A Disease Sample Pixel Clustering for Accurate Infection Prediction Using Image Processing and Data Mining

Mr. Milind H. Padgavankar[#1], Mr. Shrikant P. Akarte[*2]

[#1]ME (CSE) ,Second Year,Department of CSE, Prof. Ram Meghe Institute Of Technology and Research, Badnera, Amravati
Sant Gadgebaba Amravati University, Amravati, Maharashtra, India – 444701.

[*2] Assistant Professor, Department of CSE, Prof. Ram Meghe Institute Of Technology and Research, Badnera, Amravati.
Sant Gadgebaba Amravati University, Amravati, Maharashtra, India – 444701.

Abstract— The liver is one of the most important organs of human body. Liver diseases result heavy death worldwide. In helping doctors and surgeons treating liver diseases, computer aided liver disease diagnosis and liver surgical planning systems are playing important roles nowadays. The proposed detection methodology makes use of MRI, CT and USG scan imagery. K-means clustering technique is adopted so as to segment the images in order to capture the region of interest. Later, Haar wavelet transform is considered to compute the threshold values for the region of interest. It provides accuracy over 85%.

Key words: -Data mining, Haar wavelet transform, Image processing, k-means clustering, Liver cancer.

I. INTRODUCTION

Liver is one of the most important organs of human body. Liver diseases result heavy death worldwide. In helping doctors and surgeons treating liver diseases, computer aided liver disease diagnosis and liver surgical planning systems are playing important roles nowadays. An accurate and automatic segmentation approach of liver parenchyma, vessel and tumors is crucial to a computer aided liver disease diagnosis and liver surgical planning system such as a system for liver transplantation. However, due to the highly varying shape of liver, low contrast and intensity in homogeneity inside liver, weak boundaries to its adjacent organs such as heart and stomach, and intensity homogeneity to adjacent organs, liver segmentation becomes a challenging task that has attracted much research attention recently. The prime facie of this disease is due to the regular or excessive consumption of alcohol, intake of contaminated food and drugs, injecting drugs with shared needles. Apart from these, having low immunity, inherited liver diseases, L-carntine deficiency, smoking, etc.
.

II. LITERATURE REVIEW

Research is a process of identification of problem in a particular field by adopting any of the existing or introducing new techniques. In this paper a method is introduced to diagnose liver cancer which is considered to be novel and pragmatic. Various scientist has put forward their utmost effort in the B. Prem et al. [1] put forth a proposal for Liver surgery based on the Vascular Territories in which data related to portal and hepatic vein were used for experimentation. Smriti Sahu et al. [2] proposed an image enhancement technique for ultra sound images such as contrast stretching, shock filters and so on so as to provide an effective clinical diagnosis approach. Pradeep Kumar B.P et.al [3] proposed a fully automatic segmentation of Ultrasound liver images using Peak and valley method which is a new nonlinear, non iterative multidimensional filter for impulsive like noise reduction. S.S. Kumar, Dr. R. S. Moni et al [4] proposed a computer aided analytical system for the diagnosis of benign and malignant liver tumors from computed tomography (CT) images using curvlet transform based multiresolution texture feature extraction and neural network. Ekong V.E. et al. [5] have developed a fuzzy cluster means system to support the diagnosis of liver diseases using a set of clinical signs and symptoms with LFTs. The experimental results showed a quality enhanced liver diseases diagnosis, but with a time complexity. Robin Martin, et al. [6] implemented a semi-automatic method based on region-growing to isolate the liver part. Smitri et al, [7] uses the image enhancement technique along with filtering technique over the ultrasound images so as to enhance and read the scan image well for proper detection. Himadri Nath Moulick and Moumita Ghosh [8] presented how images can be compressed so as to reduce the memory size and how image can be clustered using k means clustering technique. Piotr Porwik and Agnieszka Lisowska [9] in their paper gave a brief on wavelet transform of which Haar wavelet transformed is elucidated. Pansnur M.A and P.S. Malge [10] represent the technique of image retrieval using Haar wavelet transform and K means clustering technique.

2.1 Data Mining

The most commonly accepted definition of "data mining" is the discovery of "models" for data. A "model," however, can be one of several things. Statisticians were the first to use the term "data mining." Originally, "data mining" or "data

dredging" was a derogatory term referring to attempts to extract information that was not supported by the data. For ex.: Suppose our data is a set of numbers. This data is much simpler than data that would be data-mined, but it will serve as an example. A statistician might decide that the data comes from a Gaussian distribution and use a formula to compute the most likely parameters of this Gaussian. The mean and standard deviation of this Gaussian distribution completely characterize the distribution and would become the model of the data. More recently, computer scientists have looked at data mining as an algorithmic problem. In this case, the model of the data is simply the answer to a complex query about it. There are many different approaches to modeling data

2.2 K-means Clustering

k-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. *k*-means clustering aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells. The problem is computationally difficult (NP-hard); however, there are efficient heuristic algorithms that are commonly employed and converge quickly to a local optimum. These are usually similar to the expectation-maximization algorithm for mixtures of Gaussian distributions via an iterative refinement approach employed by both algorithms. Additionally, they both use cluster centers to model the data; however, *k*-means clustering tends to find clusters of comparable spatial extent, while the expectation-maximization mechanism allows clusters to have different shapes.

2.3 Haar- Wavelet Transform

The **Haar wavelet** is a sequence of rescaled "square-shaped" functions which together form a wavelet family or basis. Wavelet analysis is similar to Fourier analysis in that it allows a target function over an interval to be represented in terms of an orthonormal basis. The Haar sequence is now recognised as the first known wavelet basis and extensively used as a teaching example. The Haar sequence was proposed in 1909 by Alfréd Haar. Haar used these functions to give an example of an orthonormal system for the space of square-integrable functions on the unit interval [0, 1]. The study of wavelets, and even the term "wavelet", did not come until much later. As a special case of the Daubechies wavelet, the Haar wavelet is also known as Db1. The Haar wavelet is also the simplest possible wavelet. The technical disadvantage of the Haar wavelet is that it is not continuous, and therefore not differentiable. This property can, however, be an advantage for the analysis of signals with sudden transitions, such as monitoring of tool failure in machines. The Haar wavelet's mother wavelet function $\psi(t)$ can be described as

Its scaling function $\phi(t)$ can be described as

$$\phi(t) = \begin{cases} 1 & 0 \le t < 1, \\ 0 & \text{otherwise.} \end{cases}$$

III. PROPOSED SYSTEM

Figure shows the block diagram of the system. This includes the objective to process and analyze the images that are framed from the CT scan, MRI scan [12] and generate results whether the said image contains cancer cells or not. These images are collected from various multispecialty hospitals and diagnostic centers. The experimentation procedure makes use of MATLAB R2013 a software in order to process the images. The overall methodology is elucidated by the segmentation process represented. As the images that are framed via the MRI/CT scan exist as a grayscale image. It creates a discrepancy in identifying the cancer cell which may mislead the experimentation process, therefore the image is again converted to RGB image which makes it easier to identify the cancer cell based on the color. Now, as the cancer cell is considered to be the region of interest segmenting [18] the liver alone from the abdominal CT image is difficult due to the fact that the image includes other organs like kidney, spleen, pancreas etc very close to the liver. In order to amass only the liver part and analyze the cancer cell the experimentation makes use of image segmentation using *K*-means clustering [10]. The clusters thus formed by the above mentioned process. Now the clustered image that depicts the extracts of cancer cell which is further used for detection process. So as to analyze and judge if the given image is a cancer cell or not the feature is extracted by cropping the region the interest, and for these images the threshold range is to be fixed.
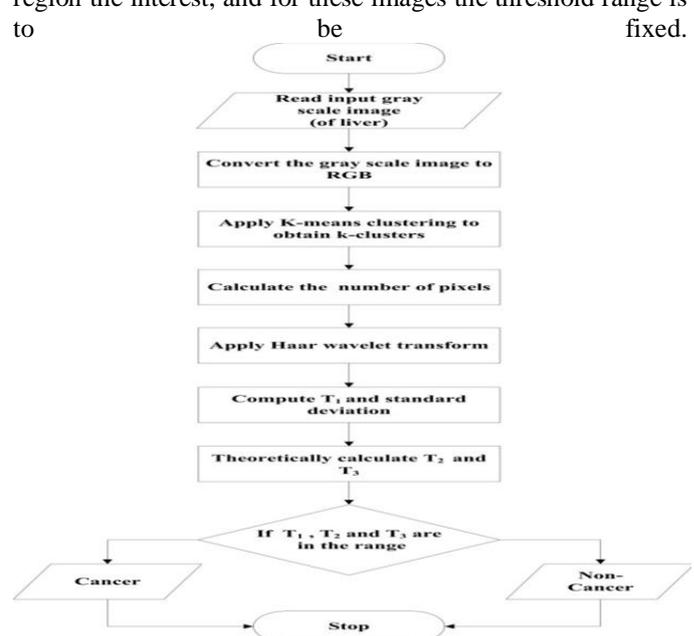

Figure 3.1: Proposed System Architecture

IV. METHOD IMPLEMENTED

To detect cancer cell in the images which are given by the CT scans and MRIs. first we have to select the image where we have to detect the cancer cell which are the input images. In the first part infected input image which is selected we have to perform the first part image processing section which we will gives us the parameter mean intensity and entropy of the image is given. In the second section which is background removal. we will filter the image the black portion will be removed from the images. The Third section which is the K-means clustering where the clusters are created. The standard type of patterns are generated of the input image. Where we can scan the pattern if none of the pattern is useful then we can select the original image and process. In the next section which is the Noise filter removal where more enhance images are given with the noise removal and time taken by software and etc. In the next section which is the compute infection section where we can see which type of infection is present in images which are given. In the next section which is the images registration where we can register the pattern which are infected and where we can sent it to database for the future references. The second part of the process is the Pattern Matching where we can analyze the infected pattern which are stored in our database and we can easily diagnose symptoms which are required for the infection detection of the cancer cells. Here can give the number symptoms which are usefull in the detection of the cancer. The detected patterns and the symptoms which are compared so that we can save the disease names to our database so that we can use it in future.

V. EXPERIMENTAL ANALYSIS

5.1 Requirement Analysis

For the implementation of this system we used MATLAB 2013 and Visual Studio2013. **MATLAB** (**mat**rix **lab**oratory) is a multi-paradigm numerical computing environment and fourth-generation programming language. A proprietary programming language developed by Math Works, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, Java, Fortran and Python. Although MATLAB is intended primarily for numerical computing, an optional toolbox uses the MuPAD symbolic engine, allowing access to symbolic computing abilities. An additional package, Simulink, adds graphical multi-domain simulation and model-based design for dynamic and embedded systems.

5.2 Hardware and Software Requirements

A. Hardware Requirement:
Processor : Dual Core or Onwards
RAM : 4 GB RAM
HDD : 40 GB
LAN : Enabled
B. Software Requirement:
Operating Platform :Windows 7
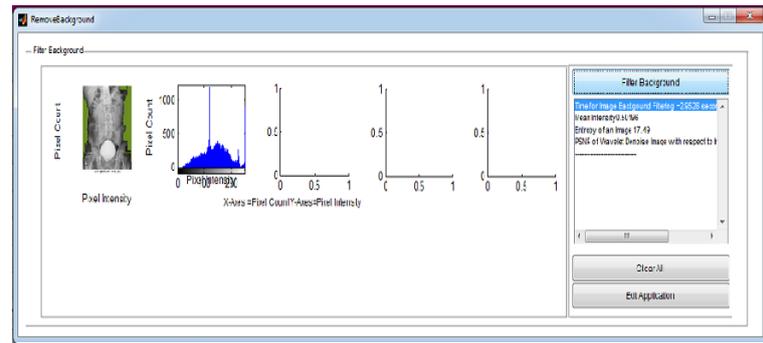Front End : MATLAB

VI. RESULT

6.1 Background Removal

Fig.6.1 Background Removal

This is the background removal process. where the input image which scanned and the the unwanted black pixel are removed and set to the green color.
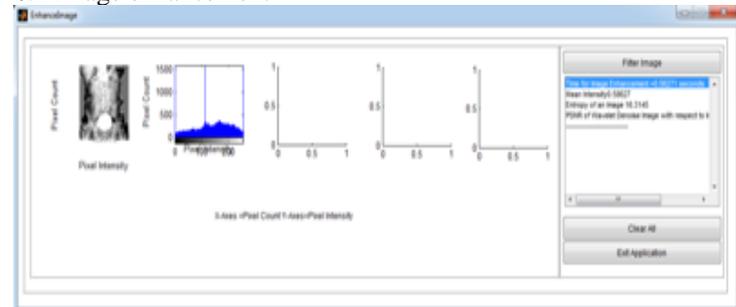
6.2 Image enhancement

Fig.6.2 Image enhancement

This consist of the Image Enhancement Process. Where we are going to calculate the pixel values of the image which are nearer to 255. If the values of the pixels ranges in 245 to 255 region we are making it white.

6.3 Infection detection
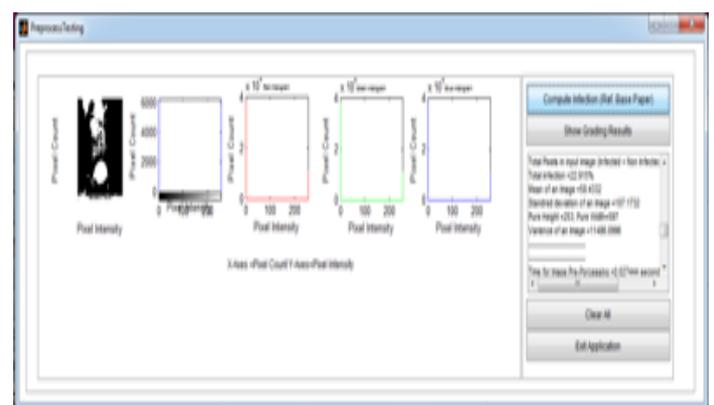

Fig. 6.3 Infection detection

This is Infection Detection here are going to count the Whiter pixel. Here we are going to count then pixel which are

1213

ISSN: 2278 – 1323

International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 5, Issue 4, April 2016

near to 255. We are going to mark it as the infected pixel. The Algorithm scans the image it same form and if it is near to 255 pixels then it is the infected pixel.
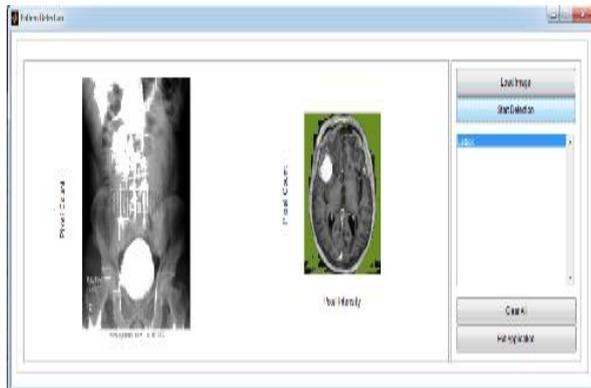
6.4 Pattern Matching

Fig.6.4 Pattern Matching

The pattern matching process where we are going to compare the infected pixels with the registered image. Various infected pattern are scanned with the image which is given by our software. In grading result these infected patterns are showed by the blue box in infected region.
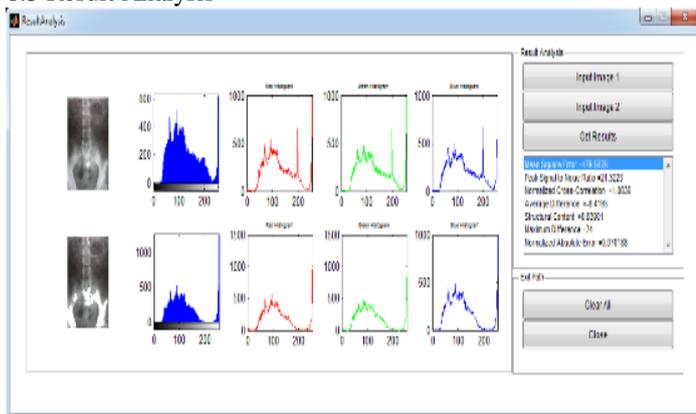
6.5 Result Analysis

Fig6.5 Result Analysis

In this section two images are scanned. First image is the input image and the second image is the registered image. These two images are compared and the value are generated. These values are given in graph below.
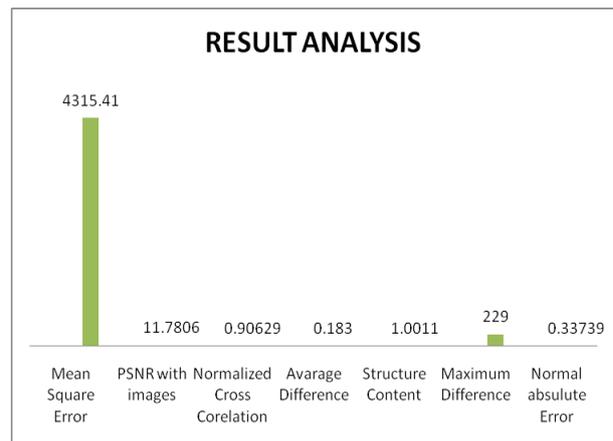

Fig.6.5.1 Result Analysis.

VII. CONCLUSION

This model contributes by providing a computer-aided diagnostic system for the diagnosis of the liver cancer using the images framed through the MRI, CT scan of certain patients. The proposed System works on the various Algorithm such as Background Removal, Image Enhancement, Pattern Matching which are used to enhanced the images and compute infection in the images. And doing so it is found that the algorithm works correctly and gives the accurate infection prediction of the images with less amount of time. The accuracy at which we can see the infected pattern lies in between 80-85%.

VIII. FUTURE SCOPE

The work can be further extended to an have inclusion of detection of the cancer cell with other the related diseases. The other diseases like cancer cells in the kidney region, brain region, etc. This system can work in the any diseases where MRIs and Ct-scan images such the grey images are computed for infection detection. we can improve the time required to analysis the images in future.

REFERANCES

[1] P. Sheba lin, V. Keilis - Borok , A. Gabri elov, I. Zaliapin, D.Turcotte, "Short-term earthquake prediction by reverse analysis of lithosphere dynamics," *ELSEVIER Tectonophysics* , 2006, pp.63 – 75

[2] Neeti Bhargava, V. K. Katiyar, M. L. Sharma and P. Pradhan, "Earthquake Prediction through Animal Behaviour: A Review," *Indian Journal of Biomechanics*: Special Issue NCBM 2009, pp.159-165.

[3] G. Molchan and L. Romashkova, "Earthquake prediction analysis: The M8 algorithm," *physics.geo-physics*, 2010, pp. 1-20

[4] Sajjad Mohsin, and Faisal Azam, "Computational seismic algorithmic comparison for earthquake prediction," *International Journal Of Geology* Issue 3, Volume 5, 2011, pp. 53-59.

[5] Chieh-Hung Chen, Horng-Yuan Yen,Chung-Ho Wang, Yih-Hsiung Yeh, Jann-Yenq Liu, Yee-Ping Chia, Chen Liu, Yetmen Wang and Wen-Tzong
Liang, "Identification of earthquake signals from groundwater level records using the HHT method," *Geophysical Journal International*, 2010, pp.1231–1241.

[6] D. Giaouris, J.W. Finch, "Denoising using wavelets on electric drive applications," *ELSEVEIR*, Electric Power Systems Research, 2008, pp.559–565.

[7] Radomir S. Stankovic and Bogdan J. Falkowski "The Haar wavelet transform: its status and achievements," *ELSEVEIR,* Computers and Electrical Engineering, 2003, pp.25–44.

[8] Burhan Ergen, "Signal and Image Denoising Using Wavelet Transform," Advances in Wavelet Theory and Their Applications in Engineering,
Physics and Technology, 2012, pp.495-514.

[9] Piotr Porwik and Agnieszka Lisowska, "The Haar–Wavelet Transform in Digital Image Processing:Its Status and Achievements," *Machine Graphics and Vision,* 2004, pp.79-98.

[10] Dr. Micheal Sek, "Frequency Analysis Fast Fourier Transform, Frequency Spectrum," Victoria University, pp.1