International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-8, June 2020

Recognition of Lung Tumor Area Based on Watershed Segmentation and CNN

M. Rama Bai, G. Meghana

Abstract: ***Early recognition of tumor would assist in saving an enormous number of lives over the globe frequently. Study and remedy of lung tumor have been one of the greatest troubles faced by humans over the latest few decades. Effective recognition of lung tumor is a vital and crucial aspect of image processing. Several Segmentation methods were used to detect lung tumor at an early stage. An approach is presented in this paper to diagnose lung tumor from CT scan images. The input image (CT scan image) will be preprocessed initially using median filter to remove the noise. After applying preprocessing technique, the Dual-Tree Complex Wavelet Transform (DTCWT) segmentation technique is used for the edge detection. The Gray-Level Co-occurrence Matrix (GLCM) features are calculated based on the pixel values of the extracted image. These features can be compared with database images using Convolutional Neural Network (CNN) which facilitates in categorizing it as tumorous. After confirming that the affected area is tumorous, watershed segmentation algorithm is used to get the color features of the tumor.***

Keywords: Preprocessed, DTCWT, GLCM, Convolutional Neural Network, Watershed Segmentation Algorithm

I. INTRODUCTION

Lung tumor is the end result of cellular dying in lung tissue. Lung tumor is a lethal disease in the evolving countries. Many imaging techniques are used for detecting lung tumor like Computed Tomography (CT) [5] and Magnetic Resonance Imaging (MRI).

The possibility of survival at the intense stage is less whilst compared to the treatment and lifestyle to survive cancer remedy whilst identified on the early stage of the tumor. By implementing image processing techniques, diagnosis system and guide analysis may be improved. A variety of scrutinize on the image processing techniques [1] for detection are available. But the identity of tumor is not greatly improved.

Neural network plays a crucial role within the recognition of the tumor cells amongst the normal tissues, which in turn affords a productive tool for detecting tumor. The tumor treatment will be effective only while the tumor cells are exactly segregated from the ordinary cells and training of the neural network using algorithm. CNN is used to classify the lung tumor.

Revised Manuscript Received on June 30, 2020.
***** Correspondence Author

Dr. M. Rama Bai, Professor, Dept. of Computer Science and Engineering, Mahatma Gandhi Institute of Technology, Hyderabad, India. mramabai_cse@mgit.ac.in

Ms G. Meghana*, Computer Science & Engineering department, Mahatma Gandhi Institute of Technology, Hyderabad, India. gaddammeghana2402@gmail.com

II. RELATED WORK

Classifying the lung tumors as tumorous has been done by diverse methodologies. The Deep Neural Network consists of more than one hidden layers consisting of convolutional layer is known as CNN. To get a normalized layer, pooling layer and fully connected are used. In the convolutional layer the weights of the CNN is used minimizing the memory footmark and will increases the performance of the network. The important characteristics of CNN lie with the 3D volumes of neurons, shared weights and local connectivity. By the usage of convolution layer throughout convolution of different sub regions of the image which is taken as an input with a learned kernel, a feature map is produced.

In pooling layer, a region of the image is selected and the pixel with maximal value among them is selected as the illustrative pixel so that a single scalar value by lowering a 2x2 or 3x3 grid. This results a massive depletion within the sample size. In conjunction with the convolutional layers concerning the output stage, the Fully-Connected (FC) layer can be used.

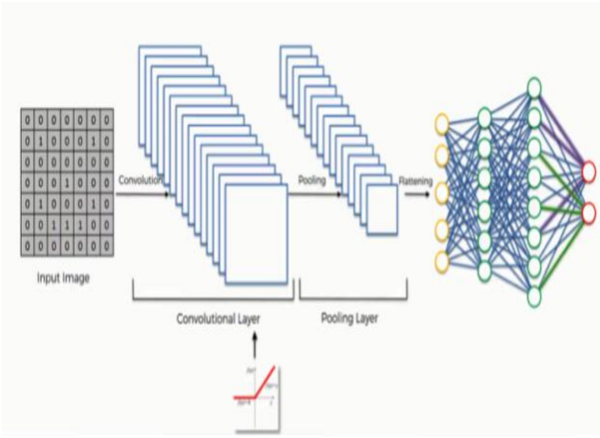


Fig 1. Convolutional Neural Network

The input is an image given to a convolutional layer. There are n filter kernels and can vary for each kernel in convolutional layer, that are convolved with the input image to provide l feature maps as shown in fig 1. Each map is then sub sampled with mean or max pooling is put in advance the sub sampling layer. Another method is the CAD (computer Aided Diagnosis). CAD of lung CT image has been an absolute step within the early identifying of lung diseases. The finest technique of applying computer aided recognition for medical image scrutiny is first to preprocess the image so as to segment it. In CAD of lung computed tomography patient image is usually to first segment the area of interest and then scrutinize individually each vicinity acquired, for a tumor.

This is usually much less difficult way, due to the fact the region used for setting the proper recognition, is getting little with the approach of segmentation, so the obstetrician can focus his observation most effective on precise records inside the precise region.

Lung segmentation approach is used to precisely segment the lung tissue of lung CT images, which can help obstetrician in early recognizing lung diseases.

III. METHODOLOGY

The system basically takes the input images as CT scan images. Then practice preprocessing for better image than the input image by the conversion RGB to Gray to B/W. Then for the segmentation processes, DTCWT is used for the edge detection processes and deliver the feature extraction processes by using GLCM for the identification of the problem as shown in fig 2.

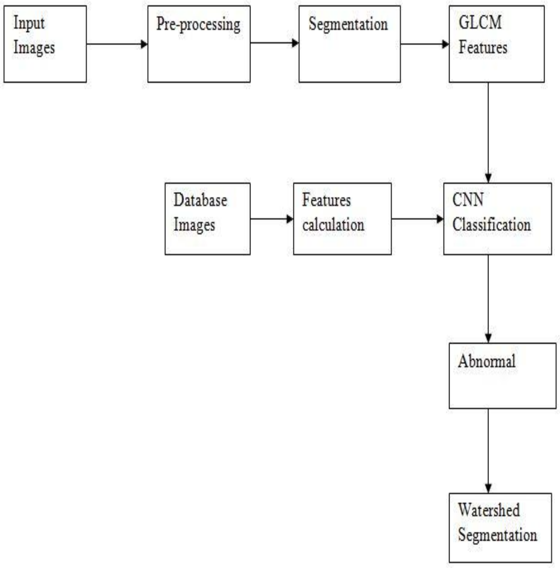


Fig 2. System Architecture

The Gray-Level Co-occurrence Matrix features are extracted from the processed image to form feature vector. These features can be compared with database images using classifier as neural networks. After identifying the effected disease as tumorous, watershed segmentation is used to get color features.

IV. PROPOSED WORK

The techniques used within the proposed system are as follows:

A. Preprocessing techniques:

Preprocessing is a method that is used for operations with images at the lowest degree of extraction each input and output are meaningful images. Suppressing undesirable distortions or enhances some image features of an image is crucial for further processing is the principle aim of preprocessing. The pre-processing techniques are resize, conversion, filtering.

Out of all the pre-processing techniques, filtering method is used to get rid of the noise from the desired image. Filtering techniques are mean filtering, median filtering.

Median Filtering:

The median filtering is used for disposing of noise from the image. Such noise depletion is a traditional pre-processing step to upgrade the outcomes of later processing like edge detection on an image. In digital image processing, median filtering is used under specific conditions, it preserves edges while eliminating noise. The following equation 1 shows the median filter process.

```
T = imread('file');
T = im2double(T);
[a b] = size(T);
Md = [];
for i=2:a-1
  for j=2:b-1
      Md(1) = T(i-1,j-1);
      Md(2) =T(i-1,j) ;
      Md(3) = T(i-1,j+1);
      Md(4) = T(i,j-1);
      Md(5) = T(i,j+1);
      Md(6) =T(i+1, j-1);
      Md(7) = T(i+1,j);
      Md(8) = T(i+1,j+1);
      T(i,j) = median(Md);
  end
end
```
…. (1)

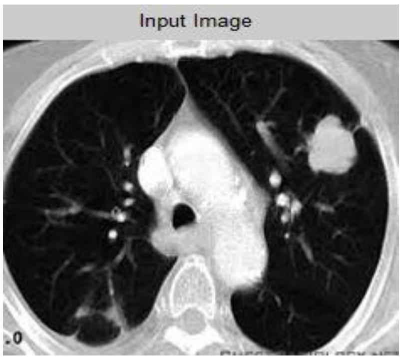


Fig 3. Input Image

The above mentioned fig 3 is the input image.

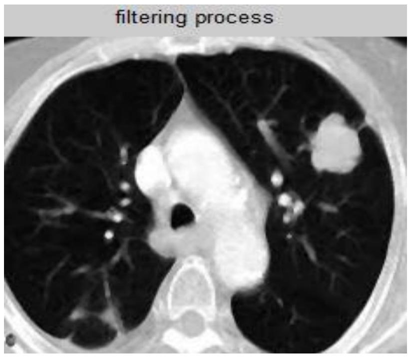


Fig 4. Median Filtering

The above mentioned fig 4 is the output for fig 3.

B. CT Scan Images:

In this model the input images are taken as the CT scan images. CT scan image provides more data when as compared to X-ray images. CT scan is nothing but Computed tomography scan. CT scan image provides the detailed statistics of images of internal organs.

The detailed statistics of internal organs that cannot be visible in conventional X-rays are found out by the CT scan. Comparing to conventional X-ray the dosage of the radiation of a CT scanner is the best one, but the details received from a CT scan are regularly a lot considerable.

C. Segmentation:

The process which is used for dividing a digital image into multiple segments is known as segmentation. To alternate the depiction of an image into something this is more significant and less difficult to examine is the main objective of the segmentation. Image segmentation is commonly used to find objects and obstacles in images. Each of the pixels in a location is indistinguishable with few characteristics.

Dual Tree Complex Wavelet Transform:

DTCWT is applied after preprocessing the image. DTCWT is mainly used to detect the edges. Calculating the complex transform of a signal using two distinct DWT decompositions is the dual-tree complex wavelet transform. The aggregate of traits of two DWT's is DTCWT.

```
[LL LH HL HH] = dwt2(mfima,'db');
bb = [LL LH;HL HH];
[LL1 LH1 HL1 HH1] = dwt2(LL,'db');
[LL2 LH2 HL2 HH2] = dwt2(LL1,'db');
[LL3 LH3 HL3 HH3] = dwt2(LL2,'db');
bb1 = [LL3 LH3;HL3 HH3];
bb2 = [bb1 LH2;HL2 HH2];
bb3 = [bb2 LH1;HL1 HH1];
bb4 = [bb3 LH;HL HH];
imshow(bb2,[]);                                    …. (2)
```

The above referred to equation 2 is for the DTCWT segmentation.

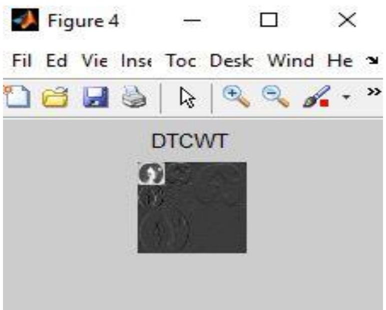


Fig 5. DTCWT

The above cited fig 5 is the output for the fig 4.

D. Gray-Level Co-occurrence Matrix Features:

After segmenting the image by using DTCWT, GLCM features are calculated based at the pixel values. The gray-level co-occurrence matrix (GLCM) is a technique used for scrutinize texture that considers the spatial affiliation of pixels of the image. The other name for GLCM is the gray-level spatial dependence matrix.

Identifying the character of an image by calculating how the set of pixel with precise values and in a selected position arise in an image by means of developing a GLCM, after which extricating numerical values from this matrix which can be utilized by the GLCM functions.

```
cnntest(qfeat,dfeatures);                          …. (3)
```

The above cited function is used for comparing the features of trained and the test images.

E. Convolutional Neural Network:

CNN compares the input image and dataset images and it detects the image as tumorous one. The set of rules which assigns the importance to different objects within the image and capable of specifying one from the other by means taking an imput image is known as convolutional neural network (ConvNet/CNN). Comparing to other classification algorithms the pre-processing needed in a ConvNet is much lower. ConvNets have the capability to learn essential techniques, filters with sufficient training the filters/characteristics.

```
net.ffc=zeros(enum,1);
net.ffv=(rand(enum,gvnum)-0.5)*2*sqrt(6/(enum+gvnum));                          …. (4)
```

As the above mentioned, gvnum represents the number of output neurons at the last layer, ffc represents the weights of the output neurons, ffv represents the weights between the last layer and the output neuron. The last layer is understood for the fully connected to the output layer.

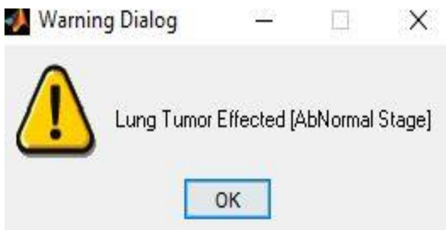


Fig 6. Warning Dialog

The fig 6 is the output for the fig 5.

F. Watershed Segmentation:

Watershed Segmentation is used to get the color features of the image. The area which is effected may be highlighted by using watershed segmentation as shown in the fig 3. A watershed is used for the modification defined on a grayscale image. For watershed there are exclusive technical definitions. Watershed lines may be defined on the nodes, at the edges, or on both nodes and edges utilized in graphs. In image processing, Watershed algorithm is used essentially for segmentation purposes.

```
L = watershed(gmag);
Lrgb = label2rgb(L);
imshow(Lrgb)                                       …. (5)
```

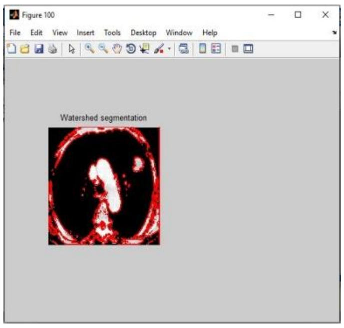


Fig 7. Watershed Segmentation

The above mentioned fig 7 is the output obtained.

V. RESULTS

Model is trained by giving a complete data on which CNN can be done. The below mentioned image composed of a set of CT scan images. The CT scan images are taken as input images as shown in fig 8 and techniques mentioned inside the proposed work are applied to identify the tumor.

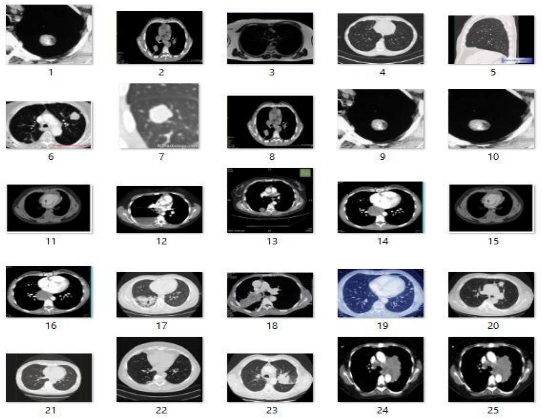
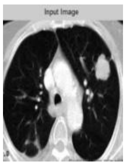


Fig 8. Dataset sample for CT scan images

After identifying the effected disease as tumorous, the tumorous image is segmented by using watershed segmentation to get color features of tumor after. CNN are applied for the recognition of lung tumor.

The following table I and table II shows the positive and negative test cases.

Table-I: Positive Test Cases

Sample Input	Expected Output	Actual Output
	Tumor Segment Area: 25.5016 Sensitivity: 50 Specificity: 87.5637 Accuracy: 87.5623 MSE: 9.8742 PSNR: 38.1858 Entropy: 1.4507 Correlation: 0.8009	Tumor Segment Area: 25.5016 Sensitivity: 50 Specificity: 87.5637 Accuracy: 87.5623 MSE: 9.8742 PSNR: 38.1858 Entropy: 1.4507 Correlation: 0.8009

As above noted contents the expected output suits with the actual output. So it's far taken into considered as positive test case.

Table-II: Negative Test Case

Sample Input	Expected Output	Actual Output
	Tumor Segment Area: 25.5016 Sensitivity: 50 Specificity: 87.5637 Accuracy: 87.5623 MSE: 9.8742 PSNR: 38.1858 Entropy: 1.4507 Correlation: 0.8009	Tumor Segment Area: 14.5828 Sensitivity: 50 Specificity:75.8162 Accuracy:75.8154 MSE: 11.3476 PSNR: 37.5818 Entropy: 0.1853 Correlation: 0.1561

$$tarea = (sqrt(Pcount))* 0.264; \quad .... (6)$$

The equation 6 is for calculating tumor segment area.

$$Sen = (Tp/(Tp + Fn)) * 100; \quad .... (7)$$

$$Spec = (Tn/(Tn + Fp)) * 100; \quad .... (8)$$

$$Accuracy = ((Tp + Tn)/(Tp + Tn + Fp + Fn)) * 100; \quad .... (9)$$

The above noted are used for finding out sensitivity, specificity and accuracy. These values are calculated primarily on actual values and predicted values. The fig 9 is the output for the input image (fig 3) which shows the area which is affected using watershed segmentation and the fig 10 shows the different attributes derived from the output image.

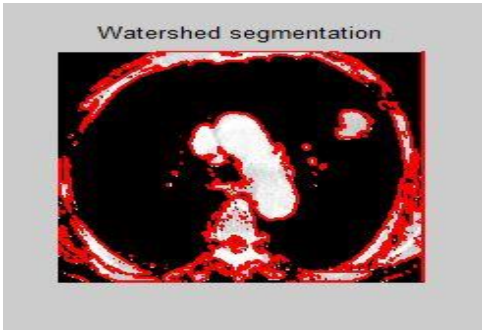


Fig 9. Effected area

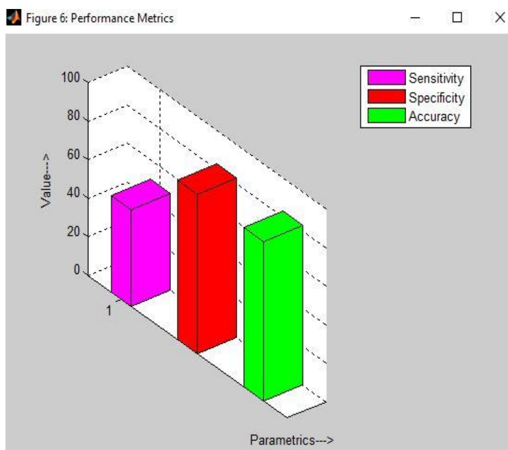


Fig 10. Attributes with the values

The following table III shows the parameters and their obtained values.

Table III: Parameters and Values

S. No	Parameters	Values
1	Tumor Segment Area	25.5016
2	Sensitivity	50
3	Specificity	87.5637
4	Accuracy	87.5623
5	MSE	9.8742
6	PSNR	38.1858
7	Entropy	1.4507
8	Correlation	0.8009

VI. CONCLUSION

Diagnosing lung tumor from CT scan images is not a new hassle but the arrival of Convolutional Neural Network and Watershed Segmentation Algorithm (WSA) has shifted the goalpost of this problem to a new level. The proposed system has substantially shown that CNN and WSA is the most expeditious way to recognize early identification of lung tumor diseases. Sample set of images are fed as input to the trained model and the model is capable of notify the existence of tumor and locate the tumor region in the images. The procedure involves preprocessing, segmentation, CNN, WSA, feature extraction in building a model for successful identification of the tumor spot.

REFERENCES

1. Anita Chaudhary, Sonit Sukraj Singh, Lung Cancer Detection on CT Images(2012).
2. http://globocan.iarc.fr/Pages/fact_sheets_ most cancers.aspx.
3. BC Lee, JB Keeland, PT Cahill, “MR recognition of supratentorial tumors”.
4. S Yohe, IT Yeh, “Missed diagnosis of phyllodes tumor on breast biopsy: pathologic clues to its recognition”.
5. Y Yu, Y Yao, H Yan, R Wang, “A tumor specific microRNA recognition system facilitates the accurate targeting to tumor cells by magnetic nanoparticles”.
6. SA Patil, MB Kuchanur, “Lung cancer classification using image processing”.
7. Mir Rayat Imtiaz Hossain, Imran Ahmed, “Automatic Lung Tumor Detection Bsed on GLCM Features”.
8. L Li, Y Wu, Y Yang, L LI, B Wu, “A new strategy to detect lung cancer on CT images”.

AUTHORS PROFILE

Dr M Rama Bai, is working as Professor in Department of CSE at MGIT, Hyderabad, India. She has published many papers in international journals. Her research areas of interest include Image Processing, Machine Learning, Artificial Intelligence and Data Science.

Ms G. Meghana currently pursuing B.Tech final year in Computer Science & Engineering at MGIT, Hyderabad, India. Her areas of interest include Image Processing and Data Science.