International Journal of Electrical Electronics & Computer Science Engineering
Volume 6, Issue 4 (August, 2019) | E-ISSN : 2348-2273 | P-ISSN : 2454-1222
Available Online at www.ijeecse.com

Detection and Classification of Lung Cancer Stages using Image Processing Techniques

Nwe Ni Kyaw[1], Kyaw Kyaw Naing[2], Phyu Myo Thwe[3], Kyawt Kyawt Htay[4], Hanni Htun[5]
[1, 5]Assistant Lecturer, [4]Lecturer, Faculty of Computer Science, Myanmar Institute of IT, Mandalay, Myanmar
[2]Lecturer, Physics Department, Shwebo University, Shwebo, Myanmar
[3]Assistant Lecturer, Faculty of Computer Science, Computer University, Kalay, Myanmar
[1]nwe_ni_kyaw@miit.edu.mm, [2] kyawkyawnaingshwebo2007@gmail.com, [3]phyumyothwe1@gmail.com,
[4]kyawt_kyawt_htay@miit.edu.mm,[5]hanni_htun@miit.edu.mm

Abstract: In current days, image processing techniques are widely used in many medical areas for improving earlier detection and treatment stages, especially in various cancer nodules such as the lung cancer, breast cancer, brain cancer and so on. This paper shows the detection and classification of lung cancer stages based on CT Scan Images. The median filter algorithm is used for image processing. In this paper, morphological operations are used to detect lung cancer nodule. And then, extracts low-level features from the detected nodule. This paper uses seven features area, perimeter, eccentricity and four texture features using Gray-level Co-occurrence Matrix (GLCM). Finally, the extracted features from the detected regions are given as input to 3-layer Artificial Neural Network (ANN) classifier to classify the detected lung cancer nodule into stages. Diagnosis is mostly based on CT (computed tomography) images. The lung cancer CT scan images for each stage obtain from the internet.

Keywords: Preprocessing, Morphological Operations, Gray Level Co-occurrence Matrix, Artificial Neural Network.

I. INTRODUCTION

Lung cancer is the leading cause of cancer deaths in the world. It is a large family of diseases that involve abnormal cell growth with the potential to spread to other parts of body. Anatomy of lung is shown in Fig.1. Lung cancer is a disease of abnormal cells multiplying and growing into a nodule. Fig.2 describes the beginning of the cancer. The types of lung cancer are divided into four stages. In stage I, the cancer is confined to the lung. In stages II and III, the cancer is confined to the chest (with larger and more invasive tumors classified as stage III). Stage IV cancer has spread from the chest to other parts of the body. Of all types of cancer, lung cancer is the most common cause of deaths, accounting for 1.3 million deaths annually. An estimated 159,260 people are expected to become from lung cancer in 2014, accounting for approximately 27 percent of all cancer. Early detection of lung cancer can increase the chance of survival among people

There are many techniques to diagnose the lung cancer, such as Chest Radiograph (x-ray), Computed Tomography (CT), Magnetic Resonance Imaging (MRI scan) and Sputum Cytology. However, most of these techniques are expensive and time consuming. A CT scan uses x-rays to make detailed cross-sectional images of your body. Instead of taking one picture, like a regular x-ray, a CT scanner takes many pictures as it rotates around you while you lie on a table. A computer then combines these pictures into images of slices of the part of your body being studied.

CT scans are more likely to show lung tumours than routine chest x-rays. They can also show the size, shape, and position of any lung tumours and can help find enlarged lymph nodes that might contain cancer that has spread from the lung. When a low-dose CT scan of the chest is done for lung cancer screening, it’s common to find small, abnormal areas (called nodules or masses) in the lungs, especially in current or former smokers. Most lung nodules seen on CT scans are not cancer. They are more often the result of old infections, scar tissue, or other causes.

There are two types of lung cancer non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). Non-small cell lung cancer is most commonly appearing than SCLC and it generally increases and spreads more slowly. Smoking is also related with SCLC and increases and spreads more quickly and form large tumors that can spread through the body. Generally, they start many times in the bronchi near the center of the chest. Lung cancer death rate depends on total amount of cigarette smoking. The lungs are a pair of sponge-like, cone-shaped organs [1]. The right lung has three lobes, and is larger than the left lung, which has two lobes. Anatomy of lung is shown in Fig.1. Lung cancer is a disease of abnormal cells multiplying and growing into a nodule Therefore, there is a great need for a new technology to diagnose the lung cancer in its early stages. Image processing techniques provide a good quality tool for improving the manual analysis. The use of image processing techniques can assist radiologists and doctors in diagnosing diseases and to offer a rapid access to medical information gained importance in a short time. In this paper, MATLAB has been used through every procedure made.

In technical literature done by A. Amutha and R.S.D Wahidabanu [3], Level Set-Active Contour Modeling was used as a method in diagnosing lung tumor. First step was removing noise from image using kernel based non-local neighborhood denoising function and done feature extraction based on histogram to classify between normal and abnormal classes. At the final step or in tumor detection, level set-active contour modeling with minimized gradient to the image was introduced. In another study [4], Auto enhancement, Gabor filter and Fast Fourier transform (FFT) were used to enhance the image and used Thresholding and Watershed segmentation to segment the image. While for feature extraction, Binarization and Masking approach were applied. N.A. Memon et. al [5] proposed thresholding method which select the threshold based on the object and background pixel means. Region growing is used then to extract the exact cavity region with accuracy.

The rest of the paper is organized as follows: we describe related works in Section 2, present our implementation in Section 3, provide experimental results in Section 4, and conclude in Section 5.

Fig. 1. Anatomy of lung [2]

II. Related Works

In the past, several methods have been proposed to detect and classify lung cancer in CT images using different algorithms. For example, Camarlinghi et al. [6] have used three different computer aided detection techniques for identifying pulmonary nodules in CT scans. Abdulla and Shaharum [7] used feed forward neural networks to classify lung nodules in X-Ray images albeit with only a small number of features such as area, perimeter and shape. Kuruvilla et al. [8] have used six distinct parameters including skewness and fifth & sixth central moments extracted from segmented single slices containing 2 lungs along with the features mentioned in [7] and have trained a feed forward back propagation neural network with them to evaluate accuracy for different features separately. In Bellotti et al. [9] , the authors have proposed a new computer-aided detection system for nodule detection using active contour based model in CT images. The paper reports a high detection rate of 88.5% with an average of 6.6 false positives (FPs) per CT scan on 15 CT scans. In the recent past a comparison between six different methods for detecting nodules in lungs was done by Ginneken et al. [10] that also proposed a method to combine the output of multiple systems for effectively detection of pulmonary nodules.

In Riccardi et al. [11] the authors presented a new algorithm to automatically detect nodules with an overall accuracy of 71% using 3D radial transforms. In the recent years, there has also been a renewed interest in the field of deep learning and the latest research in area of medical imaging using deep learning shows promising results. One such study is of Suk et al. [12] in which the authors propose a novel latent and shared feature representation of neuroimaging data of brain using Deep Boltzmann Machine (DBM) for AD/MDC diagnosis. The methods outperforms the competing methods and achieve a maximal diagnostic accuracy of 95.52% (AD vs. NC); Wu et al. [13] use deep feature learning for deformable registration of brain MR images demonstrating that a general approach can be built to improve image registration by using deep features. A stacked auto encoder (a type of deep learning architecture) was used by Fakoor et al. [14] to diagnose and classify different types of cancer based on gene expression data, which eventually out performs contemporary methods for different datasets.

III. Methodology

A. Lung Cancer Nodules Segmentation:

In this paper, the first step takes a gathering of lung cancer CT scan images (stage1, stage2, stage3, stage4) by downloading from the internet. The second step applies median filter for image pre-processing to get best level of quality and clearness. Finally, image segmentation is performed by using morphological operations. This work involves the methods in the following sequence: pre-processing, segmentation, feature extraction and classification. The block diagram of the work is as shown in Fig 2.

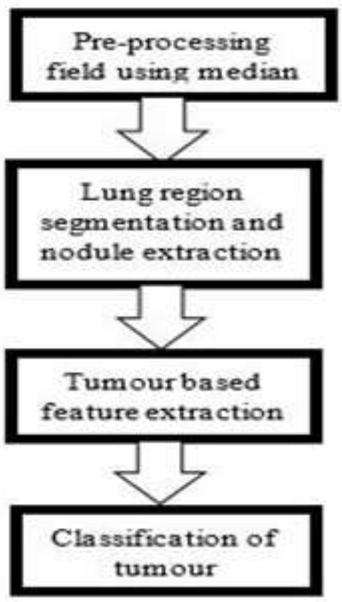


Fig. 2. Block Diagram of the Method

The CT scan image is chosen because CT images are more sensitive in finding the tumour size and the lymph nodes of lung. In the figure 3, shows the beginning of the lung cancer.

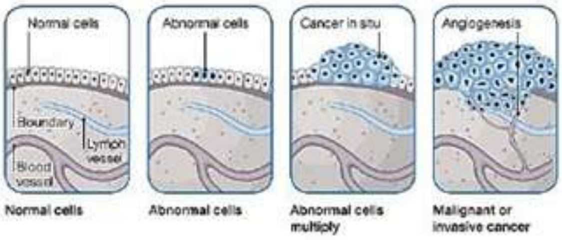
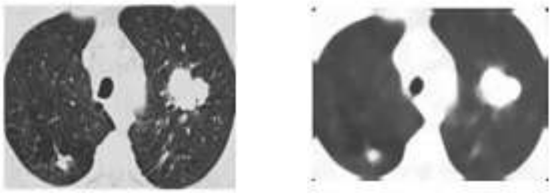


Fig. 3. The Beginning of Cancer

Many medical images have noise present in the images such as white noise, salt and pepper noise etc. In order to get an accurate or clearer result we apply the median filter in the image pre-processing stage to filter out noise that has corrupted image. Median filtering is a nonlinear operation often used in image processing to reduce "salt and pepper" noise. It often does a better job than the mean filter of preserving useful detail in the image. The median is calculated by first sorting all the pixel values from the surrounding neighbourhood into numerical order and then replacing the pixel being considered with the middle pixel value. If the neighbourhood under consideration contains an even number of pixels, the average of the two middle pixel values is used. In general, 3x3 mask size of filter is mostly used. In this paper, the mask size of filter is 10x10 because the larger the mask size, the more eliminate the noise. The output of median filtered image is shown in Figure 4. Median filter is used to remove the noise of images. This pre-processing image is used as the input for image segmentation.



(a) (b)

Fig. 4. (a). Original CT Lung Image Output o Median Filter

In the third step, segmentation of images is performed by using morphological operations to obtain individual lung and to remove unnecessary parts. Morphology is a technique of image processing based on shapes. In this paper, arbitrary shape mask structuring element is used for morphological operations. The basic morphological operations are dilation and erosion. The value of the output pixel is the maximum value of all the pixels in the input pixel's neighborhood. In a binary image, if any of the pixels is set to the value 1, the output pixel is set to 1. Dilation adds pixels to the boundaries of objects in an image. It grows or thickens objects in binary image. The formal definition of dilation of a set *A* by another set *B* is denoted A⊕B, and defined by:

$$A \oplus B = \{z \mid (\hat{B})_z \cap A \neq \phi\}$$

Where, $\hat{B}$ is the reflection of B. This definition means that dilation of A by B is done by reflecting B and then shifting B over A by z. Then all the displacements of B are set such that B and A overlap by at least one element, which gives the dilation. Set B is also referred to as the dilation mask or structuring element (STREL). Erosion is an operation that 'shrinks' or 'thins' objects in a binary image. Erosion produces an opposite effect of dilation. In other words, erosion of A by B is set of all points traversed by center of B such that B is totally contained within A at all times.

$$A \Theta B = \{\mathrm{z} \mid (\mathrm{B})_{\mathrm{z}} \cap \mathrm{A}^{\mathrm{C}} \neq \phi\}$$

After applying the morphological operations, on the output of pre-processing image is shown in Figure 5 and 6.

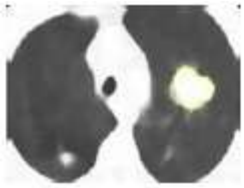


Fig. 5. Output of Morphological Operations for Abnormal Lung

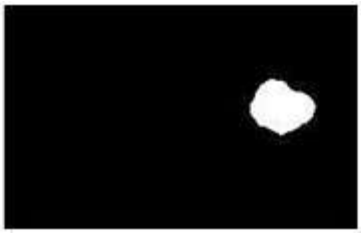


Fig. 6. Output of Segmented Lung Nodule

B. Feature Extraction:

After the segmentation is performed, the segmented lung nodule is used as an input for feature extraction step. Lung cancer nodule has many numbers of features. Basically there are two types of ways to extract features i.e. textural and Structural [15]. The features like geometric and intensity-based statistical features are extracted. This measurement information is very helpful in detecting lung nodule as cancer or not. Shape measurements are physical dimensional measures that characterize the appearance of an object. For feature extraction basic characters are required which are measured in scalar [16]. The physical dimensional measures are defined as follows:

1) Area: The area is obtained by the summation of areas of pixel in the image that is registered as 1 in the binary image within the curly brackets.

2) Perimeter: The perimeter [length] is the number of pixels in the boundary of the object. Perimeter P is measured as the sum of the distances between every consecutive boundary points [17]. Mathematically,

International Journal of Electrical Electronics & Computer Science Engineering
Volume 6, Issue 4 (August, 2019) | E-ISSN : 2348-2273 | P-ISSN : 2454-1222
Available Online at www.ijeecse.com

$$P = |S_n S_1| + \sum_{i=1}^{n-1} |S_i S_{i+1}|$$

where, s= {s1,…,s n} is a set of the boundary points.

3) Eccentricity: The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 and1.After the segmentation is performed, the segmented nodule is used for feature extraction. A feature is significant information extracted of the image. Gray Level Co-Occurrence Matrix (GLCM) has proved to be a popular statistical method of extracting textural feature from images. Haralick also offered different measures i.e. entropy, energy, contrast, correlation. Statistical parameters calculated from GLCM values are as follows:

4) Contrast: Measures the local variations in the GLCM. It calculates intensity contrast between a pixel and its neighbor pixel for the whole image. Contrast is 0 for a constant image.

$$\text{Contrast} = \sum\sum (i-j)^2 \, p\,(i,j)$$

where, p (i, j) = pixel at location(i, j)

5) Correlation: Measures the joint probability occurrence of the specified pixel pairs.

$$Correlation = \frac{\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i \sigma_j}$$

6) Energy: Provides the sum of squared elements in the GLCM.It is also known as uniformity or the angular second moment.

$$\text{Energy} = \sum\sum (p\,(i,j))^2$$

7) Homogeneity: Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

$$Homogeneity = \sum_{i,j} \frac{P(i,j)}{1 + |i-j|}$$

C. Artificial Neural Network:

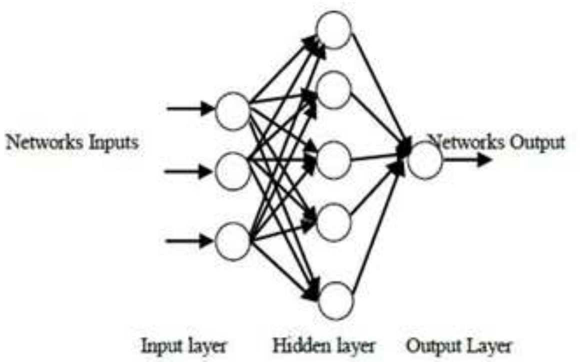


Fig. 7. Architecture of a General ANN

Artificial neural network is one of the classification methods commonly used in image processing techniques. ANN is collections of mathematical models that emulate the real neural structure of the brain [18]. ANN has three layers. They are input layer, hidden layer and output layer. Architecture of a general ANN is shown in Fig 7.

The common terminologies used in ANN include weight, bias and activation functions.

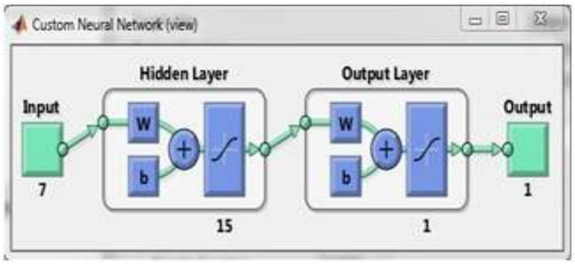


Fig. 8. Artificial Neural Network

In this paper, 3-layers feed-forward neural network toolbox is used for lung cancer nodule stages classification. The feed-forward neural networks are the simplest type of artificial neural networks devised. In this network, the features information moves in only one direction, forward from the input nodes, through the hidden nodes (if any) and to the output nodes. The input layers consider seven features from the feature extraction step. The output layers contain four stages. The hidden layers present 15 layers. For training step, four images on CT Scan Images are used for stage1, stage2, stage3 and stage4. The extracted features for four train images are shown in Table.1. The Artificial Neural Network training process is shown in Fig 9.

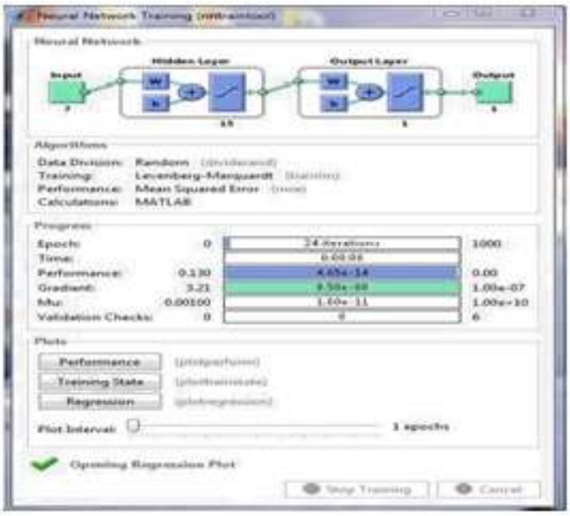


Fig. 9. Neural Network Training

IV. RESULTS AND DISCUSSION

In this paper, the initial step takes a gathering of lung cancer CT scan images (stage1, stage2, stage3, stage4) by downloading from the internet. In the pre-processing step, the median filter is used to remove noise from the input images. According to the outputs of image pre-processing as shown in the Figure 4, median filter is more suitable because the main advantage of median filtering is that even after pixel intensity values are changed the edges of the images are preserved. The increasing mask size is more effective in minimizing the impact of noise. The mask size of this figure result is 10x10. In the segmentation step, morphological operations are used to get individual lung cancer nodule. By doing morphological operations, it gets not only the individual lung but also apparent the lung nodule is shown in the Figure 6. In the feature extraction step, the seven low level features are extracted from the detected lung cancer nodules.

Table I. Features Extracted from the Detected Lung Cancer Nodules for Training

Features	**Stage 1**	**Stage 2**	**Stage 3**	**Stage 4**
Area	12457	320	76	34
Perimeter	590.914	75.523	30.062	23.62
Eccentricity	0.327220 341	0.56802 5683	0.58380 9475	0.870963 098
Contract	6.05E-04	0.00037 4948	7.42E-05	0.001448 646
Correlation	0.983957 564	0.91543 7109	0.86838 3943	0.705156 242
Energy	0.961697 379	0.99519 1239	0.99936 2064	0.993640 188
Homogeneity	0.999697 614	0.99981 2526	0.99996 2901	0.999275 677

The above table shows the result of seven extracted low-level features for each of the stage1, stage2, stage3 and stage4 on four CT scan images. The seven low level features area, perimeter, eccentricity, contract, correlation, energy and homogeneity are calculated from the above formulas in session 3.2. The extracted features are given as input to the ANN classifier for training in classification to classify lung cancer nodule stages. This system uses four CT scan images for training and another four CT scan images for testing purposes. Finally, the performance of the system is evaluated. This system offers the accuracy of 92% with epoch 24 and minimum gradient is reached.

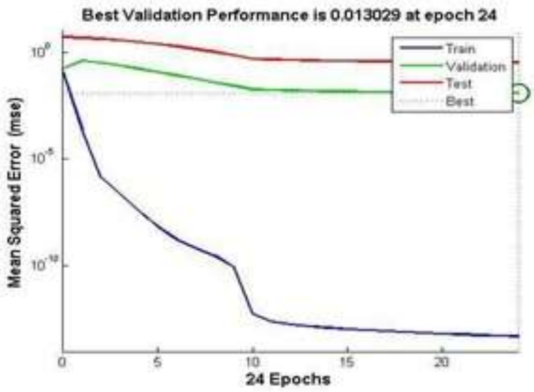


Fig. 10. Neural Network Training Performance Plot

V. CONCLUSIONS

Lung cancer detection technique is proposed here using different image processing techniques. It primarily uses CT images. Images are pre-processed first. The second step applies median filter for image pre-processing to get best level of quality and clearness. The third step is image segmentation using morphological operations to segment lung cancer nodule on CT scan images. This system uses 7 low level features extracted from the detected lung nodule is given as input to 3-layer ANN classifier to classify cancer stages. Determining the nodule features provide to know more information of the condition of lung cancer at the early stages. This system offers the accuracy of 92% with epoch 24 and minimum gradient is reached. This technique helps the radiologists and the doctors by providing more information and taking correct decision for lung cancer patient in short time with accuracy. Therefore, this method is not expensive and few time consuming.

VI. REFERENCES

[1] Non-Small Cell Lung Cancer, [online available], http://www.katemacintyrefoundation.org, Adapted from National Cancer Institute (NCI) and Patients Living with Cancer (PLWC), 2007, accessed on 13 July 2013.

[2] Anatomy of lung picture and beginning of cancer, [online available], www.allreferhealth.com, accessed on 4 June 2013.

[3] A. Amutha and R.S.D Wahidabanu, "A Novel Method for Lung Tumor Diagnosis and Segmentation using Level Set- Active Contour delling", European Journal of Scientific Research, Vol.90, No.2, November 2012, pp.175-187.

[4] Mokhled S. Al-Tarawneh, "Lung Cancer Detection Using Image Processing Techniques", Leonardo Electronic Journal of Practices and Technologies, Issue 20, January- June 2012, p.147-158. ISSN 1583-1078.

[5] N.A. Memon et. al, “Segmentation of Lungs from CT Scan Images for Early Diagnosis of Lung Cancer,” World Academy of Science, Engineering and Technology. 2006.

[6] S. Beucher and F. Meyer, “The morphological approach to segmentation: The watershed transform” in Mathematical Morphology in Image Processing, E. R. Dougherty, Ed. New York: Marcel Dekker, 1993, vol. 12, pp. 433–481.

[7] N.A. Memon et. al, “Segmentation of Lungs from CT Scan Images for Early Diagnosis of Lung Cancer,” World Academy of Science, Engineering and Technology. 2006.

[8] N. Camarlinghi, I. Gori, A. Retico, R. Bellotti, P. Bosco, P. Cerello, G. Gargano, E. L. Torres, R. Megna, M. Peccarisi et al., “Combination of computer-aided detection algorithms for automatic lung nodule identification,” International journal of computer assisted radiology and surgery, vol. 7, no. 3, pp. 455–464, 2012.

[9] A. A. Abdullah and S. M. Shaharum, “Lung cancer cell classification method using artificial neural network,” Information Engineering Letters, vol. 2, no. 1, pp. 49–59, 2012.

[10] J. Kuruvilla and K. Gunavathi, “Lung cancer classification using neural networks for ct images,” Computer methods and programs in biomedicine, vol. 113, no. 1, pp. 202–209, 2014.

[11] R. Bellotti, F. De Carlo, G. Gargano, S. Tangaro, D. Cascio, E. Catanzariti, P. Cerello, S. C. Cheran, P. Delogu, I. De Mitri et al., “A cad system for nodule detection in low-dose lung cts based on region growing and a new active contour model,” Medical Physics, vol. 34, no. 12, pp. 4901–4910, 2007.

[12] B. van Ginneken, S. G. Armato, B. de Hoop, S. van Amelsvoort-van de Vorst, T. Duindam, M. Niemeijer, K. Murphy, A. Schilham, A. Retico, M. E. Fantacci et al., “Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the anode09 study,” Medical image analysis, vol. 14, no. 6, pp. 707–722, 2010.

[13] A.Riccardi, T. S. Petkov, G. Ferri, M. Masotti, and R. Campanini, “Computer-aided detection of lung nodules via 3d fast radial transform, scale space representation, and zernike mip classification,” Medical physics, vol. 38, no. 4, pp. 1962–1971, 2011.

[14] H.-I. Suk, S.-W. Lee, D. Shen, A. D. N. Initiative et al., “Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis,” Neuro Image, vol. 101, pp. 569–582, 2014.

[15] Bhawana Malik,Jaykant Pratap Singh,Veer Bhadra Pratap Singh and Prashant Naresh, “Lung Cancer Detection at Initial Stage by Using Image Processing and Classification Techniques”, International Research Journal of Engineering and Technology (IRJET) Volume 3, ISSN:e-2395-0056,p-2395-0072,November 2016.

[16] Bhagyashri G. Patil and Prof.Sanjeev N.Jain, “Cancer Cells Using Digital Image Processing Methods”, International Journal of Latest Trends in Engineering and Technology (IJLTET) Volume 3, ISSN: 2278-621X, March 2014.

[17] Cigdem Demir and B ¨ulent Yener., "Automated cancer diagnosis based on histopathological images: a systematic survey", Technical, Rensselaer Polytechnic Institute, Department of Computer Science, TR-05-09.

[18] Mark Hudson Beale, Martin T. Hagan, Howard B. Demuth, "Neural Network Toolbox™ User‟s Guide", R2014a Matlab, MathWorks, Inc.