

COMPUTER AIDED DIAGNOSIS FOR LIVER CANCER USING STATISTICAL MODEL

Vincey Jeba Malar.V¹, Saravana Kumar. E²

¹M.E, ²M.E (PhD), Professor, Department of CSE, Adhiyamaan college of Engineering, Hosur, India
vinjen04@gmail.com, saraninfo@gmail.com

Abstract

Liver Cancer is one of the most difficult cancer to cure and the number of deaths that it causes generally increasing. The signs and the symptoms of the liver cancer are not known, till the cancer is in its advanced stage. So, early detection is the main problem. If it is detected earlier then it can be helpful for the Medical treatment to limit the danger, but it is a challenging task due to the Cancer cell structure. Interpretation of Medical image is often difficult and time consuming, even for the experienced Physicians. Most traditional medical diagnosis systems founded needs huge quantity of training data and takes long processing time. Focused on the solution to these problems, a Medical Diagnosis System based on Hidden Markov Model (HMM) is presented. This paper describes a computer aided diagnosis system for liver cancer that detects the liver tumor at an early stage from the chest CT images. This automation process reduces the time complexity and increases the diagnosis confidence.

Keywords—HMM, Segmentation, Feature Extraction.

-----***-----

1. INTRODUCTION

Liver cancer, also known as hepatic cancer is a cancer which starts in the liver, and not from another organ which eventually migrates to the liver. In other words, there may be cancers which start from somewhere else and end up in the liver - those are not (primary) liver cancers. Cancers that originate in the liver are known as primary liver cancers.

Liver cancer consists of malignant hepatic tumors (growths) in or on the liver. The most common type of liver cancer is hepatocellular carcinoma (or hepatoma or HCC), and it tends to affect males more than females. According to the National Health Service (NHS), UK [3], approximately 1,500 people in the United Kingdom die from HCC each year. The World Health Organization (WHO) [4] says that liver cancer as a cause of death is reported at less than 30 cases per 100,000 people worldwide, with rates in parts of Africa and Eastern Asia being particularly high. Experts say that common causes of HCC are regular high alcohol consumption, having unprotected sex and injecting drugs with shared needles[1], [2].

Signs and symptoms of liver cancer tend not to be felt or noticed until the cancer is well advanced. Hepatocellular carcinoma (HCC) signs and symptoms may include Jaundice, Abdominal pain, Unexplained weight loss, Hepatomegaly, Fatigue, Nausea, Emesis (vomiting), Back pain, General itching, Fever.

Liver cancer, if not diagnosed early is much more difficult to get rid of. The only way to know whether you have liver cancer early on is through screening, because you will have no symptoms. High risk people include those with hepatitis C and B, patients with alcohol-related cirrhosis, other alcohol abusers, and those that have cirrhosis as a result of Hemochromatosis. Diagnostic tests may include Blood test, Imaging scans (either an MRI or CT scan) and Biopsy (a small sample of tumor tissue is removed and analyzed).

Unfortunately, because symptoms do not appear until the liver cancer is well advanced, currently only a small percentage of patients with HCC can be cured; according to the National Health Service (January 2010) only about 5%.

Some of the CAD system uses Support Vector Machine (SVM), Fuzzy Logic, and Artificial Neural Network (ANN) algorithm. Their disadvantages are time consumption and needed a lot of data for training. Focused on the solution to the above problem, Hidden Markov Model (HMM) is presented for getting more advantage.

2. PROPOSED METHOD

2.1 CT Scan

A computerized axial tomography scan (CT scan or CAT scan) is an x-ray procedure that combines many x-ray images with the aid of a computer to generate cross-sectional views and three-dimensional images of the internal organs and structures of the body. A CT scan is used to define normal and

abnormal structures in the body and/or assist in procedures by helping to accurately guide the placement of instruments or treatments. It is a medical imaging method that employs tomography[11]. Tomography is the process of generating a two-dimensional image of a slice or section through a 3-dimensional object (a tomogram) (Figure 1).

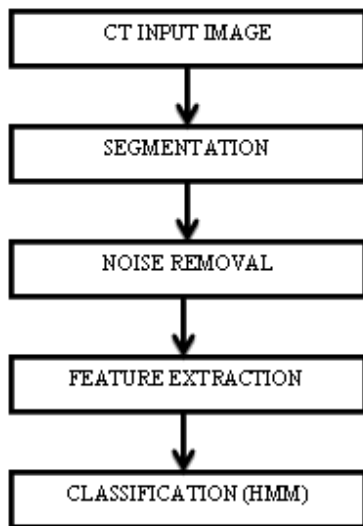


Fig-1: Block Diagram of Automatic Liver Cancer Detection

CT scans of the abdomen are extremely helpful in defining body organ anatomy, including visualizing the liver, gallbladder, pancreas, spleen, aorta, kidneys, uterus, and ovaries.

CT scans in this area are used to verify the presence or absence of tumors, infection, abnormal anatomy, or changes of the body from trauma. The Original chest CT image is shown (Figure 2).

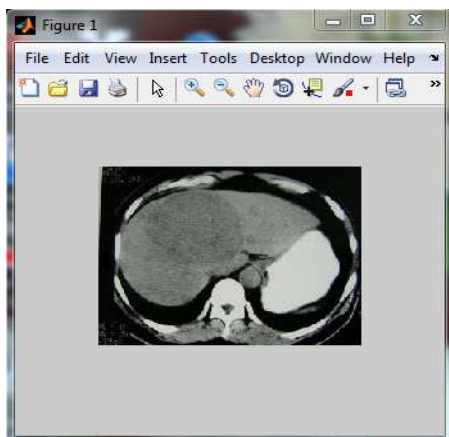


Fig-2: Chest CT image

The CT image is Sufficient for analysis for this proposed method. Moreover MRI Scan is very costly and the tissues can't able to view clearly. But the CT is not so costly but also the tissues can be clearly visible in CT scan.

2.2 Segmentation

After the primary noise removal, the segmentation has to be carried out. The goal of the segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images.

The segmentation of medical images of soft tissues into regions is a difficult problem because of the large variety of their characteristics[8]. All connected components in the eroded image are labeled and number of connected components is computed. The labeled components are segmented depending on region of interest.

Here the region growing techniques is used (Figure 3). This technique is enough for segmented the lung region. Moreover it will take less time. Here from a seed it starts growing. It will compare their neighbor-hood values and grows up.

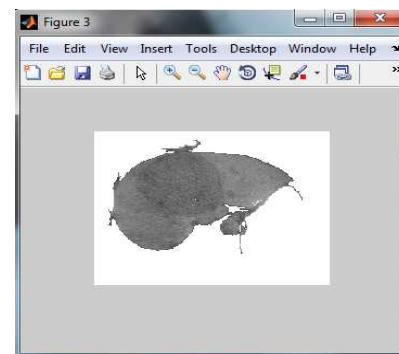


Fig-3: Segmented CT image using Region Growing

The segmentation process will result in separating the liver tissue from the rest of the image and only the liver tissues under examination are considered as the candidate region for detecting tumor in liver portion.

2.3 Noise Removal

The most important technique for removal of blur in images due to linear motion and also due to vibrations. Normally an image is considered as the collection of information and the occurrence of noises in the image causes degradation in the quality of the images. So the information associated with an image tends to loss or damage. It should be important to restore the image from noises for acquiring maximum information from images.

Since CT images contain more Gaussian noise, a Gaussian filter is used for noise removal. Gaussian filters are a class of linear smoothing filters[10]. The weights are chosen according to the shape of Gaussian function. The Gaussian smoothing filter is a very good filter to remove noise drawn from a normal distribution.

2.4 Morphological Operation

In order to smooth the liver boundaries and to retain the original liver shape, the morphological operations have to be carried out. The most basic morphological operations are dilation and erosion. Dilation means adding new pixels to the boundaries of an object in an image. Erosion means removing pixels from the boundaries of an object in an image.

The number of pixels added or removed from the object in an image depends on the size and the shape of the structural elements used to process the image.

2.5 Feature Extraction

When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (much data, but not much information) then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called feature extraction.

If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input[5],[12].

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. Analysis with a large number of variables generally requires a large amount of memory and computation power.



Fig-4: Feature Extraction

Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy.

2.6. Classification

Classification analyzes the numerical properties of various image features and organizes data into categories[6]. The classification algorithm employs two phases of processing namely training and testing.

In the initial training phase, characteristic properties of typical image features are isolated and based on these, training class is created. In the subsequent testing phase, these feature-space partitions are used to classify image features. The classification algorithm used in this proposed method is Hidden Markov Model(HMM)[7].

Hidden Markov Model(HMM)

A random sequence has the Markov property if its distribution is determined solely by its current state. Any random process having this property is called a Markov random process. For observable state sequences (state is known from data), this leads to a Markov chain model. For non-observable states, this leads to a Hidden Markov Model (HMM).

A hidden Markov model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states.

An HMM can be considered as the simplest dynamic Bayesian network. HMMs are composed of states, which are traversed according to transition probabilities. The sequence data is viewed as a series of observations emitted by the states, where an emission distribution over observations is associated with each state[13]. Formally, an HMM is characterized by three stochastic matrices, called the initial, transition and observation matrices.

The transition matrix, A , is a square matrix that holds the probabilities of transitioning from each state to any other. The probability of transitioning from state i to state j is denoted by a_{ij} . The initial distribution vector, π , is a column vector that stores the probabilities of starting in each state at the beginning of the sequence. π_i denotes the probability of starting in state i .

Finally, the observation matrix, B , defines the probabilities of observing each base pair for every state. The probability of observing observation k in state j is denoted by $b_j(k)$.

In a regular Markov model, the state is directly visible to the observer, and therefore the state transition probabilities are the only parameters. In a hidden Markov model, the state is not directly visible, but output, dependent on the state, is visible. Each state has a probability distribution over the possible output tokens. Therefore the sequence of tokens generated by an HMM gives some information about the sequence of states[9],[14].

Note that the adjective ‘hidden’ refers to the state sequence through which the model passes, not to the parameters of the model; even if the model parameters are known exactly, the model is still hidden (Figure 13).

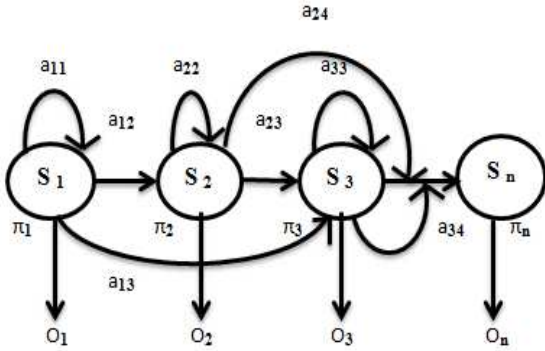


Fig-5: Structure of Hidden Markov model

Most selected Training algorithm used is the Baum- Welch re-estimation Formulas. The Hidden Markov Model is a finite set of states, each of which is associated with a (generally multidimensional) probability distribution. So according to the training of HMM with regarding the features it will reports the accrue out more effectively[9].

An artificial neural network is a structure which will attempt to find a relationship i.e. a function between the inputs, and the provided output(s), in order that when the net be provided with unseen inputs, and according with the recorded internal data (named “weights”), will try to find a correct answer for the new inputs.

The main difference could be this: In order to use a Markov chain, the process must depend only in its last state. For use a neural network, you need a lot of past data.

The Baum-Welch algorithm can be used to train an HMM to model a set of sequence data. The algorithm starts with an initial model and iteratively updates it until convergence. The algorithm is guaranteed to converge to an HMM that locally maximizes the likelihood (the probability of the training data given the model).

Since the Baum-Welch algorithm is a local iterative method, the resulting HMM and the number of required iterations depend heavily on the initial model. Of the many ways to generate an initial model, some techniques consider the training data while others do not.

The multiple observation training sequences are concatenated together to form one observation sequence for input into the Baum-Welch algorithm. All equations involved in the Baum-Welch process were obtained from Rabiner’s tutorial on

HMMs In a nutshell, the re-estimated HMM parameters π_i , a_{ij} and $b_j(k)$ are found using the following equations,

\tilde{O}_i = Expected frequency (number of times) in state N_i at time 1

$$a_{ij} = \frac{\text{expected number of transition from state } i \text{ to state } j}{\text{expected number of transition from state } i}$$

$$b_j(k) = \frac{\text{expected number of transition in state } i \text{ and observing symbol } k}{\text{expected number of times in state } i}$$

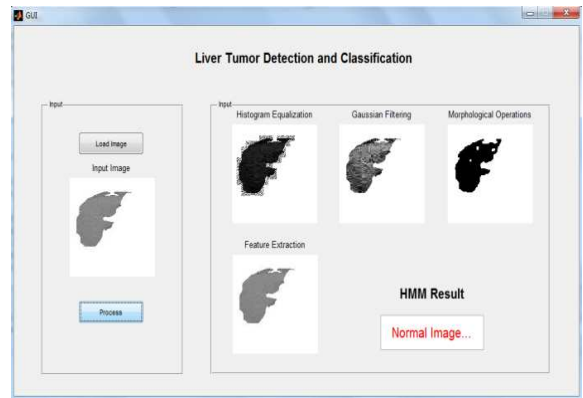


Fig-6: Liver Tumor Detection and Classification(1)

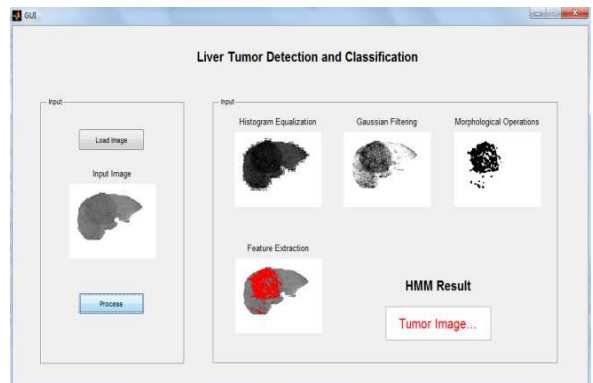


Fig-7: Liver Tumor Detection and Classification(2)

The comparison table(Table 1), bar graph(Chart 1), line graph(Chart 2)for HMM with other methods are shown below.

Table-1: Comparison table for HMM with other methods

Method	Detection Rate	Data Used
Back propagation method	89%	67 cases
Artificial neural networks	90%	11 cases

Lin	72%	29 Cases
Yamamoto	88%	7 Cases
Armato	72%	17 Cases
Hara	77%	8 cases
HMM	96.5%	100cases

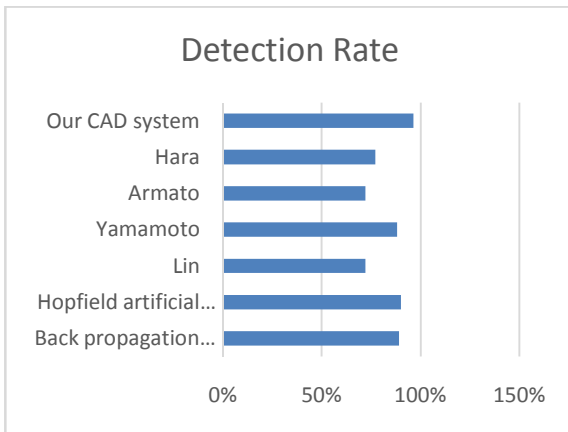


Chart-1: Bar graph for HMM detection rate with other methods

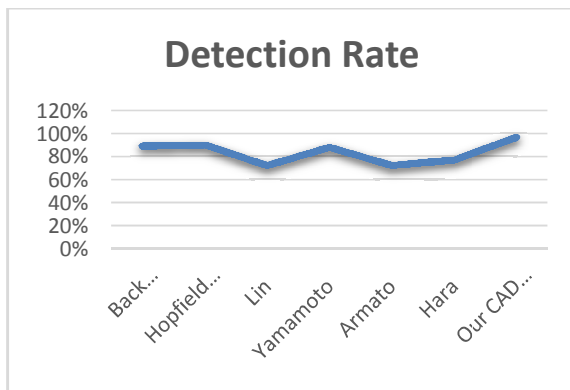


Chart-2: Line graph for HMM detection rate with other methods

CONCLUSIONS

In this paper, a novel method of segmenting the CT images been discussed. This research work carried out by taking 2-D CT images. The proposed work was carried out in 5 phases. In first phase, image acquisition of liver features and the second phase is related to the segmentation of ROI features of liver which can be determined using segmentation algorithm such as region growing approach. Third phase is removal of the noise. Fourth phase is feature extraction, it extract the corresponding liver nodule. Finally, the extracted liver nodules

are classified. In this paper we analyses the result for 2-D images (Figure 6 & 7). So early detection of Liver Cancer cells can be highly possible and it reduces the risk as well. This Bio-imaging method will enhance the proper radiotherapy treatment for Liver Cancer patients.

FUTURE SCOPE

By this process time complexity is reduced and diagnosis confidence is increased. A clear identification of liver Cancer whether it present or not by the CT images by using hidden Markov model. This process reduces the time complexity and increases the diagnosis confidence. The collected data contain noise, the noises are removed. And then segmentation of the liver images and after that the image is separated. According to the features of the defected part we can conclude the cancer is present or not and the patient is in which stage. For calculation the output image is trained by using the HMM model and the diagnosis is made from the output.

Our current investigation is to further obtaining a clear identification of Liver Cancer for 4-D CT image. And also we believe that combining CT screening with biomarker tests may help us to detect more liver cancers at earlier stage while reducing the number of biopsies or operations performed for non-cancerous abnormalities.

REFERENCES

- [1]. Cancer Research UK, liver cancer causes, signs and symptoms, <http://www.cancerresearchuk.org>
- [2]. National Cancer Institute, <http://www.cancer.gov/cancertopics>
- [3]. National Health Service, <http://www.Cancerresearchuk.org>
- [4]. World Health Organization, Liver cancer report, <http://www.cancerresearchuk.org>
- [5]. Anita chaudhary, SonitSukhraj Singh, “Lung Cancer detection on CT images by using Image Processing”, 2012 International Conference on Computing Sciences.featex
- [6]. AymnE.Khedr and Abd El-Ghany A. M. Mohmed, “A proposed image processing framework to support Early liver Cancer Diagnosis”, Life Science Journal 2012;9(4).Cla
- [7]. Blessingh.T.S, VinceyJebaMalar.V, Jenish.T, “CAD system for Lung Cancer using Statistical model and Biomarker”, CIIT International Journal Of Digital Image Processing, Volume 5, No 4, April 2013, ISSN 0974-9691.
- [8]. Clifford Samuel C, Saravanan V, Vimala Devi MR. Lung nodule diagnosis from CT image using Fuzzy Logic. International Conferenceon Computational Intelligence and Multimedia Applications. 2007.
- [9]. Ghahramani Z, Jordan MI. Factorial hidden Markov models. Machine Learning. 1997; 29: 245-273.

- [10]. Govindaraj.V, Sengottaiyan.G, “Survey of Image Denoising using Different Filters”, ISSN: 2278 – 7798, International Journal of Science, Engineering and Technology Research (IJSETR), Volume 2, Issue 2, February 2013.
- [11]. Kanazawa K, Kawata Y, Niki N, Satoh H, et al. Computer-aided diagnosis for pulmonary nodules based on helical CT images. *Compute.Med. Image Graph.* 1998; 22 (2): 157-167.
- [12]. Marius George Linguraru, William J. Richbourg, Jianfei Liu, Jeremy M. Watt, VivekPamulapati, Shijun Wang, and Ronald M. Summers, “Tumor Burden Analysis on Computed Tomography by Automated Liver and Tumor Segmentation”, *IEEE Transactions on Medical Imaging*, vol. 31, no. 10, october 2012.
- [13]. Porat S, Feldman JA. Learning automata from ordered examples. *Machine Learning* 7. 1991; p109-138.
- [14]. Rabiner LR. A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proc. IEEE.* 1989; 77: 257-286.