

Overview of HCC; Data Mining Discovery (Multi-Center Study)

Mohamed Younis¹, Abdallah Nawara², Elsayed A. Elgohary², Mohammad M. Sallam², Ali Ismael², Abd Elrazek Abd Elrazek^{3*} and Hamdy Mahfouz^{1,*}

¹Department of Tropical, GIT and Hepatology, Al Azhar Asuit Faculty of Medicine, Al Azhar University, Asuit, Egypt

²Department of Internal Medicine, Zagazig Faculty of Medicine, Zagazig University, Zagazig, Egypt.

³Department of Tropical, GIT and Hepatology, Aswan Faculty of Medicine, Aswan University, Aswan, Egypt.

Received: 21 Dec. 2016, Revised: 2 Feb. 2017, Accepted: 12 Feb. 2017.

Published online: 1 Sep. 2017.

Background and Aim: Hepatocellular carcinoma is one of the most aggressive cancers that represent a global health problem. It is the fifth most common cancer and the second most common cause of cancer related- mortality worldwide. Hence we conducted the current study using high performance- data mining technology to obscure hidden knowledge would be helpful guide for better predication, follow up and best management.

Patients and Methods: Retrospectively, 130 Egyptian patients presented with HCC were followed clinically and their data were analysed using high performance technology of data mining intellectual machine learning of Raid I software program.

Results: Amazing results were obtained using Knowledge discovered by data (KDD) such as Serum Creatinine and Total Bilirubin should be the best laboratory indicators for HCC progression. AFP may increase not for the HCC itself but for progression cirrhosis. Furthermore HCC biology should be considered than multiplicity. **Conclusion:** HCC still a major medical concern in Egypt. Data mining programs would be very helpful in such an area of oncology and future medicine.

Abbreviations: AASLD: American Association for the Study of the Liver Diseases, AI: Artificial Intelligence, AFP: Alfa-feto- protein, CA: Computational analysis, DM: Data Mining, KDD: Knowledge discovered by data, ML: Machine Learning.

Conflict of interest: No

Keywords; HCC, Egypt, Data mining, AFP, Liver cirrhosis

1 Introduction

In Egypt, the recent epidemiological analysis showed that HCC is the commonest type of cancer causing much mortality and it is ranked as the first type of cancer in males representing 33.63% and the second in females representing 13.54% of all reported cancers [1-3]. However, our recent published studies suggested high prevalence of viral hepatitis in the Egyptian population, accordingly the peak of HCC should be heavily considered in the coming few years [4-7]. Machine learning (ML), data mining (DM), and the artificial intelligence (AI) are integrated terms widely used for big data analysis, recently used in applied medicine. Bioinformatics is a problem solved by machine learning or data mining usually applied in genetic sequencing. Currently data mining is the breakthrough in applied medicine would replace the traditional statistical analysis coming decades [8-11]. Here we will shed light on the usefulness of data mining analysis in HCC medicine.

2 Patients and Methods

This retrospective study was conducted on 130 recording sheets of patients with primary hepatocellular carcinoma (HCC), at Al-Azhar university hospitals, South Egypt cancer

institute and private hospitals across Egypt, in period from January 2015 to December 2016. Each patient presented with many parameters (Big data) in aim to discover best factors related –disease progression or regression using data mining technology. All patients were diagnosed by Ultrasound, Triphasic spiral CT scan, based on their early arterial enhancement and late portal venous washout according to AASLD guidelines.

2.1 Inclusion criteria

1-The Study was done on review sheets for Egyptian patients aged above 16 years old.

2-Patients with primary hepatocellular carcinoma.

3-All patients with HCC, diagnosed according to AASLD practice guidelines for such radiological data for complete description of tumor site, number, size and extension. This was done using abdominal ultrasound, triphasic spiral CT and Doppler studies for confirmation of possible vascular invasion and AFP.

All patients will be categorized according to each HCC therapy e.g.; surgery, local ablative therapy, TACE....etc

*Corresponding author e-mail: ahmadrazek@gmail.com \ dr.hamdymm511@gmail.com

Patients were classified into (A, B and C) according to Child Classification, additionally laboratory studies were obtained including; CBC, LFTs, RFTs, AFP, and viral markers.

MELD score was calculated for each patient.

Radiological data of tumor site, number, size and extension and if is associated with vascular invasion was also obtained. The study was performed according to Ethical and Medical rules of Al Azhar Faculty of Medicine-Asuit Committee.

2.2 Exclusion criteria

1-Patients with tumors metastasized to the liver; (Secondary liver cancers).

2- Patients with incomplete data.

3 Computational Analysis

3.1 Data mining Bio-informatics analysis

Why did we use Data mining in our study instead of different statistical programs?

Data mining is an intelligent computational analysis often using widely in economics and information technology to extract the helpful applicable and effective knowledge; (KDD). Data mining can be easily and simply applied in clinical situation to predict factors related-disease progression or regression. Nevertheless, statistical analyses are those methods deal with numbers and statistical parameters, hence sometimes results obtained may not be applicable clinically or being theoretical believes in many medical circumstances. On the other hand data mining can be viewed as a result of the natural evolution of information technology. In brief, data mining is the breakthrough in computational analysis over mathematical and statistical analyses in space, labor, biology, engineering and different computer sciences, recently in medicine with particular working on genomes (bio-informatics) and future medicine.

4 Results and Discussion

The major diagnostic techniques for HCC include serum tumor markers, various imaging modalities and histological analysis [12-16]. Tumor markers; Alpha fetoprotein (AFP) and Des- γ -Carboxy Prothrombin are used in early HCC diagnosis, assessment of stage, and prognosis of cancer. Also tumor markers are used to differentiate between benign and malignant tumors [17-19]. However the role of liver biopsy in diagnosis of HCC in patients with cirrhosis, has declined during the past few years because better radiological techniques have already improved the accuracy of diagnosis [20-22]. Many studies have been designed for the study of HCC epidemiology, biology, diagnostic modalities, different medical and surgical therapies, and even prediction epidemiology, however for our knowledge most of these studies have used traditional statistical analysis. Data mining

or Knowledge mining (KDD) may discover factors never been recognized by traditional statistical methods. Simple distribution Naïve Bayes and decision tree created by Rapid I version 4.6, Germany were used in the current study. For 130 patients included in the study, incidence of HCC in male gender higher than female [figure 1], additionally HCC not related cirrhosis is reported in male group.

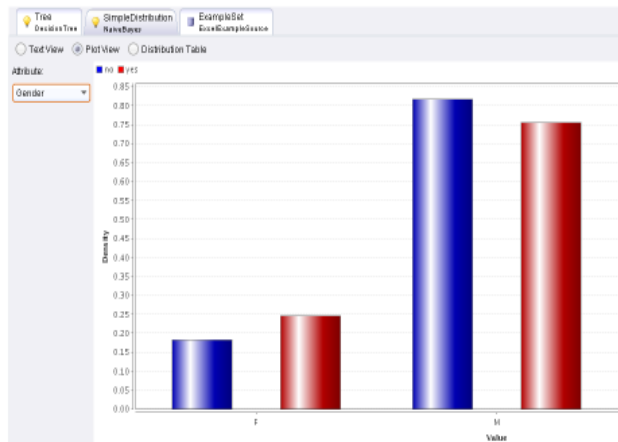


Figure 1: Incidence of HCC in Male and Female group. Blue= No cirrhosis, Red= cirrhosis.

Serum Creatinine and T. Bilirubin as a combination factor being the only predictor for advanced HCC especially in non-cirrhotic patients [Figure 2].

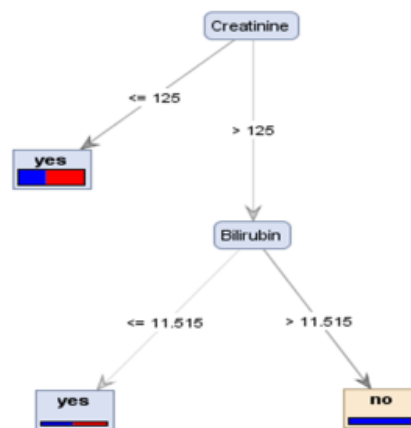


Figure 2: Decision Tree Algorithm; Higher (Creatinine & Bilirubin) means Progression of HCC especially in non-cirrhotic patients, irreversible to other Labs or Investigation Modalities; Biological Accuracy.

In simple words we can follow HCC patients using the decision tree created by Rapid I of data mining program, when increasing both Creatinine and Bilirubin that mean biological activity of the HCC with special concern for those without cirrhosis (More significant in non-cirrhotic HCC patients). Rapid I identified cut-off levels; (> 1.2 mg/dl) Equv.to 125mmol and (> 11.5 mg/dl) of S.Creatinine and T.Bilirubin respectively in HCC patients without underlying cirrhosis, should be progressive HCC biological activity.

Additionally AFP would be good predictor not only for HCC but also for the presence of underlying cirrhosis, hence AFP is rapidly progressive in those with HCC related -cirrhosis than without cirrhosis, likely cirrhosis initiate/ progress the AFP levels, hence apart of AFP is related to cirrhosis activity, [figure 3] and this explain why Japanese protocol for HCC screening depends on both AFP and PIVIKA II. Furthermore AFP proportionally related to advanced HCC biological activity than multiplicity [figure 4-6] and Table [1].

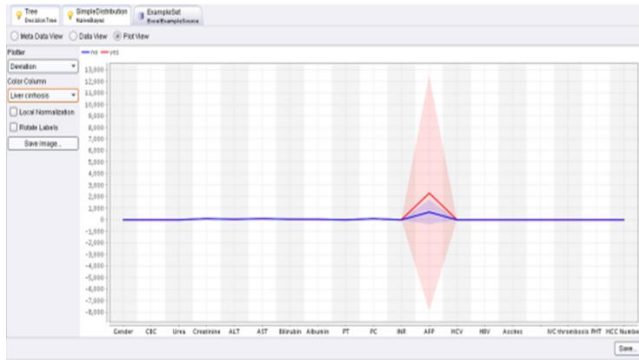


Figure 3: AFP increased significantly in HCC related-cirrhosis.



Figure 4: AFP is significantly higher in HCC related-biology more HCC related-multiplicity.

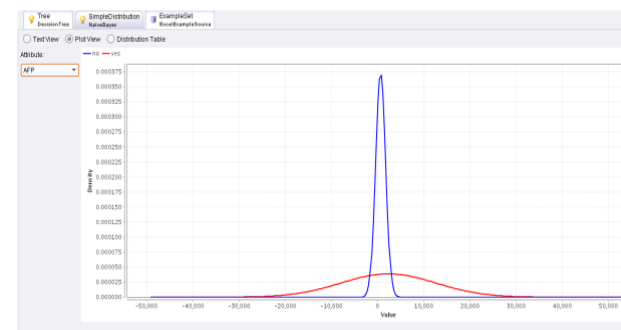


Figure 5: AFP is not only a good predictor for HCC but also for Liver Cirrhosis.



Figure 6: Blue = No Cirrhosis, Red = Cirrhosis, All are HCC patients.

Table 1: Showing the accuracy by Rapid I was 65.49% for most of the patients have been studied.

	true single	true multiple	true diffuse	class precision
pred. single	56	20	0	73.68%
pred. multiple	18	18	1	48.65%
pred. diffuse	0	0	0	0.00%
class recall	75.68%	47.37%	0.00%	

Acknowledgement

The current study was designed according to medical data mining computational technology, hence common statistical means; P value, CI and other statistical analytic parameters have been replaced accordingly.

References

- [1] 1-Abbasi A, Bhutto A, Butt N, Munir S M. Correlation of serum alpha-fetoprotein and tumor size in hepatocellular carcinoma. JPMA 62: 33; 2012.
- [2] 2-Abdel-Hamid M, El-Daly M, Molnegren V, et al. Genetic diversity in hepatitis C virus in Egypt and possible association with hepatocellular carcinoma. J Gen Virol 2007; 88(Pt 5): 1526–31.
- [3] 3-Abu El Makarem MA, Abdel Hamid M, Abdel Aleem A, Ali A, Shatat M, Sayed D, Deaf A, Hamdy L, Tony EA. Prevalence of Occult Hepatitis B Virus Infection in Hemodialysis Patients From Egypt With or Without Hepatitis C Virus Infection. Hepat Mon. 2012; 12(4):253-258.
- [4] Abu Dayyeh BK, Yang M, Fuchs BC, et al. A functional polymorphism in the epidermal growth factor gene is associated with risk for hepatocellular carcinoma. Gastroenterology 2011; 141:141.
- [5] 5-Andreana L, Isgrò G, Pleguezuelo M, Germani G,

- Burroughs AK. Surveillance and diagnosis of hepatocellular carcinoma in patients with cirrhosis. *World J Hepatol* 2009; 1(1): 48-61.
- [6] 6-Ansari D, Andersson R. Radiofrequency ablation or percutaneous ethanol injection for the treatment of liver tumors. *World J Gastroenterol.* 2012; 18(10): 1003-1008.
- [7] 7- Anwar WA, Khaled HM, Amra HA, El-Nezami H, Loffredo CA. Changing pattern of hepatocellular carcinoma (HCC) and its risk factors in Egypt: possibilities for prevention. *Mutat Res.* 2008; 659(1-2):176-84.
- [8] 8- Abd Elrazek AE, Bilasy SE, Elbanna AE, Elsherif AE. 2014 Prior to the Oral Therapy, What do we know about HCV-4 in Egypt: A Randomized Survey of prevalence and risks using data mining statistical analysis. *Medicine* 93(28):e204.
- [9] 9- Abd Elrazek AE, Eid KA, El-Sherif AE, Abd El Al UM, El-Sherbiny SM, Bilasy SE. 2015 Screening esophagus during routine ultrasound: medical and cost benefits. *Eur J Gastroenterol Hepatol* 27(1):8-12.
- [10] 10- Abd El Razek, Hamdy M Mahfouz, M. Afifi, AbdElwahhab fathy, Khaled Abdelazeem Ahmed El-Shamy, et al. 2014 Detection of Risky Esophageal Varices by Two-Dimensional Ultrasound: When to Perform Endoscopy. *Am J Med Sci* 347 (1):28.
- [11] 11- AbdElrazek M Aly AbdElrazek, Hamdy M. Mahfouz. 2013 Prediction analysis of esophageal variceal degrees using data mining: Is validated in Clinical medicine? *Global Journal of Computer Science and Technology* 13, 10:1-5.
- [12] 12-Ariff B, Lloyd C, Khan S, Shariff M, Thillainayagam A, Bansil D, Khan S, Taylor-Robinson S & Lim A. Imaging of liver cancer. *World Journal of Gastroenterology* 2009; 15 (11), 1289–1300.
- [13] 13-Arii S, Sata M, Sakamoto M, Shimada M et al. Management of hepatocellular carcinoma: Report of Consensus Meeting in the 45th Annual Meeting of the Japan Society of Hepatology (2009). *Hepatology Research* 2010; 40: 667–685.
- [14] 14-Aron R, Annunziato R, Miloh T, Wasserstein M, Sogawa H, Wilson M, Suchy F, Kerkar N. Liver transplantation for hereditary tyrosinemia type I: Analysis of the UNOS database. *Pediatr Transplantation* 2011; 15: 400–405.
- [15] 15-Arora A, Tyagi P, Ghuman S S, Sharma P et al. Hepatocellular Carcinoma Presenting as Budd–Chiari Syndrome. *Journal of Clinical and Experimental Hepatology* 2012; 2 (1) 91–92.
- [16] 16-Asim M, Sarma M P, Thayumanavan L, Kar P. Role of Aflatoxin B1 as a risk for primary liver cancer in north Indian population. *Clinical Biochemistry* 2011; 44:1235–1240.
- [17] 17-Das SK, Ray K. Wilson’s disease: an update. *Nat Clin Pract Neurol.* 2006; 2:482-493.
- [18] 18-Cucchetti A, Cescon M, Bigonzi E, Piscaglia F, Golfieri R, Ercolani G, et al. Priority of candidates with hepatocellular carcinoma awaiting liver transplantation can be reduced after successful bridge therapy. *Liver Transpl* 2011; 17:1344–1354.
- [19] 19-Cui R, He J, Zhang F, et al. Diagnostic value of protein induced by vitamin K absence (PIVKAII) and hepatoma-specific band of serum gamma-glutamyl transferase (GGTII) as hepatocellular carcinoma markers complementary to alpha-fetoprotein. *Br J Cancer* 2003;88(12):1878–82.
- [20] 20-Dancygier H. *Gross Anatomy*. In: Dancygier H (eds) *Clinical Hepatology Principles and Practice of Hepatobiliary Diseases*. Springer-Verlag Berlin Heidelberg, 2010. pp 11-1
- 21-Bassi N, Caratozzolo E, Bonariol L, Ruffolo C, Bridda A, Padoan L, Antoniutti M, Massani M. Management of ruptured hepatocellular carcinoma: Implications for therapy. *World J Gastroenterol* 2010; 16(10): 1221-1225.
- [21] 22-Behne T and Copur M S. Biomarkers for Hepatocellular Carcinoma. *Int J Hepatol*. Volume 2012, Article ID 859076, 7 pages.4. pp 1305-1350.