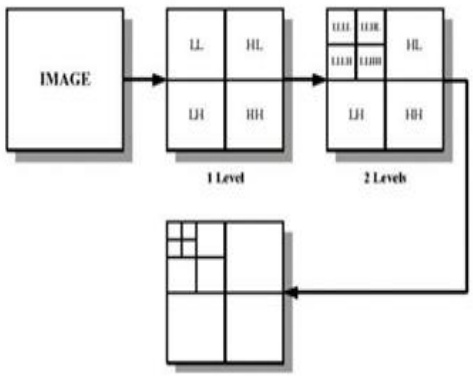

Figure 11: 2-D DWT block diagram [13]

Watershed based segmentation

Watershed segmentation is a gradient-based segmentation technique. It considers the gradient map of the image as a relief map . The

watershed transform figures catchment basins and ridgelines (otherwise called watershed lines), where catchment basins relating to image regions and ridgelines identifying with region boundaries [14]. The basic idea of watershed was placing source of water in each regional minima, after sometime different regional minima gets merged, in order to solve this barriers would be built. Watershed algorithm is useful when there are touching objects. The main drawback of watershed algorithm is oversegmentation.To overcome the oversegmentation.To problem many algorithms have been proposed like watershed by markers.

Thresholding based segmentation[15]

This is simplest image segmentation technique for partitioning images directly into regions based on intensity values with one or more thresholds.Thresholding can be of two types: local Thresholding and global Thresholding. Segmentation of images having more than two kind of regions corresponding to different objects regarded as local Thresholding .

$$g(x,y) = \begin{cases} 1 & \text{if } f(x,y) \geq TH \\ 0 & otherwise \end{cases} \quad (1)$$

Pixels with value 1 corresponds to region of interest (ROI) whereas remaining pixels that are set to zero corresponds to background of image. This type of thresholding technique is known as global Thresholding. Otsu's method is one of the global Thresholding method.

Region based segmentation

Region based segmentation techniques segments an image into different regions based on some criteria.

1) **Region Growing:**

This algorithm initially starts with the seed value. Seed point is selected by the user based on some criteria. For example if a criteria is pixel intensity threshold value then the knowledge of histogram of image would be useful to select seed point. After selection of seed points it examines neighbor pixels and finds the most similar one and if it satisfy some membership criteria it is included in the region.[16]The membership criteria could be gray level, pixel intensity, color etc[17].This process iteratively examines all unallocated neighboring pixels and add them to the region .The difference between pixel intensity value and region mean is used as a similarity measure. The pixel having smallest difference would be added to the region.

V. CLASSIFICATION

Once the tumor is detected and segmentated it has to be checked whether it is benign or malignant. For justification of cancerous or non-cancerous tumor a classifier is required. Classifier can be supervised or unsupervised. To train a classifier features are used. Features are in terms of textural features, histogram features, geometrical features,GradientandSpatialfeatures[18].Textural featureslike entropy,contrast,energy, etc.Histogramfeatureslikehistogram,standarddeviation, skewness, etc.[18]

1) **SVM classifier**:

SVM is a supervised learning classifier which classifies two classes by creating hyperplane in high dimensional feature space[19].Constructing a hyperplane having largest distance to the nearest point in any class gives good separation. The vector value for the neighbor boundaries is known as supporting vectors and that calculated distances from that vector and hyper-plane is called as the margins.

- Linear SVM [19]: It is the simplest one, in which the training patterns are linearly separable. A linear function of the form is given below:
  $f(x) = w^T X + b$ (2)
- Non Linear SVM [19]:In a nonlinear SVM classifier, a nonlinear operator is used to map the input pattern x into a higher dimensional space H. The nonlinear SVM classifier is defined as:
  $f(x) = W^T \langle I \rangle(x) + b$ (3)

Linearly separable data can be analyzed by hyperplane whereas linearly non-separable data can be analyzed by different kernel functions like higher order polynomials and quadratic. The few advantages are:

- It Contains a nonlinear transformation.
- It provides a good generalization capability.
- The problem of over fitting is eliminated.
- Reduction in computational complexity.
- Simple to manage decision rule complexity and Error frequency.

Suren Makaju, A. K. Singh [5] have classified cancerous and non-cancerous tumor using SVM having classifier accuracy is 86.6%. Deep Prakash Kaucha, P.W.C. Prasad, Abeer Alsadoon, A. Elchouemi, Sasikumaran Sreedharan [4] used SVM classifier and the accuracy was 95.16% and sensitivity of 98.21% and specificity of 78.69%.

2) **Convolutional Neural Network(CNN):**

Convolutional neural network is a type of deep learning algorithms. CNN is having an input layer, an output layer and multiple hidden layers. The hidden layer is mainly having three layers: Convolutional layer, max pooling and fully connected layers.Convolutional layers apply a convolution operation to the input, passing the result to the next layer. After the convolution process, an image downsampling is performed by the max-pooling layer. While such down-sampling can be done by averaging, empirical results have suggested that by taking the maximum in each sub-region can yield better performance in most cases [20]. Fully connected layers connect every neuron in one layer to every neuron in another layer.CNN does not require feature extractor it only takes image itself as an input. The deep CNN uses stochastic gradient descent (SGD) to update the weight parameters [20]. Hirano,Noriaki Hashimoto [2] used CNN for classification and accuracy was 95.2%.

3) **Backpropogation Neural Network[21]:**

During the design of neural network weights should be given by user. Sometimes after assigning weights model gives an output but that is not the actual output. In this case we need to somehow explain the model to change the parameters (weights), such that error becomes minimum. The Backpropagation algorithm looks for the minimum value of the error function in weight space using a technique called the delta rule or gradient descent. The weights that minimize the error function is then considered to be a solution to the learning problem.

Backpropogation is having one hidden layer. There may be two passes: forward pass and backward pass. In forward pass Backpropogation simply calculates the output and checks whether it is same as actual output or not. If the actual output is not equal to desired output than the network is adjusted by modifying the weights and calculates error in backward pass. This way model will try to reduce the error by changing the values of weights and biases. Some disadvantages found in Backpropogation:

- Can be sensitive to noisy data and outliers
- Training time for BPNN is high
- Structure of the BPNN is highly complex.

LilikAnifah,Haryanto,RinaHarimurti,Zaimah Permatasari [3] have used Backpropogation NN as a classifer for which they have extracted features like homogeneity, energy, contrast, correlation, and variance having an accuracy of 80%.

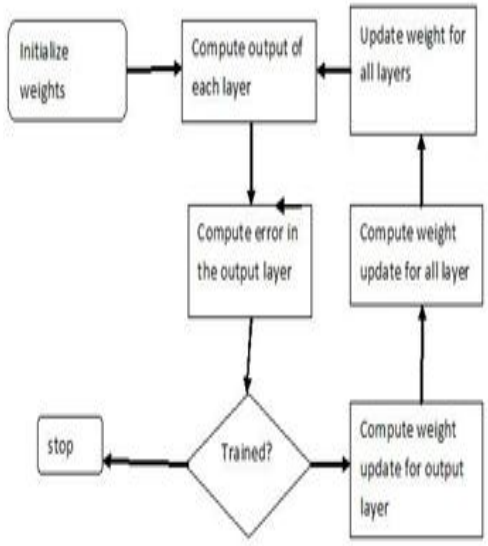

Figure 12: Backpropogation Training Algorithm[21]

Table 3 Comparison between different segmentation techniques

Technique	Description	Advantage	Disadvantage
Edge based	based on discontinuity detection	• Good contrast between objects	• not suitable for wrong detected or too many edges
Watershed algorithm	It consider image as a topological surface having picks and valleys	• Detects continuous boundaries • good results for high frequency image	• Over segmentation • complex calculation of gradients
Thresholding based	based on the histogram peaks of the image to find particular threshold values	• No need of previous information • Computationally inexpensive • Can be used in real time applications	• Dependent on peaks of histogram • For image having broad valleys or without peak it doesn't work well • Noise sensitive • Wrong selection of threshold can result into under or over segmentation
Region based	based on partitioning image into homogeneous regions	• Flow from inner point to outer regions generates clear object boundaries • Proper selection of seed value gives better result than other methods.	• Quite expensive in terms of memory and time • To define stopping criteria is difficult. • Selection of improper seed can get flawed segmentation

VI. CONCLUSION

In this paper we have described different stages of detecting lung cancer. The preprocessing, segmentation and classification are the major components for CAD system. We have discussed different methods for preprocessing, segmentation as well as classification. These methods are helpful for radiologists for their opinion. This review gives an overview of current algorithms that are used for detection of cancer.

REFERENCES

[1]. Sheenam Rattan,Sumandeep kaur,Nishu kansal,Jaspreet kaur,"An Optimized Lung Cancer Classification System For Computed Tomography Images",IEEE, 2017 Fourth International Conference on Image Information Processing (ICIIP).

[2]. Shoji Kido,Yasusi Hirano,Noriaki Hashimoto,"Detection and Classification of Lung Abnormalities by Use of Convolutional Neural Network (CNN) and Regions with CNN Features (R-CNN)",IEEE,2018.

[3]. Lilik Anifah,Haryanto,Rina Harimurti,Zaimah Permatasari," Cancer Lungs Detection on CT Scan Image Using Artificial Neural Network Backpropagation Based Gray Level Coocurrence Matrices Feature",IEEE,2017.

[4]. Deep Prakash Kaucha, Abeer Alsadoon, A. Elchouemi, Sasikumaran Sreedharan," Early Detection of Lung Cancer using SVM Classifier in Biomedical Image Processing", IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI-2017).

[5]. Suren Makaju, P.W.C. Prasad, Abeer Alsadoon, A. K. Singh, A. Elchouemi," Lung Cancer Detection using CT Scan Images", Elsevier,2017,107–114.

[6]. K.Gopi," Lung tumor Area Recognition and Classification using EK-Mean Clustering and SVM", 978-1-5090-5913-3/17,IEEE,2017.

[7]. Parminder Pal Kaur,Tejinderpal singh,"Speckle Noise Reduction in Ultrasound Images:Performance Analysis And Comparision",International Journal Of Engineering Research and Technology,Vol.3,ISSN:2278-0181.

[8]. Abdalla Mohamed Hambal, Dr. Zhijun Pei, Faustini Libent Ishabailu," Image Noise Reduction and Filtering Techniques", International Journal of Science and Research (IJSR), Volume 6,ISSN: 2319-7064.

[9]. Y.Ramadevi, T.Sridevi, B.Poornima, B.Kalyani," SEGMENTATION AND OBJECT RECOGNITION USING EDGE DETECTION TECHNIQUES", International Journal of Computer Science & Information Technology (IJCSIT), Vol 2, No 6.

[10]. Aarti Maan, Dr. Sukhvir Singh," A Review Paper on Image Segmentation Using Edge Detection Techniques and Threshold in MATLAB", Journal of Network Communications and Emerging Technologies (JNCET), Volume 7, Issue 6, ISSN: 2395-5317.

[11]. Marina Nikolic, Eva Tuba and Milan Tuba," Edge Detection in Medical Ultrasound Images Using Adjusted Canny Edge Detection Algorithm",IEEE, 2016.

[12]. M.RamaKrishna,Fahmeeda,G.DaizyFlorence,K.Sravani ," Brain Tumor Image Segmentation Based On Discrete Wavelet Transform and Support Vector Machine", International Journal for Modern Trends in Science and Technology Volume: 03, Special Issue No: 02, ISSN: 2455-3778.

[13]. A. Rezaeian , S. Homayouni , A. Safari," SEGMENTATION OF POLARIMETRIC SAR IMAGES USIG WAVELET TRANSFORMATION AND TEXTURE FEATURES", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-1/W5, 2015, doi:10.5194.

[14]. Pooja Thakur,Dr. Kuldip Pahwa,Dr. Rajat Gupta," BRAIN TUMOR DETECTION, SEGMENTATION USING WATERSHED SEGMENTATION AND MORPHOLOGICAL OPERATION", International Journal of Advanced Research in Electronics and Communication Engineering, Volume4, Issue 6, ISSN: 2278 – 909X.

[15]. Muhammad Ali Qadar, Yan Zhaowen," Brain Tumor Segmentation: A Comparative Analysis", National Natural Science Foundation of China (NSFC).
[16]. Nihad Mesanovic , Mislav Grgic , Haris Huseinagic , Matija Males , Emir Skejic , Muamer Smajlovic ," Automatic CT Image Segmentation of the Lungs with Region Growing Algorithm",Researchgate,2015.
[17]. Vinay Parmeshwarappa,Nandish S,"Segmentation of Lung Cancer using Image Enhancement Techniques and Region Growing
[18]. Algorithm",International Journal of Engineering Research and Technology(IJERT),ISSN:2278-0181,vol.3 Issue 2,2014.
[19]. G.Niranjana," A Review on Image Processing Methods in Detecting Lung Cancer using CT Images", 2017 International Conference on Technical Advancements in Computers and Communications,IEEE, DOI 10.1109.
[20]. Hari Babu Nandpuru," MRI Brain Cancer Classification Using Support Vector Machine", 2014 IEEE Students' Conference on Electrical, Electronics and Computer Science.
[21]. Xulei Yang, Si Yong Yeo, Jia Mei Hong, Sum Thai Wong, Wai Teng Tang, Zhen Zhou Wu, Gary Lee, Sulin Chen, Vanessa Ding, Brendan Pang, Andre Choo, Yi Su," A DEEP LEARNING APPROACH FOR TUMORTISSUEIMAGECLASSIFICATION",Researc hgate,2016,DOI: 10.2316/P.2016.832-025.
[22]. Md. Ali Hossain, Md. Mijanur Rahman, Uzzal Kumar Prodhan, Md. Farukuzzaman Khan," Implementation Of Back-Propagation Neural Network For Isolated Bangla Speech Recognition", International Journal of Information Sciences and Techniques (IJIST) Vol.3, No.4, July 2013, DOI: 10.5121.

Aarohi Trivedi "A Comparative Study Onpreprocessing,Segmentation And Classification Techniques For Cancer Detection "'International Journal of Engineering Research and Applications (IJERA) , vol. 8, no.10, 2018, pp 01-08