A Pragmatic approach for Detecting Liver Cancer using Image Processing and Data Mining Techniques

Anisha P R
Stanley College of Engg. & Tech. for Women, Hyderabad, India
E-mail: pranisha@stanley.edu.in

Kishor Kumar Reddy C
Stanley College of Engg. & Tech. for Women, Hyderabad, India
E-mail: kishoar23@gmail.com

Narasimha Prasad L V
Institute of Aeronautical Engineering Hyderabad, India
E-mail: lvnprasad@yahoo.com

***Abstract*— Cancer diagnosis and treatment has a great significance due to the prevalent episodes of the diseases, high death rate and reappearance after treatment.. On the world scale, cancer stands in the fifth position which causes death. Among the various cancers, liver cancer stands in the third position. Liver cancer is generally diagnosed by three different test like blood test, image test and biopsy. To make the task of detecting the liver cancer simpler, less time consuming, an effective and efficient approach is adopted for the same. In this research a computer aided diagnostic system for detecting liver cancer is put forward. 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. The experiment put forth gave an average accuracy of 82% besides reducing the time complexity and computational complexity of the test.**

Keywords– Data mining; Haar wavelet transform; Image processing; k-means clustering; Liver cancer.

I. INTRODUCTION

Cancer which is clinically referred as a malevolent neoplasm is a extensive group of diseases, involving unregulated cell growth. In cancer, cells subdivide and grow hysterically, forming malignant tumours, and invade nearby parts within the body. These tumours can grow and hinder the digestive, nervous, and circulatory systems and releases hormones that may alter the body functionality. There are about 200 different known cancers that are shown in human. Each of these are characterised by the type of the cell that is first affected.

Liver is considered to be one of the major internal organs in the human body, which plays a major role in the body's metabolism, serving several vital functions that supports every other organ in the body. Research is carried out across the globe, looking at ways to prevent the occurrence of liver cancer. As there are very few effective ways to prevent this disease, at this juncture, there is always a great deal of research going on in the area. Scientists are experimenting for causes and ways to thwart liver cancer, on the other hand doctors are striving hard to improve the nursing procedure. The most effectual approach to reduce the worldwide burden of liver cancer is to avert it from happening in the first place. Some scientists believe that vaccinations and improved treatments for hepatitis could prevent about half of liver cancer cases worldwide. Researchers are studying ways to prevent or treat hepatitis infections before they cause this disease.

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. Fig.1 shows the physical appearance of liver, the cause of infection, the toxic damages, and the immunological damages resulting this fatigue infection.

.

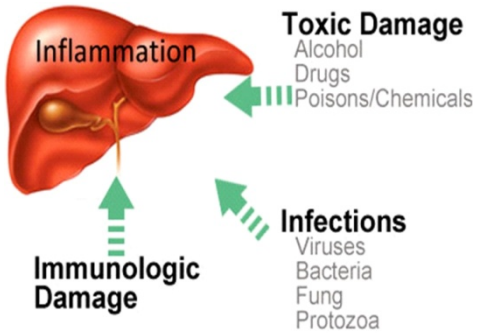

Fig.1: Structure of liver area where it could get damaged.

Endurance of a cancer patient depends heavily on early detection, and thus, developing technologies applicable for sensitive and specific methods is an inevitable task for cancer researchers. Fig. 2 depicts the percentages of various causes that give birth, develop this disease. The bar graph shows that over consumption of Alcohol and Hepatitis C virus stands as the major cause for this growth and development of this cancer cell.

Existing cancer screening methods include: (1) the Papanicolau test for women to detect cervical cancer and mammography to detect breast cancer, (2) occult blood detection for colon cancer (3) endoscopy, (4) prostate-specific antigen (PSA) level detection in blood sample for men to detect prostate cancer, and CT scans, X-ray, ultrasound imaging [2] and MRI for various cancer detection. These traditional diagnostic methods however are not very powerful methods when it comes to cancer detection at very early stages. Apart from this, some of

the screening methods are quite expensive making it unaffordable and unavailable to the common man.

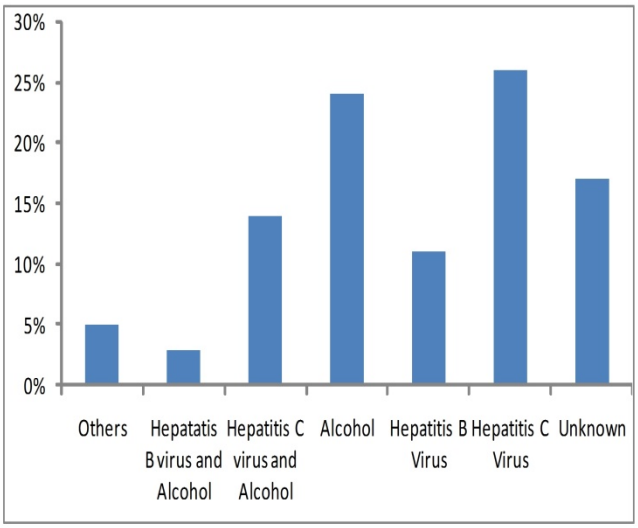

Figure 2: Causes of liver cancer and their damages (%)

Therefore, with the growth in technology it has become an utter importance to having a mechanism that is specific and reliable for detecting cancers at early stages and is easily accessible so that it can function as the first-line guidance. With an increase in cancer effect and the death toll due to this disease and a lack of early detection it gave a motivation to present an idea that would not only be a novel approach, less time complex, less computational complex but also could be made available for all section of the people.

The liver cancer can be diagnosed in 3 ways namely Blood test, Imaging test and Biopsy. To get a computer assisted technique, the focus is mostly given to the Imaging tests. The Imaging tests used in the diagnosis of liver cancer are Ultrasound, Computed tomography (CT), Magnetic resonance Imaging (MRI) and Angiography. Computer assisted liver tumour detection and classification which is based on image analysis techniques provides more useful information. The conventional methods for liver cancer tissue classification consist of three steps: Firstly, segmentation [3] [5] of liver and tumour from CT abdominal images, secondly features are extracted and finally classification is done using classifiers. Over the years the characterisation of liver images based on texture study techniques is been developed over the years. Researchers have shown that, although the wavelet transforms are very effective while they are representing objects with isolated point singularities, but when line singularities are considered they are not so efficient. Many have worked on the ridgelet transform, curvelet transform and other transformation techniques previously.

In the present research, we adopted Haar wavelet transform to analyse the images. The Haar technique has two advantages over other techniques. Firstly Haar automatically converts a greyscale image to RGB image and secondly it considers any discontinuous image and compresses it accordingly.

The methodology followed in this paper can be explained as follows. Firstly we consider a greyscale image of the liver. To get better knowledge about the image we convert it to an RGB image. The liver part is extracted from the image and clusters are formed using the k-means clustering [15] technique. The most appropriate cluster is selected from these clusters and the number of pixels for the cancer part is calculated. These values are compared with the theoretical values. Finally, a mean is computed based on calculated and theoretical value. If all the three values fall in the same range, it is identified as liver cancer, and this is explained clearly in section III.

The paper is structured as follows: Section II points out various methodologies followed previously related to the same or similar aspects. Section III describes the experimentation methodology followed in this paper. Section IV gives the experimental results. Conclusion and future enhancements is presented in Section V.

II. RELEVANT WORK

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 multi-resolution 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. Wu Qiu et al [6] made use of Fuzzy technique for the image enhancement by making use of frequency domain and spatial domain methods. Alexandra Branzan Albu et al. [7] describes a segmentation technique for 2D interventional MR images of liver tumours in order to extract the targeted tumour with high accuracy and reliability. Katia Passera et al. [8] made use of radiofrequency ablation which is a non surgical treatment for hepatic tumors which put forward a

realistic system aiming at providing a more objective tool for the evaluation of RFA coverage.

Narendra D Londheal et al. [9] proposed a method using region based coding of liver cancer images framed through CT scan. Masayuki Matsuo et al. [10] put forth a study that compare the detectability of malignant heptatic tumors on ferumoxides-enhanced MRI using five gradient-recalled echo sequences at different TEs. K. Mala et al [11] gave a system which is used to segment the tumor with so as to generate the required basis for the detection. Bi-orthogonal wavelet based texture features can be extracted which can be used to train the PNN to classify the liver tumor as hepatocellular carcinoma, cholangio carcinoma, hepatocellular abdomen and hemagioma with better performance that supports the radiologists and the medical specialists during their medical decision.

Robin Martin, et al. [13] implemented a semi-automatic method based on region-growing to isolate the liver part. Smitri et al, [14] 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 [15] 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 [16] in their paper gave a brief on wavelet transform of which Haar wavelet transformed is elucidated.

Pansnur M.A and P.S. Malge [17] represent the technique of image retrieval using Haar wavelet transform and K means clustering technique. Pushpa R. Suri and Mahak [18] in their work presented a brief on the techniques for Image Segmentation of which the number of clusters that can be used for the segmentation purpose is represented. Sanjeev Kumar and Varun Sood [19] examines a set of wavelet functions (wavelets) so as to implement the same on a still image compression system and briefed certain features of wavelet transform that include the quality of images when compressed and decompressed..

III. EXPERIMENTATION METHODOLGY

The objective of this research work is 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 R2011a software in order to process the images. The overall methodology is elucidated by the segmentation process represented in Fig. 3. As the images that are framed via the MRI/CT scan exist as a grayscale image as shown in Fig. 4, it creates a discrepancy in identifying the cancer cell which may mislead the experimentation process, therefore the image is again converted to RGB image as shown in Fig. 5, 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 [18]. The clusters thus formed by the above mentioned process is shown in Fig. 6. Now the clustered image that depicts the extracts of cancer cell which is further used for detection process, shown in Fig. 7. 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.

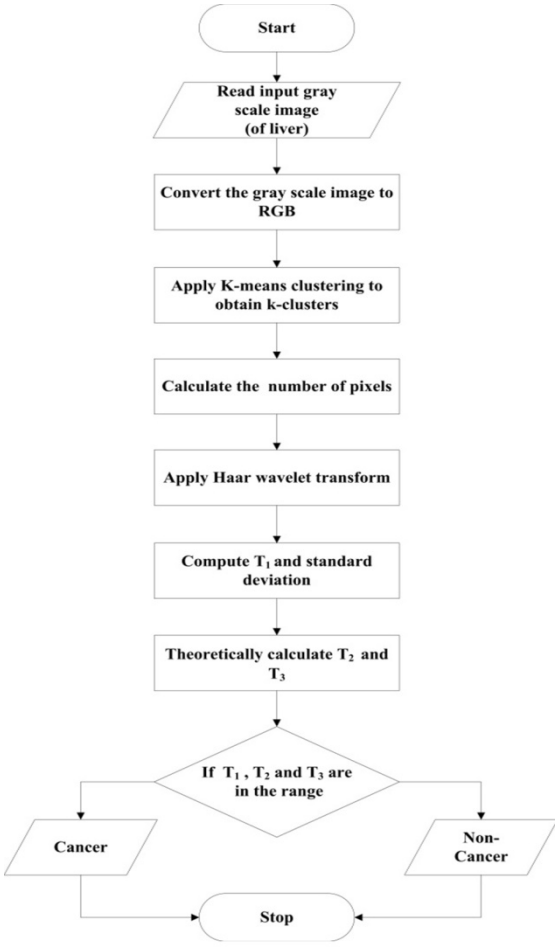

Fig. 3. Methodology to detect cancer cells

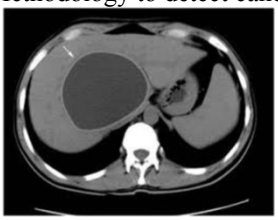

Fig. 4. Original Grayscale Liver Cancer Image

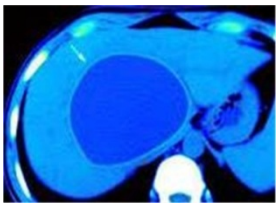

Fig. 5. Converted RGB Liver Cancer Image

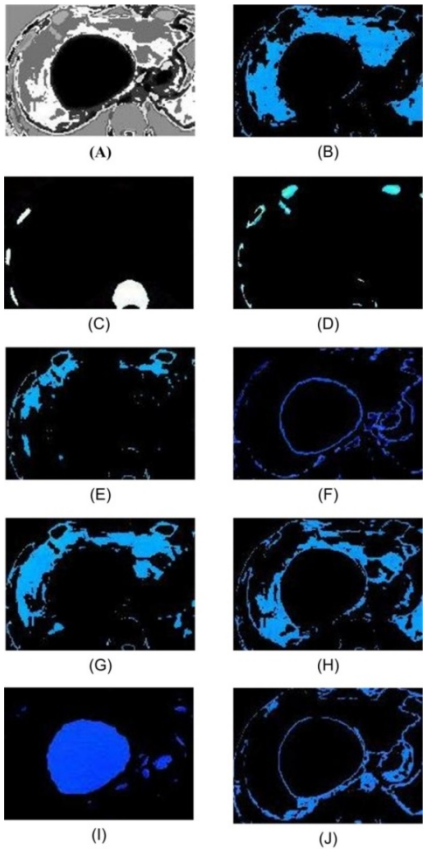

Fig. 6. Segmented images for RGB liver Cancer Image

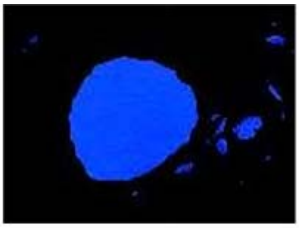

Fig. 7 Clustered Region of Interest

As the area of the affected part is to be calculated, the number of pixels has to be determined which is done using MATLAB software. Further on processing the clustered image, the mean value for the pixels is calculated using which the variance is obtained. Finally the obtained values are hence forth substitute to find the final calculated threshold using equation 1, this threshold value is calculated using the square root of 2 multiplied with the product of variance and log(n) value which is divided to n and then multiplied to k where n represents the number of pixels which is obtained by the clustered image by area calculation and k represents the number of clusters formed.

$$T_2 = \sqrt{\frac{2*\sigma^2*\log(n)}{n}} * k \qquad (1)$$

Further the MATLAB obtained threshold value and the calculated value are compared and finally the mean of these two threshold values is calculated, using equation 2, which further results in generating a final threshold value to justify if the given clustered image carrying the affected part is indeed a cancer cells or not. Here finally images are justified by analysis and comparison method wherein, if the threshold value obtained through the software (T_1), the calculated threshold value (T_2) and finally the average value of the above thresholds (T_3) falls within the same range with a marginal variation which can be neglected, depicts the existence of cancer cell or not.

$$T_3 = \frac{T_1+T_2}{2} \qquad (2)$$

If all the threshold value falls within the same range the image is said to have cancer cells effected, if the variations are very much out of boundary then it is considered to be a no cancer part. The results are established and tabulated in Table I. Consider Image 1, its T1 value is 8.50, T2 value is 9.10 and T3 value is 8.80. All the three values fall in the same range hence the experimental result is cancer. Further, we compared the experimental result with historical data. For Image 1, both experimental and historical results are cancer and hence, the final prediction is true.

IV. Experimental Results

Cancer detection is one of the challenging tasks that researchers of present day are experiencing. After a lot of research, still there is a lack of an accurate model as the detection is not unidisciplinary but is a multi-disciplinary task as it depends on various parameters. A lot of research has been performed on the development of accurate techniques for the detection of cancer at the earlier stage.

Table I. Computed and Established Threshold Values

Image No	Number of Pixels (N)	Variance	T_1	T_2	T_3	Experimental result	Historical Data	Prediction
Image 1	15.81	5.46	8.50	9.10	8.80	cancer	cancer	TRUE
Image 2	10.28	3.72	8.81	8.56	8.69	cancer	cancer	TRUE
Image 3	14.85	3.74	7.83	7.68	7.76	cancer	cancer	TRUE
Image 4	8.36	2.88	7.38	7.97	7.67	cancer	cancer	TRUE
Image 5	10.71	3.23	7.88	7.88	7.88	cancer	cancer	TRUE
Image 6	7.27	1.88	6.69	6.67	6.68	cancer	cancer	TRUE
Image 7	21.41	6.03	8.25	8.80	8.52	cancer	cancer	TRUE
Image 8	15.27	3.46	7.38	7.33	7.35	cancer	cancer	TRUE
Image 9	8.40	3.73	7.13	9.06	8.09	cancer	cancer	TRUE
Image 10	18.76	5.00	8.00	8.24	8.12	cancer	cancer	TRUE

This paper put forward a novel approach for the detection of cancer at an early stage which is comparatively accurate, less time complex and easily calculative.

Table I shows the exemplary experiment that is carried out, consider image 2 whose T_1 value is 8.81, T_2 value is 8.56 and T_3 value is 8.69, all the three threshold values falls within the same range with a marginal variation which can be negligible, on analysis they said that image is considered to be cancer effected and based on comparison with the historical assessed data the image resulted to be a cancer affected image. This made ways for accurate prediction.

The same procedure was followed and tested with the rest of the collected clustered image as shown in Table I. Finally experimentation is done on various such images and the result is shown in Table II making path with an accuracy of about 82%, by far reducing the time and computational complexity for the detection of cancer.

V. CONCLUSION

Research is said to be a continuous, unavoidable, necessary, experimenting, an innovative and result based process, In brief to the fact, this paper this 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. This diagnostic application makes use of MATLAB software for processing of the image, by making use of Haar wavelet transformed and clustering techniques. The whole analysis is done based on the threshold values and the images are justifying by checking if the threshold falls within the same range estimated for each image.

The result obtained found to be a pragmatic approach for the early and accurate detection of cancer cells. The experimentation gave an accuracy of about 82% besides being less time complex, reducing the computational complexity for the purpose of detection.

REFERENCES

[1] P. Shebalin, V. Keilis-Borok, A. Gabrielov, 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] Claudio Satriano, Yih-Min Wu, Aldo Zollo, Hiroo Kanamori, "Earthquake early warning: Concepts, methods and physical grounds," *ELSEVEIR Soil Dynamics Earthquake Engineering* 2010, pp. 1-13.

[7] Hiroo Kanamori, "Magnitude Scale and Quantification of Earthquakes*,"Elseveir Scientific Publishing Company*, Amsterdam. Tectonophysics, 1983, pp.185-199

[8] John Stockwell, "Some empirical relations in earthquake seismology," *Center for Wave Phenomena Colorado School of Mines*, 2001, pp.1-6

[9] Giuseppina Immè and Daniela Morelli, "Radon as Earthquake Precursor*," Dipartimento di Fisica e Astronomia Università di Catania - INFN Sezione di Catania*, Italy, 2012, pp.143-160.

[10] Arvind Kumar, Vivek Walia, Surinder Singh, Bikramjit Singh Bajwa, Sandeep Mahajan, Sunil Dhar, and Tsanyao Frank Yang, "Earthquake precursory studies at Amritsar Punjab, India using radon measurement techniques," *International Journal of Physical Sciences*, 9 November, 2012, pp.5669-5677.

[11] W. H. K. Lee, J. M. Espinosa-Aranda, "Earthquake Early-Warning Systems: Current Status And Perspectives," United States Geological Survey (USGS), 2010, pp.409-423.

[12] Kuo-Liang Wen, Tzay-Chyn Shin, Yih-Min Wu, Nai-Chi Hsiao and Bing-Ru Wu, "Earthquake Early Warning Technology Progress in Taiwan," *Journal of Disaster Research* Vol.4 No.4, 2009, pp 202-210.

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

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

[15] 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.

[16] 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.

[17] Dr. Micheal Sek, "Frequency Analysis Fast Fourier Transform, Frequency Spectrum," Victoria University, pp.1-12.

[18] Lynn R. Sykes, Bruce E. Shaw and Christopher H. Scholz," Rethinking Earthquake Prediction," *Pure applied geophysics*, 1999, pp-207–232.

[19] Robert J**.** Geller," Earthquake prediction: a critical review," *Geophysical Journal International***,** 1997, pp- 425-450.

[20] Stefan Wiemer, "Earthquake Statistics and Earthquake Prediction Research," *Institute of Geophysics,* ETH Hönggerberg, Zürich, Switzerland, 2003, pp.1-11.

[21] Hiroo Kanamori, "Earthquake Prediction: An Overview," *International Handbook of Earthquake and Engineering Seismology*, 2003, pp.1205-1216.

[22] Toshi Asda, "Earthquake Prediction Study in Japan," *Proceedings of Ninth World Conference on Earthquake Engineering*, 1988, pp.13-19.

[23] C. G. Sammis and D. Sornette, "Positive feedback, memory and the predictability of earthquakes," 2002, pp 2501–2508.

Image No	T_1	T_2	T_3	Experimental Result	Historical Result	Prediction
1.jpg	15.75	9.18	12.46	Non-cancer	Non-cancer	TRUE
2.jpg	15.75	8.6	12.17	Non-cancer	Non-cancer	TRUE
3.jpg	15.75	10.87	13.31	Non-cancer	Non-cancer	TRUE
4.jpg	15.69	10.88	13.29	Non-cancer	Non-cancer	TRUE
5.jpg	14.13	8.02	11.07	Non-cancer	Non-cancer	TRUE
6.jpg	15.81	8.66	12.23	Non-cancer	Non-cancer	TRUE
7.jpg	15.56	8.59	12.07	Non-cancer	Non-cancer	TRUE
8.jpg	15.5	8.18	11.84	Non-cancer	Non-cancer	TRUE
9.jpg	15.25	10.04	12.64	Non-cancer	Non-cancer	TRUE
10.jpg	15.75	10.11	12.93	Non-cancer	Non-cancer	TRUE
11.jpg	15.88	9.61	12.75	Non-cancer	Non-cancer	TRUE
12.jpg	14.56	9.98	12.27	Non-cancer	Non-cancer	TRUE
13.jpg	6.69	6.18	6.43	Cancer	Cancer	TRUE
14.jpg	6.19	6.94	6.56	Cancer	Cancer	TRUE
15.jpg	6.75	6.12	6.44	Cancer	Cancer	TRUE
16.jpg	6.69	6.27	6.48	Cancer	Cancer	TRUE
17.jpg	6.38	3.58	4.98	Cancer	Non-cancer	FALSE
18.jpg	8.75	10.09	9.42	Cancer	Non-cancer	FALSE
19.jpg	5.25	4.19	4.72	Cancer	Non-cancer	FALSE
20.jpg	6.63	5.43	6.03	Cancer	Non-cancer	FALSE
21.jpg	6.38	5.41	5.89	Cancer	Non-cancer	FALSE
22.jpg	7.06	5.55	6.31	Cancer	Non-cancer	FALSE
23.jpg	6.38	5.35	5.86	Cancer	Non-cancer	FALSE
24.jpg	5.81	4.96	5.39	Cancer	Non-cancer	FALSE
25.jpg	6.25	6.96	6.61	Cancer	Cancer	TRUE
26.jpg	6	6.17	6.08	Cancer	Cancer	TRUE
27.jpg	7.44	7.73	7.58	Cancer	Cancer	TRUE
28.jpg	7.5	7.32	7.41	Cancer	Cancer	TRUE
29.jpg	8.5	8.41	8.46	Cancer	Cancer	TRUE
30.jpg	9.06	9.08	9.07	Cancer	Cancer	TRUE
31.jpg	8.88	8.77	8.82	Cancer	Cancer	TRUE
32.jpg	7.25	6.06	6.65	Cancer	Cancer	TRUE
33.jpg	9	9.53	9.27	Cancer	Cancer	TRUE
34.jpg	8.06	8.18	8.12	Cancer	Cancer	TRUE
35.jpg	7.13	7.32	7.22	Cancer	Cancer	TRUE
36.jpg	6.38	6.78	6.58	Cancer	Cancer	TRUE
37.jpg	6	6.02	6.01	Cancer	Cancer	TRUE
38.jpg	6.25	6.24	6.25	Cancer	Cancer	TRUE
39.jpg	8.25	8.8	8.52	Cancer	Cancer	TRUE
40.jpg	7.06	7.86	7.46	Cancer	Cancer	TRUE

Table II. Experimenatl results for Liver Cancer