Prediction of Cancerous Cell by Cluster Based Biomedical CT Image and Analysis

Sunita Agrawal [1], Bharati Raidu[1], Kusum Agrawal[1], R. C. Barik[2*]

[1,2]*Department of Computer Science & Engineering, Vikash Institute of Technology, Bargarh, Odisha , India*

Abstract: *Medical and health arena is advanced in recent years with the technological influence especially using image processing techniques and algorithms. Biomedical Image processing resolves many cons of manual disease recognition. In this paper we have depicted the automated clinical diagnosis for tumor detection based on segmentation of CT scan images towards lungs cancer, ovary cancer and liver cancer. Tumor is an exceptional expansion generated by human cells reproducing themselves in an unconstrained manner. Accurate detection of size and location of tumour plays a vital role in the diagnosis of tumor. Clustering plays a specific role in Image object segmentation both in Gray and RGB based Bio-medical images. We have taken different CT scanned Image of three main sources of diseases processing for detection of tumor based on three major life killer disease with the steps of image processing keeping prominence on noise removal, contrast enhancement by stretching.*

Keywords: *CT Image, Diagnosis, K-means Clustering, Probability Density Function (PDF), Segmentation*

I. Introduction

Image segmentation is one of the widely used methods to organize the pixels of an image accurately in a decision oriented application. It divides an image into a number of discrete fields such that the pixels have high resemblance in each region and high distinction between regions. It is an important tool applied to different problem domain including health care, image processing, traffic image, pattern recognition etc. There are different approaches for image segmentation like threshold based, edge based, cluster based, neural network based. From the different methods one of the most adequate methods is the clustering method .There are distinct types of clustering: K-means clustering, C–means clustering, mountain clustering method and subtractive clustering method. One of most popular clustering algorithm is k-means clustering. It is simple and computationally faster than the other clustering even though the centres are prior decided. And it can also work for large number of variable. It is required to initialize the proper number of number of cluster, k^2. Again, it is required to load the k number of centroid. Distinct value of initial centroid would result distinct cluster. So, alternatives of proper initial centroid are also a significant task. Now a day's image segmentation becomes one of important tool in medical area where it is used to extract the region of interest from the background of biomedical images. So medical images are segmented using different techniques and process outputs are used for the further analysis in medical. Mathematically an image is a 2-D spatial function

$$\mathbf{f(x, y) = r(x, y) \times i(x, y)} \tag{1}$$

Equation 1 represents the finite discrete values in the form of amplitude of f at any pair of spatial or plane coordinates. (x, y) is called the intensity or grey level of the image at that point [1] where i(x,y) is the illumination property and r(x, y) is the reflectance property.

As per the standard size of a grey Image 512×512and 256 × 256 have 262144 and 65536 intensity features respectively. Image clustering and segment clustering or labelling is a major issue in image processing.

II. Literature Survey

2.1 Image Acquisition

Images are procured using CT scan and these scanned images are displayed in a two dimensional matrices having pixels as its elements. These matrices are conditional on matrix size and its field of view. Images are stored in Image file and shown as a gray scale image. The approaches of a gray scale image are ranging from 0 to 255, where 0 displays total black colour and 255 displays pure white colour. Entries between these ranges vary in intensity from black to white. Even though for acquisition of an image highly sophisticated scanning machine used but still for prediction only one scanned image is not sufficient to diagnose life killer diseases. But scanning the same image may not be same due to camera aperture setting, focal length. So enhancement is the ultimate solution before knowledge based prediction or detection [5, 6].

2.2 Pre-processing Stage for Image Enhancement

In this stage image is upgraded in the way that finer details are refined and noise is detached from the image. Most commonly used enhancement and noise degradation techniques are accomplished that can give finest possible results. Enhancement will result in more prominent edges and a sharpened image is obtained, noise will be shortened thus dropping the blurring effect from the image. In addition to enhancement, image segmentation will likewise be adapted. This refined and enhanced image will help in detecting edges and refining the superiority of the overall image. Edge detection will lead to the discovery of the exact location of tumour. In general the image smoothing can be mathematically denoted in equation 2.

$$g(x, y) = \frac{\sum_{i=-m}^{m} \sum_{j=-n}^{n} w_{i,j} \cdot f(x + i, y + j)}{\sum_{i=-m}^{m} \sum_{j=-n}^{n} w_{i,j}} \tag{2}$$

Where $g(x, y)$ is enhanced gray level pixel intensity image, $w_{i,j}$ is weighted mask and $f(x + i, y + j)$ gray level pixel Intensity function in both spatial x and y coordinate. The value of m and n depends on the mask size of -1 to 1 for 3×3 mask, -2 to 2 for 5×5 mask and so on.

2.2.1 Image Smoothing

Smoothing of the image can be done by using any of the low pass linear and non-linear filters such as box filter, weighted average filter, Median filter etc. These filters can remove the noise by smoothing the image texture.
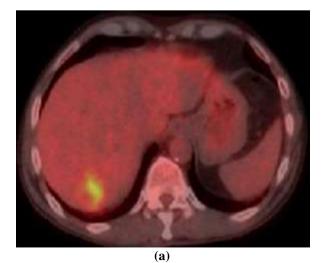
2.2.2 Image Sharpening

Sharpening of the image can be done by using different high pass filters. Noise is being eliminated by using different low pass filters, the image need to be sharpened because the image need sharp edges which will help to identify the boundary of the tumour. Gaussian high pass filter gives very high rated results and used very extensively to enhance the finer specifics of the project.

2.2.3 Histogram Processing

Histogram processing is a important phase of image enhancement to magnify the contrast using Histogram Equalization by calculating $P_s(s)$ probability density function (PDF) of s which is output or processed gray level and $P_r(r)$ probability density function of input gray level r. To equally distribute the gray level Cumulative density function (CDF) need to be calculated [5, 6]. Another most important Histogram processing technique is Histogram specification or Histogram Matching to match a target histogram by finding cumulative probability density function of z as $P_z(z)$ using the equilibrium as $\cong s \cong z$.Mathematically Histogram Equalization can be denoted in equation 3.

$$s = T(r) = (l - 1) \sum_{i=0}^{k} P_r(r_i) \tag{3}$$

Figure 1 shows the PDF of the grey level of RGB plane of the input image. Figure 2 show the same image in grey plane or black and white plane.
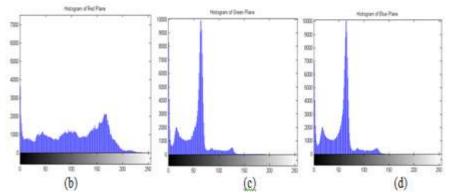

(a)

Figure 1 (a) Original 2-D Biomedical Liver Image with infectious cancerous disease. (b) Histogram of the Red plane (c) Histogram of the Green Plane (d) Histogram of the Blue plane. From figure 1(b) it is clearly shows the supremacy of red plane and its Probability Density Function (PDF) is equally distributed to all the grey levels. In general the RGB image has the intensity level distribution as 66:32:2 i.e. 66% for red.
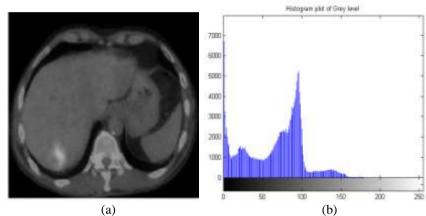

Figure 2 (a) Original Grey Image (b) Histogram plotting shows the Probability Density Function.

Figure 2(b) shows the PDF distribution which is not equally distributed to all the grey level. So the image should be passed to Histogram Equalization and Specification phase.

2.3 Noise Reduction
Noise is an external source of unwanted signal diluted or making contamination to the original signal. CT scan image is being noisy from different sources which can be reduced using image processing low pass and high pass filter in both spatial and frequency domain. Removing noises using computational approach is still manual but So many automated algorithms were being proposed in recent past [3, 4, 5]. The noises such as Gaussian noise, Rayleigh noise, Salt and Pepper noise etc. are eradicated using many spatial domain and frequency domain filter. Figure 3 demonstrates the Gaussian noise over the original image.
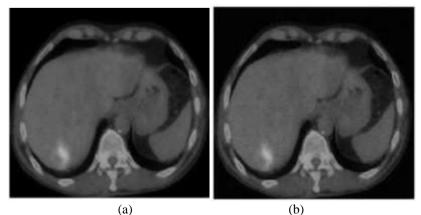

Figure3 (a) Original grey level image (b) Gaussian noise over original image

2.4 Segmentation using K-Means Clustering

Clustering is a process of allotment or ordering a given sector unlabelled pattern into a number of clusters such that identical patterns are authorized to a group, which is examined as a cluster. Figure 1 clearly depicts the cluster of four data groups. Segmentation is an elementary process to extract information from complex medical images. The crucial objective of the image segmentation is to isolate an image into commonly modified and exhausted regions such that each region of importance is spatially continuous and the pixels within the region are homogeneous with esteem to a predefined standard.

K means is one of the famous clustering methods because it is simpler and easier in computation. It is the simplest unsupervised learning algorithms that solve the well known clustering problems. The algorithm calculates the latent features from the input data or signals as a vector space and performs the natural clustering [2]. Mathematically K-means represented in equation 4. The points are categorized around centroids or cluster centres $\mu_i \forall_i = 1, 2, \dots k$ that are obtained by minimizing the objective or distance

$$V = \sum_{i=1}^{k} \sum_{x_j \in s_i} (x_j - \mu_i)^2 \qquad (4)$$

Where there are k clusters $s_i, i = 1, 2, \dots, k$ and μ_i is the centroid or mean point of all the points $x_j \in s_i$.

In figure 4(a) it is clearly depicted that the similar characteristic data set are grouped under one group. The groups are interlinked using some common characteristics. The figure 4(b) depicts a generalized k-means clustering for two random data set keeping emphasis on the cluster centres and its surrounding data points.
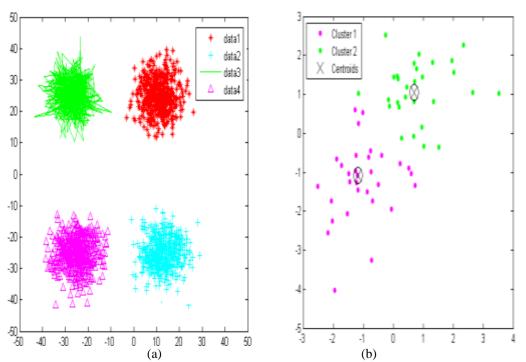

Figure 4: (a) Shows Group or cluster formation among four different dataset. (b) K-means clustering applied to two random data pattern.

III. Materials & Method

Acquiring the same CT scan image for better diagnosis of disease is depends upon the lens and aperture settings. So for an optimize prediction of the diagnosis result better contrast, sharpened image is required which will make ease of segmentation process. Segmentation of the CT scanned images for detection of tumors using K-Means clustering technique. A cluster can be described as a collection of pixels where all the pixels in assured group defined by identical relationship. Clustering is unsupervised learning because the algorithm naturally classifies objects based on user given condition. Here K-Means clustering algorithm for segmentation of the image is used for tumor detection from the CT scanned images. The proposed block diagram is as shown in figure 5.
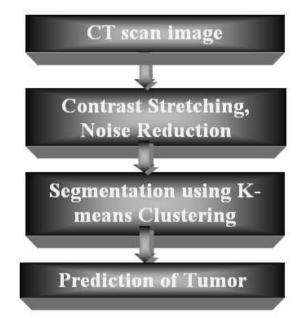
Figure 5: Proposed method block diagram

CT Scanned images of the human organs forms the input images for our system where the gray-scale CT Scanned input images are given as the input. The fetching stage will convert the RGB input image to gray-scale. Noise present if any, will be detached using a median filter. The fetched image is given for image segmentation using K-Means clustering algorithm.

The better choice is to position them as much as possible far away from each other. The next step is to take each point belonging to a given data set and collaborate it to the nearest centroid. When no point is continuing, the first step is completed and an early grouping is done. At this point we need to re-calculate k new centroids as barycentre of the clusters emerging from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the adjacent new centroid. A loop has been accomplished. As a result of this loop we may consider that the k centroids change their location step by step until no more changes are done. In other words centroids do not move anymore [8]. Finally, this algorithm marks at minimizing an objective function, in this case a squared error function. The objective function is represented in equation 5.

$$J = \sum_{j=1}^{k} \sum_{i=1}^{x} \left\| x_i^{(j)} - c_j \right\| \qquad (5)$$

Where $\left\| x_i^{(j)} - c_j \right\|^2$ is a selected distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, is an symbol of the distance of the n data points from their corresponding cluster centres [8].

Let us examine an image with resolution of x×y and the image has to be gathered into k number of cluster. Let p(x, y) be an input pixels to be cluster and c_k be the cluster centres. The algorithm for k-means clustering is resulting as:

1. Compute number of cluster k and centre.

2. For each pixel of an image, calculate the Euclidean distance d, between the centre and each pixel of an image applying the relation given in equation 6.
$$d = \left\| p(x, y) - c_k \right\| \qquad (6)$$

3. Assign all the pixels to the closest centre based on distance d.

4. After all pixels have been selected, recalculate new position of the centre using the relation given in equation 7.

$$ck = k^{-1} \sum_{y} \in c_k \sum x \in c_k \, p(x, y) \qquad (7)$$

5. Repeat the process until it captures the resistance or error value.

6. Reform the cluster pixels into image.

IV. Results & Analysis

With some gray-scale image, which are affected with tumor and which are altered it into RGB image, by this it indicates which area is affected by tumor, as shown in figure the yellow portion is the affected area. Manual diagnosis of tumor in Liver, Lungs, and Ovary is replaced by automated highly sophisticated Image processing software where different algorithms are used. Segmentation procedures partition an image into its constituent parts or objects. A rugged segmentation procedure brings the process a long way toward successful solution of imaging problems that require objects to be identified individually. Faulty segmentation algorithms almost always guarantee eventual failure [5]. Segmentation are of different types such as Region based Segmentation , Color Based Segmentation such as Segmentation in HSI space , Segmentation in RGB vector space, Edge detection based segmentation etc.

4.1.1 Case Study 1- Liver Tumor Detection

A CT scan of the abdomen can identify many types of liver tumors. It can yield absolute knowledge about the size, shape and location of any tumors in the liver. Figure 6 demonstrates a tumor is detected or predicted in the abdomen by clustering based segmentation after necessary image enhancement procedure. The clustering is applied to colour based segmentation.
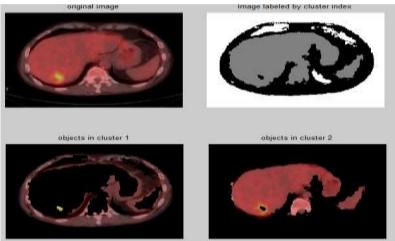

Figure 6: Original CT image of Liver infectious Tumor detection by automated segmentation using K-means clustering [9].

4.1.2 Case Study 2- Lungs Tumor

Lung cancer is one of the most threatening diseases in the world.CT scan images are cured from different hospitals. These images include less noise as compared to X-ray and MRI images. The CT captured images are processed. The region of interest i.e., tumor is identified accurately from the original image. Cluster 2 in Figure 7 shows clearly the tumor.
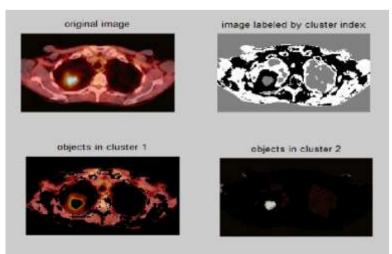

Figure 7: Original CT image of Lung infectious Tumor detection by automated segmentation using K-means Clustering [10].

4.1.3 Case Study 3- Ovary Tumor

Formation of Tumor in Ovary is a life killer disease. Basically the ovary tumors grow up to its maximum. CT scan image gives a realistic view of tumor in the image but automated cluster based segmentation will leads to prediction of tumor with its size, shape and location. Figure 8 shows that cluster based segmentation to predict the tumor.
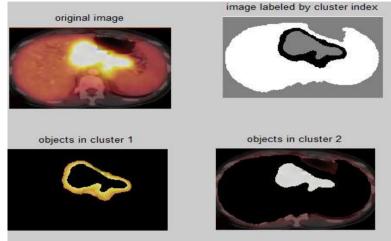

Figure 8: Original CT image of Ovary infectious Tumor detection by automated segmentation using K-means Clustering [11].

V. Conclusion

Cancer is a threat for human society. Prediction of cancer in the form of a tumor is a major area for researcher in last decade. It is achieved by many Image Processing computation techniques. Detection of tumor is not equivalent with detection of Malaria. This paper gives more focus on three major types of cancer diseases using the popular k-means based clustering approach. In spite of highly sophisticated scanning machine used in image acquisition still for better diagnosis so many images of the same objects are required. Again some images may be noisy, poor contrast which needs to be passed for Image enhancement by Smoothing, Sharpening, Histogram processing, Contrast Stretching. After enhancement the image is passed to automated clustering based segmentation phase. This paper gives a narrative approach of all the major phases of knowledge based image description and representation.

Reference

[1]. R.C.Barik, R. Pati and H. S. Behera, "Robust Signal Processing Compression for Clustering of Speech Waveform and Image Spectrum", IEEE Conference on Communication and Signal Processing.pp 1816 -1820.

[2]. R. C. Barik, R. Mishra, "Comparative Analogy on Classification and Clustering of Genomic Signal by a Novel Factor Analysis and F-Score Method", Advances in Intelligent Systems and Computing, Springer,pp 399-409

[3]. Jang Hwan Cho, Jeffrey A. Fessler "Regularization Designs for Uniform Spatial Resolution and Noise Properties in Statistical Image Reconstruction for 3-D X-ray CT", IEEE Transactions on Medical Imaging Vol: 34, Issue: 2, Feb. 2015 pp: 678-689.

[4]. AnjaBorsdorf , Rainer Raupach , Thomas Flohr , Joachim Hornegger ,"Wavelet Based Noise Reduction in CT-Images Using Correlation Analysis", IEEE Transactions on Medical Imaging (Volume: 27, Issue: 12, Dec. 2008).

[5]. R.C. Gonzalez, R.E. Woods, Digital Image Processing, 3rd Edition, Pearson Education.

[6]. R. C. Gonzalez, Woods and Eddins, Digital Image Processing using Matlab, 2nd Edition, Tata McGraw Hill.

[7]. Minsoo Chun, Young Hun Choi ,Jong Hyo Kim, "Automated measurement of CT noise in patient images with a novel structure coherence feature" , Physics in Medicine & Biology 60 (2015) 9107–9122.

[8]. Malathi R, Nadirabanu Kamal A. R., "Brain Tumor Detection and Identification Using K-Means Clustering Technique" International Journal of Advanced Networking and Applications. pp 14-18.

[9]. http://img.medscapestatic.com/pi/meds/ckb/08/17608tn.jpg

[10]. https://d2t4uon7z5xj4k.cloudfront.net/images_20nov2012/08-09-13-p8-c0154613-lung_cancer,_ct_and_pet_scans-spl.jpg

[11]. http://www.petscaninfo.com/zportal/portals/phys/cancer/ovarian/case01/Ovarian1a.jpg.