Discovered Facts Related to Liver Cancer Diagnosis – A Research Perspective

Manish Tiwari
Research Scholar
Department of Computer
Science and Engineering, Mewar University,
Chittorgarh-312901, Rajasthan, India

Prasun Chakrabarti
Department of Computer Science and Engineering
Sir PadampatSinghania University, Udaipur-
313601, Rajasthan, India

ABSTRACT
The authors have proposed some discovered facts related to liver cancer diagnosis .The analysis of intensity of liver cancer growth can be governed by artificial neural modeling. The progress of treatment of liver cancer based on past event (intensities of cancer growth at specific observed timing instants) can be computed on the basis of neuro-associator. The augmentation or expansion of features indicating liver cancer growth can be quantified and realized based on Markov property based state transition. The investigation related to liver cancer detection can be governed by the fundamental concept of geometric distribution. The investigation related to significance of parameters responsible for liver cancer can be realized in the light of the Cobb-Douglas model. Liver cancer detection can be realized based upon the fundamental principle of information gain. The reliability and mean time to failure of liver cancer testing system can be realized in the light of parallel system configuration. The effect of alcohol consumption leading to liver cancer can be sensed using the concept learning approach.

Keywords
neuro-associator, Markov property based transition, Cobb-Douglas model, information gain, reliability, concept learning.

1. INTRODUCTION
As in 20th century cancer is one of the uncured diseases in human body. Cancer is describing as uncontrolled growth of cells that attack on healthy tissues. Researchers have given many proposals on cancer classification [1]. One of the cancer dieses is Hepatcellular carcinoma (HCC), Predication of HCC is difficult due to Information of liver cancer comes to know at last stage[2,3]. The cancer start in the liver is known as hepatic cancer or primary liver cancer, it consist of malignant hepatic tumors growths in or on liver. Hepatcellular carcinoma is most common type of liver cancer it affects male more than females[4]. According to National Health Service (NHS) 1,500 people die from HCC in United Kingdom[5]. According to World Health Organization, death rate in part of Africa and Eastern Asia is particular high[6]. HCC is cause of regular high alcohol consumption, having unprotected sex and injecting drugs with shared needles. Symptoms of Hepatocellular carcinoma (HCC) are jaundice, abdominal pain, unexpected weight loss etc. However till date the analysis of liver cancer in the light of concept learning, simulation models and machine intelligence has seldom been carried out [7,8].

2. ARTIFICIAL INTELLIGENCE BASED MODELING OF LIVER CANCER GROWTH INVESTIGATION

2.1 Fact 1- The analysis of intensity of liver cancer growth can be governed by concept learning and artificial neural modeling.

Justification -The objective is to investigate the intensity of liver cancer growth. In the view of above, the proposed concept learning based 3-tuple hypothesis indicated $(£,€,α)$ where £ indicates any value that is accepted , € indicates a necessary and sufficient value that will indicate the margin representing benign, α indicates a value representing malignant. Let the random variables $μ1, μ2, μ3$ and $μ4$ denote the probabilistic measure of cancer due four distinct symptoms. Accordingly the variables are to be mapped with the 3-tuple form $(£, €, α)$. Let us take X is a potential value that matches with $μ3$ reporting benign only. Since no trend analysis is available, the alternate hypothesis leads to be mapped 3-tuple representation $(μ3, φ, Max(μ1, μ2, μ4))$ where φ indicates null attribute. Based on the statistical assertion, the concept learning based mapped representation is as follows–

$(£, Max(μ1, μ2, μ3, μ4), φ)$ if X is less intense representing benign.

$(£, φ, Max(μ1, μ2, μ3, μ4))$ if X is more intense representing malignant.

As preposed artificial neural based model based on the input parameters (quantified value in 1-point scale as per hypothesis) related to intensity computation based on symptoms. The output received by the soma of artificial neuron is

$$δ = \sum_{i=1}^{n} wi * xi \quad \text{……………………………………..…(1)}$$

where wi is the corresponding weight of the symptom xi.

The bias is the output received by the soma of the artificial neuron. Hence, the actual output of the proposed neural model can be represented as the bivalent form (-1, +1) based on the following conditions-

If $δ >$ bias, then output is +1 indicating more intensity i.e. malignant and if $δ <$ bias ,then output is -1 indicating less intensity i.e. benign. The aforesaid mathematical model facilitates the treatment of a patient suffering from liver.

2.2 Fact 2 – The progress of treatment of liver cancer based on past event (intensities of cancer growth at specific observed timing instants) can be computed on the basis of neuro-associator.

Justification – Let $PF_{(Ei)}$ be the priority factor of event E_i (i=0,1,…n) that may be either improvement or deterioration. Based on optimum priority factor, probability of intensity forecasting of cancer growth in future timing instant increases considerably.

Cancer growth stability normally exhibits a decay effect and therefore $PF_{(Ei)}$ varies with time. As per a monotonically decreasing function for each priority factor,

$$PF_{(Ei)}(t) = \max(0, PF_{(Ei)}(t-1) - \beta)\dots\dots\dots\dots\dots\dots\dots..(2)$$

Where a small positive constant $\beta < 1$ signifies the rate of minor deviation in future growth stability relative to the past event E_i.

The training set in this future forecasting perspective may be viewed as follows-

$$T = \{(RC_{Ei}, RUC_{Ei}) : i = 0,1,\dots.n\}\dots\dots\dots\dots\dots\dots\dots(3)$$

where each pattern being bivalent {0,1} ;
0 signifies impact factor on quantitative predictive growth estimate less than or equal to bias indicating improvement and response to treatment,

1 signifies that exceeding bias value £ indicating deterioration

RC_{Ei} being impact estimate in Boolean form of event E_i in present state and

RUC_{Ei} impact estimate in Boolean form of event E_i in future state.
The priority weight is computed as
$$W_{x,y}E_i = 1/3 [RC_{Ei,x} * RUC_{Ei,y} * PF_{(Ei)}] ; i=0,1\dots., n\dots\dots(4)$$

where x, y denotes mapping of xth sequence of actual event $E(i)$ with j^{th} sequence of hypothetical event $E(i)$; x=0,1,…m and y=0,1,….n (m not necessarily equal to n).Now as per normal decay in cancer growth stability ,$W_{x,y}E_i$ also changes accordingly from time to time. Therefore with respect to time,

$$W_{x,y}E_i(t) = (1-\beta) W_{x,y} E_i(t-1) + RC_{Ei,x} * RUC_{Ei,y} * PF_{(Ei)}(t-1)\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(5)$$

where $(1-\beta)$ denotes the attenuation factor. Hence it is justified to state that "The progress of treatment of liver cancer based on past event (intensities of cancer growth at specific observed timing instants) can be computed on the basis of neuro-associator'.

2.3 Fact 3 – The augmentation or expansion of features indicating liver cancer growth can be quantified and realized based on Markov property based state transition.

Justification – The augmentation or expansion of features indicating liver cancer growth can be quantified based on the following relationship

$$Q_{s,t} = \frac{|t_p - t_a|}{t_a} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(6)$$

where t_p be the predicted estimation of viability of the augmented measure at future timing instant t, t_a be the predicted estimation of viability of the existing cancer growth rate at timing instant t based on trend analysis and $Q_{s.t}$ be the measure of feature deviation indicating the probable state transition from benign to malignant (in case of deterioration) or malignant to benign (in case of improvement). After n successive stages of sensing predicted estimation of viability of the augmented measure of liver cancer growth, the effective feature deviation is represented as follows-

$$£_e = \frac{\max(t_{p_1}, t_{p_2}, t_{p_3}\dots t_{p_n}) - t_a}{t_a} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(7)$$

Therefore, the mean deviation of t_{pi} about t_a is given by

$$£_{t_a} = \frac{1}{n} \times \sum_{i=1}^{n} |t_{p_i} - t_a| \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(8)$$

Let s=$(s_1, s_2\dots.s_n, s_n)$ denotes a sequence of states of prediction estimates of the augmented measure, f_i^* be the frequency of state transition over a specific period of observation and $w_{s_i > s_{i+1}}$ be the weight of state transition from i^{th} state to $(i+1)^{th}$ state In this perspective, the model resembles that of Markov property and the weight computations are stated as below-

$$w_{s_1 > s_2} = \frac{f_i^* \times |t_{p_1} - t_a|}{|t_{p_1} - t_a| + |t_{p_2} - t_a| + \cdots |t_{p_n} - t_a|}\dots\dots\dots\dots\dots\dots\dots(9)$$

In generalized version,

$$w_{s_i > s_{i+1}} = \frac{f_i^* \times |t_{p_{i+1}} - t_{p_i}|}{|t_{p_1} - t_a| + |t_{p_2} - t_a| + \cdots |t_{p_n} - t_a|}\dots\dots\dots\dots\dots\dots\dots(10)$$

Hence, it is justified to state that "The augmentation or expansion of features indicating liver cancer growth can be quantified and realized based on Markov property based state transition".

3. DISCOVERED FACTS OF LIVER CANCER IN THE LIGHT OF GEOMETRIC DISTRIBUTION, COBB-DOUGLAS MODEL AND INFORMATION GAIN

3.1 Fact 1 – The investigation related to liver cancer detection can be governed by the fundamental concept of geometric distribution.

Justification– The role of the doctor is to perform continuous independent trials and observe the timing instants of success. In this context, success indicates that the liver cancer of an individual has been detected successfully at specific timing instant.

Let p be the probability of success and occurrence of event S and m be the number of independent trails till S is incident. Hence m is discrete with range 1,2,3,…. and an event {m=n} indicates the validity of the sequence FF…….FS. Hence, by independence, the above fact has the probability $p(1-p)^{n-1}$. This signifies the validity of the basic principle of geometric distribution. Let X is the probability that the liver cancer can be detected successfully by independent trial sand Y is the probability that liver cancer detection has a superior competitor of proper diagnosis.

Therefore,

$$\alpha = E[m|S]X + E[m|F]Y = 1 * p + (1 + \alpha)(1 - p)....(11)$$

Hence, from Eq (11) we get α = 1/p.

This reveals that the conditional expectation, that the geometric distribution based diagnosis is the best fit technique, can be derived from the success probability of the technique.

3.2 Fact 2 –The investigation related to significance of parameters responsible for liver cancer can be realized in the light of the Cobb-Douglas model.

Justification– Understanding from literature survey, genetic and addiction (alcohol / tobacco) are the two prime factors leading to liver cancer. Correlating the above fact with the Cobb-Douglas model,

$$S = \beta M^x R^y ………………………………………..(12)$$

where S is the liver cancer growth measure , β is a constant , M is genetic relevance , R is intake of elements related to addiction , and exponents being x and y .When **x = y = 1**, then the percentage change in output S equals the percentage change in any one of the two variables M and R.

3.3 Fact 3 –Liver cancer detection can be realized based upon the fundamental principle of information gain.

Justification– The classified age groups are 1-20 and 21-40.

Assumptions (1-20 age)

1. Genetic factor is absolutely important for the occurrence of liver cancer.
2. Tobacco, Alcohol consumption and Diet are not so prominent in this age.
3. During this age normally the individual does not consume alcohol or tobacco too much.
4. Enhancement of diet will help in gain of the energy at the growing stage.

Table1. Effect on age group (1-20)

S. No.	Genetic	Alcohol	Tobacco	Diet	Chance of liver cancer
1	High	Low	Low	Low	High
2	High	Low	Low	High	High
3	High	Low	High	Low	High
4	High	High	Low	Low	High
5	High	High	High	Low	High
6	High	High	Low	High	High
7	High	Low	High	High	High
8	High	High	High	high	High
9	low	Low	Low	low	Low
10	low	Low	Low	high	Low
11	low	Low	High	low	Low
12	low	High	Low	low	Low
13	low	High	High	low	Low
14	low	High	Low	high	Low
15	low	Low	High	high	Low
16	low	High	High	high	High

$$Info_{(1-20)}(D) = -\frac{9}{16}\log_2\frac{9}{16} - \frac{7}{16}\log_2\frac{7}{16}$$

$$Info_{(1-20)}(D) = .988$$

Hence, the expected information needed to classify a tuple in D if the tuples are partitioned according to *Genetic* is

$$Info_{genetic}(D) = \frac{8}{16} \times \left(-\frac{8}{8} \times \log_2\frac{8}{8} - \frac{0}{8} \times \log_2\frac{0}{8}\right) + \frac{8}{16}\left(-\frac{1}{8} \times \log_2\frac{1}{8} - \frac{7}{8} \times \log_2\frac{7}{8}\right)$$

$$Info_{genetic} = .272\ bits$$

Hence, the gain in information from such a partitioning would be

$$Gain(genetic) = Info(D) - Info_{genetic}(D)$$
$$Gain(genetic) = .988 - .272 = .716$$

Similarly for alcohol and tobacco, the calculation is as follows.

$$Info_{Alcohol}(D) = \frac{8}{16} \times \left(-\frac{5}{8} \times \log_2\frac{5}{8} - \frac{3}{8} \times \log_2\frac{3}{8}\right) + \frac{8}{16}\left(-\frac{4}{8} \times \log_2\frac{4}{8} - \frac{4}{8} \times \log_2\frac{4}{8}\right)$$
$$Info_{Alcohol} = .984\ bits.$$

$$Gain(alcohol) = Info(D) - Info_{alcool}(D)$$

$$Gain(alcohol) = .988 - .984 = .004$$

$$Info_{tobacco}(D) = \frac{8}{16} \times \left(-\frac{5}{8} \times \log_2\frac{5}{8} - \frac{3}{8} \times \log_2\frac{3}{8}\right) + \frac{8}{16}\left(-\frac{4}{8} \times \log_2\frac{4}{8} - \frac{4}{8} \times \log_2\frac{4}{8}\right)$$

$$Info_{tobacco} = .984\ bits.$$

Hence, the gain in information from such a partitioning would be
$$Gain(tobacco) = Info(D) - Info_{tobacco}(D)$$

$$Gain(tobacco) = .988 - .984 = .004$$

$$Info_{diet}(D) = \frac{8}{16} \times \left(-\frac{5}{8} \times \log_2 \frac{5}{8} - \frac{3}{8} \times \log_2 \frac{3}{8}\right)$$
$$+ \frac{8}{16}\left(-\frac{4}{8} \times \log_2 \frac{4}{8} - \frac{4}{8} \times \log_2 \frac{4}{8}\right)$$

$$Info_{diet} = .984 \ bits.$$

Hence, the gain in information from a partitioning related to diet would be

$$Gain(diet) = Info(D) - Info_{diet}(D)$$

$$Gain(diet) = .988 - .984 = .004$$

Assumptions   (21-40 age)
1. Genetic factor is most important for occurrence of liver cancer
2. High alcohol and tobacco consumption also leads to liver cancer.
3. During this age liver metabolism becomes highest.

Table2. Effect on age group (21-40)

S. No.	Genetic	Alcohol	Tobacco	Diet	Chance of liver cancer
1	High	Low	Low	low	High
2	High	Low	Low	high	High
3	High	Low	High	low	High
4	High	High	Low	low	High
5	High	High	High	low	High
6	High	High	Low	high	High
7	High	Low	High	high	High
8	High	High	High	high	High
9	low	Low	Low	low	Low
10	low	Low	Low	high	Low
11	low	Low	High	low	Low
12	low	High	Low	low	Low
13	low	High	High	low	High
14	low	High	Low	high	High
15	low	Low	High	high	High
16	low	High	High	high	High

$$Info_{(21-40)}(D) = -\frac{12}{16}\log_2\frac{12}{16} - \frac{4}{16}\log_2\frac{4}{16}$$

$$Info_{(21-40)}(D) = .811$$

The expected information needed to classify a tuple in *D* if the tuples are partitioned according to Genetic is

$$Info_{genetic}(D) = \frac{8}{16} \times \left(-\frac{8}{8} \times \log_2 \frac{8}{8} - \frac{0}{8} \times \log_2 \frac{0}{8}\right)$$
$$+ \frac{8}{16}\left(-\frac{4}{8} \times \log_2 \frac{4}{8} - \frac{4}{8} \times \log_2 \frac{4}{8}\right)$$

$$Info_{genetic} = .5 \ bits$$

Hence, the gain in information from such a partitioning would be

$$Gain(genetic) = Info(D) - Info_{genetic}(D)$$

$$Gain(genetic) = .811 - .5 = .311$$

$$Info_{Alcohol}(D) = \frac{8}{16} \times \left(-\frac{7}{8} \times \log_2 \frac{7}{8} - \frac{1}{8} \times \log_2 \frac{1}{8}\right)$$
$$+ \frac{8}{16}\left(-\frac{5}{8} \times \log_2 \frac{5}{8} - \frac{3}{8} \times \log_2 \frac{3}{8}\right)$$

$$Info_{Alcohol} = .746 \ bits.$$

$$Gain(alcohol) = Info(D) - Info_{alcool}(D)$$

$$Gain(alcohol) = .811 - .746 = .065$$

$$Info_{tobacco}(D) = \frac{8}{16} \times \left(-\frac{7}{8} \times \log_2 \frac{7}{8} - \frac{1}{8} \times \log_2 \frac{1}{8}\right)$$
$$+ \frac{8}{16}\left(-\frac{5}{8} \times \log_2 \frac{5}{8} - \frac{3}{8} \times \log_2 \frac{3}{8}\right)$$

$$Info_{tobacco} = .746 \ bits.$$

$$Gain(tobacco) = .811 - .746 = .065$$

$$Info_{diet}(D) = \frac{8}{16} \times \left(-\frac{7}{8} \times \log_2 \frac{7}{8} - \frac{1}{8} \times \log_2 \frac{1}{8}\right)$$
$$+ \frac{8}{16}\left(-\frac{5}{8} \times \log_2 \frac{5}{8} - \frac{3}{8} \times \log_2 \frac{3}{8}\right)$$

$$Info_{diet} = .746 \ bits.$$

$$Gain(diet) = Info(D) - Info_{diet}(D)$$

$$Gain(diet) = .811 - .746 = .065$$

4. LIVER CANCER ANALYSIS IN THE LIGHT OF RELIABILITY ESTIMATION AND CONCEPT LEARNING THEORY.

4.1 Fact 1-The reliability and mean time to failure of liver cancer testing system can be realized in the light of parallel system configuration.

Justification –The mathematical computation has been carried out for two cases – (i) associated components of equal weightage (ii) associated components of different weight age

(i) Case- Associated components of equal weight age

Let assume an initial hypothesis that reliability (R_s) should be high and fix R_s =.99 For parallel system:

R_s=1-(1-$e^{-\lambda t}$)^3 …………………………………………..(13)

Where λ = failure rate, t = time.

In this context, the three associated components include - Blood Test, Imaging, and Biopsy as major classes of liver cancer test. If reliability is maximum, then from Eq.(13) we get :

$.99$=1-(1-$e^{-\lambda t}$)^3

or,   1-.99=(1-$e^{-\lambda t}$)^3

or, $\sqrt[3]{.01}$=1-$e^{-\lambda t}$ or ,-λt = ln .7846

Therefore,

-λt=-.2425

Now   if t=1 , then λ =.2425

As per the formula,

MTTF=$\frac{5}{6\lambda}$ =3.4364

(ii) Case- Associated components of different weight age

Let assume an initial hypothesis that reliability (R_i) should match with .70 as prescribed by the various literature sources.

Then, the failure rate of the blood test is represented as follows-

$R_i = e^{-\lambda T}$ …………………………………………………..(14)

or $.70 = e^{-\lambda_1 t}$

or, $\lambda_1 = .357$

The failure rate of imaging studies is represented as below :`

$$.979 = e^{-\lambda_2 t}$$
$$\ln .979 = -\lambda_2 t$$
$$\lambda_2 = .021$$

The failure rate of biopsy studies is represented as below :

$$.907 = e^{-\lambda_3 t}$$
$$\ln .907 = -\lambda_3 t$$
$$\lambda_3 = .098$$

Therefore,

$R_s(t) = 1 - ((1-e^{-\lambda_1 t}) \, (1-e^{-\lambda_2 t}) \, (1-e^{-\lambda_3 t}))$ …..........................(15)

$R_s(t) = e^{-\lambda_1 t} + e^{-\lambda_2 t} + e^{-\lambda_3 t} - e^{-(\lambda_1+\lambda_2)t} - e^{-(\lambda_1+\lambda_3)t} - e^{-(\lambda_2+\lambda_3)t} + e^{-(\lambda_1+\lambda_2+\lambda_3)t}$

$\lambda_1 = .357$,

$\lambda_2 = .021$,

$\lambda_3 = .098$ (assuming unit t time)

$R_s(t) = e^{-(.357)\times 1} + e^{-(.021)\times 1} + e^{-(.098)\times 1} - e^{-(.357+.021)\times 1} - e^{-(.357+.098)\times 1} - e^{-(.021+.098)\times 1}$
$\quad\quad\quad + e^{-(.357+.021+.098)\times 1}$

$R_s(t) = .699 + .979 + .907 - .685 - .634 - .888 + .621 = 0.999$

The above result indicates that the reliability of the diagnosis of liver cancer is highly acceptable.

Mean Time To failure (MTTF) $= \frac{1}{\lambda_1} + \frac{1}{\lambda_2} - \frac{1}{\lambda_1+\lambda_2} - \frac{1}{\lambda_1+\lambda_3} - \frac{1}{\lambda_3+\lambda_2} + \frac{1}{\lambda_1+\lambda_2+\lambda_3}$ ……………...(16)

Hence,

MTTF = 39.272

4.2 Fact 2 - The effect of alcohol consumption leading to liver cancer can be sensed using the concept learning approach.

Justification -Let P(x) be the probability of patient suffering from liver cancer,

y be the fact that person consumes more than three unit alcohol on regular basis,

P(y/x) be the probability of patient suffering from liver cancer and consuming more than three unit alcohol on regular basis,

P(~y/x) be the probability of patient suffering from liver cancer and consuming no alcohol,

P(~y/~x) be the probability of patient who neither suffers from liver cancer nor consumes any alcohol,

P(y/~x) be the probability of patient not suffering from liver cancer but consuming alcohol.

The exact posterior probabilities can be determined by normalizing the above notations as follows:

P(x/y) = (P(x) * P(y/x)) / ((P(y/~x)*P(~x) +

P(x) * P(y/x))……………………(17)

and

P(~x/y) = (P(y/~x)*P(~x)) / ((P(y/~x)*P(~x) +

P(x) *P(y/x))………......……….(18)

The aforesaid Bayesian inference justifies the statement "The effect of alcohol consumption leading to liver cancer can be sensed using the concept learning approach."

5. CONCLUSION
The paper entails how the liver cancer growth can be governed by artificial neural modeling, Markov property based state transition, geometric distribution, Cobb-Douglas model, the fundamental principle of information gain, reliability estimation and concept learning theory. The claims stated will definitely facilitate in medical diagnosis.

6. ACKONWLEDGEMENT

7. REFERENCES
[1] Bharathi, A., Natarajan, A. M. , "Efficient Classification of Cancer using Support Vector Machines and Modified Extreme Learning Machine based on Analysis of Variance Features", *American Journal of Applied Sciences*, 8(12), pp.1295-1301,2011.

[2] Feitelson MA, Sun B, SatirogluTufan NL, Liu J, Pan J, Lian Z., "Genetic mechanisms of hepatocarcinogenesis". *Oncogene*,21,pp.2593–2604,2002.

[3] Wong BW, Luk JM, Ng IO, Hu MY, Liu KD, Fan ST., "Identification of liver intestine cadherin in hepatocellular carcinoma--a potential disease marker" ,*BiochemBiophys Res Commun.*, 311(3),pp. 618-624,2003.

[4] Ms.Vincey Jeba Malar, V. , "Computer Aided Diagnosis for liver Cancer Feature Extraction", *The International Journal Of Engineering And. Science* ,2(11), pp. 27-30,2013.

[5] National Health Service, http://www.Cancer research uk.org.

[6] World Health Organization, Liver cancer report, http://www.cancer researchuk.org.

[7] Cancer Research UK, liver cancer causes, signs and symptoms, http://www.cancer researchuk.org.

[8] National Cancer Institute, http://www. cancer.gov /cancer topics.