Journal of Health Informatics in Developing Countries

www.jhidc.org
Vol. 8 No. 2, 2014

Submitted: August 9, 2014 *Accepted: September 28, 2014*

Using Intelligent Data Analysis in Cancer Care: Benefits and Challenges

Niloofar MOHAMMADZADEH[a,1], Reza SAFDARI[b], and Farshid MOHAMMADZADEH[c]

[a] *PhD Health Information Management Tehran University of Medical Sciences, Tehran, Iran*
[b] *PhD, Associate Professor, Head of Health Information Management Department Tehran University of Medical Sciences, Tehran, Iran*
[c] *BSc, Bachelor of Electronics Engineer, Shahabe Danesh University, Iran*

Abstract. Access to accurate, comprehensive, and timely relevant cancer data for study the causes of cancer, detect cancer earlier, prevent or determine the effectiveness of treatment, specify the reasons for the treatment ineffectiveness, and cancer control programs is necessary. Physicians find it difficult to make accurate decisions when overwhelmed with hundreds of data. Intelligent data analysis (IDA) has benefits in different types of cancer for physicians, patient such as: cancer detection, classification of cancer, prediction of clinical outcome of patients after cancer surgery, prediction of survival in types of cancer and has advantages for managers, policymakers and researchers like benefits in epidemiology and assess healthcare resource utilization. Intelligent data analysis because of diversity of health data and specific circumstances of health domain confront enormous challenges. Lack of physician's trust on facts that generated by software, limits the number of disease-related data , lack of electronic guidelines and expert system that helps to physician in decision making, difficulties in mining methodology, user interaction, and efficiency and scalability areas are some of these challenges. IDA with potential benefits definitely have significant role in improve cancer care, prevention, treatment, increase speed and accuracy in diagnosis and treatment, reduce costs, clinical outcomes assessment, design and implementation of clinical guidelines. The aim of this review article is to survey application, opportunities and barriers of intelligent data analysis as an approach to improve cancer care management.

Keywords. Intelligent Data Analysis, Cancer Care, Cancer Management.

1. Introduction

Cancer as an important public health problem has social and economical burden in different countries.[1] More than 50 percent of new cancer cases occur in developing countries [2], so diagnosis, prevention as well as early detection of cancers leads to better health outcomes and resource usage [1].It is expected more different types of cancer will appear in the future because of population growth and aging and impose more difficulties to the societies. Hence cancer control programs and reducing cancer burden must be priority of countries [3]. Access to accurate, comprehensive, and timely relevant cancer data for study the causes of cancer, detect cancer earlier, prevent or

[1] Corresponding Author. Niloofar Mohammadzadeh. PhD Health Information Management Tehran University of Medical Sciences, Tehran, Iran. Email: nmohammadzadeh@razi.tums.ac.ir

determine the effectiveness of treatment, specify the reasons for the treatment ineffectiveness, and cancer control programs is necessary [4].

Information systems capture and store huge amounts of data elements in variety format like text, video, image. Data can be captured fully structured, semi-structured or in free text [5, 6]. Physicians find it difficult to make accurate decisions when overwhelmed with hundreds of data [7]. On the other hand developing countries faced to health care resource limited. Access to many of the most expensive diagnostic and therapeutic technology in these countries are not available. So find the information about which health care interventions have best result and improve the survival could be important [8].

New generation of computational theories and tools to assist people in extracting useful information and knowledge from the rapidly growing volumes of different type of cancer data, and apply decision support and intelligent systems is inevitable [9]. Intelligent data analysis (IDA) is data analysis with artificial intelligence methods. It is important tool that can provide access to hidden knowledge in large databases of clinical and administrative data in hospitals. Also improve decision support, prevention, treatment and diagnosis planning with provides useful knowledge for health care personnel. Use of intelligent systems lead to facilitate, accelerate and improve health care services, decision support in the areas of cancer diagnosis and treatment particularly in home care and telemedicine [10]. IDA has ability to prepare and present complex relations between symptoms, diseases, and medical and treatment consequences [11].

Intelligent data analysis with potential benefits definitely have significant role in improve cancer care, prevention, treatment, increase speed and accuracy in diagnosis and treatment, reduce costs, clinical outcomes assessment, design and implementation of clinical guidelines [12-14]. Generally, the aim of this review article is to survey application, opportunities and barriers of intelligent data analysis as an approach to improve cancer care management.

2. Intelligent Data Analysis Advantages in Cancer Care Management

Intelligent data analysis is very similar to data mining, but uses prior domain knowledge to conduct the analysis with interactive and iterative method. Apply prior knowledge increase the efficiency of the knowledge discovery process and avoid delivering trivial results to the physicians. Many of the same methods in traditional analysis, data mining, statistical classifiers such as support vector machines and hidden Markov models, probabilistic methods such as Bayesian classifiers , neural networks, genetic algorithms, tree-building and rule-discovery methods are used in intelligent data analysis [5].

Intelligent data analysis has benefits for physicians, patient, policy makers and researchers in different types of cancer. Some IDA Capabilities for physicians and patient include:

1- Cancer detection: various intelligent data analysis techniques aid to clinicians in the different cancer diagnosis [15-17]. Proteomic patterns in serum may be show pathological changes in an organ or tissue. Proteomics as a powerful approach help to identify new tumor makers. Data mining techniques are very helpful for uncover the differences in complex proteomic patterns [18]. Micro calcification detection in breast cancer diagnosis is another example of IDA advantage in cancer diagnosis. Complicated nature of surrounding of breast tissue, the variation of micro calcifications (MCs) in shape, orientation, brightness and size make difficult use of computer-aided systems for the accurate identification of MCs. knowledge-discovery mechanism and intelligent analysis are very useful in effective MC detection in digital mammograms and have high performance in terms of the success rate in MC detections measured by both true positive rate and false positive rate [19].

2- Classification of cancer: Symptoms of prostate cancer and benign prostatic hyperplasia (BPH) are very similar and their differentiation is very crucial. Intelligent data analysis methods can present a foresight for the diagnosis of the patients with prostate cancer or BPH [20].Detecting divergence between oncogenic tumors plays a pivotal role in cancer diagnosis and therapy. Data mining methods can use to design a computational strategy to predict the class of lung cancer tumors from the structural and physicochemical properties of protein sequences [21]. Mammogram breast X-ray imaging is an inexpensive and most effective method for early detection of breast cancer, but accurate diagnosis about benign and malignant breast lesion are exist as a challenge. Intelligent data analysis can be assisted to oncologist for detecting breast cancer and abnormalities faster than traditional screening program [22].

3- Prediction of clinical outcome of patients after cancer surgery: for example after breast cancer surgery, prediction of clinical outcome of patients plays an important role in medical tasks such as diagnosis and treatment planning. Artificial neural networks are powerful tool for analyzing cancer data and help to clinician for search through large datasets seeking subtle patterns in prognostic factors, and that may further assist the selection of appropriate treatments for the individual patient [23].

4- Prediction of Survival in types of cancer is another benefit of IDA methods [24-27].

5- Intelligent data analysis for cancer patient monitoring and home care: health care systems in different countries use telemedicine and home care in order to improve the efficiency of health resources, reduce costs and to promote wider access to healthcare services [28]. In recent years intelligent data analysis methods apply effectively in patient monitoring, remote monitoring, telemedicine and home care [29-31]. A physiology signal monitoring system can help medical staff to remote monitor and analyze physiology signal effectively in people with chronic disease like cancer. Reduce medical cost and avoid having to visit doctors in hospital are some benefits of this way. Adopts system on chip techniques to develop an embedded human pulse monitoring system with intelligent data analysis mechanism for disease detection and long-term health care is very useful. This system can monitor and analyze pulse signal in daily life, provides aids long-distance medical treatment, exploring trends of potential chronic diseases and data analysis scheme based on the modified cosine similarity measure to diagnose abnormal pulses for exploring potential chronic diseases

[32]. Also intelligent monitoring system is powerful tool in serious condition like intensive care units. These systems have ability of continuous learning of time series from the physiological variables that allows the monitor patient in real time [33].

General advantages of intelligent data analysis for managers, policymakers and researchers are including: 1- Epidemiology: advance analysis techniques lead to enhance the performance of cancer information retrieval, manage large volume data about cancer [34], help to testing hypothesis about cancer, describing the distribution of cancer [35], and epidemiological information discovery, prediction of epidemic dynamics and possible health care risks [31]. 2- Assess healthcare resource utilization: intelligent data analysis methods help to policy makers, managers and researchers to decision making about utilization, allocation and evaluation of health care resources in cancer domain through predict early hospital readmission [36], determine length of stay in hospital [37], predict disease risks of individuals based on their medical diagnosis history [38],predict accurate life expectancy prediction for clinical decision making [39] and predict unplanned hospital admission of patient [40].

3. Intelligent Data Analysis Advantages in Cancer Care Management

Despite the strong role of intelligent data analysis in improvement of cancer care management, the difficulties associated with using this advance technique remain. Intelligent data analysis because of diversity of health data and specific circumstances of health domain confront enormous challenges.The most important challenges in applying intelligent data analysis in cancer include: lack of physician's trust on facts that generated by software. Doctors more rely on evidence-based medicine [41], diversity of data, gathering data from different places, limits the number of disease-related data and lack of electronic guidelines and expert system that help to physician in decision making, imbalanced class distribution, take a long time to extract knowledge from special case reports [42], lack of efficient automatic preprocessing tools, lack of suitable tools for large rich and complex data sets, lack of truly integrated data analysis environment and lack of user friendly of post processing tools [43], optimization problem in so that data mining systems generate only interesting patterns, how to scale up a statistical method over a large data set for data mining, handling complex types of data in variety databases, and difficulties in mining methodology, user interaction, efficiency and scalability areas [44].

4. Conclusion

Cancer as a chronic disease is costly for healthcare systems in many countries. Proper cancer management and suitable control programs is very important especially in developing countries with limit resources and require accurate and comprehensive information. Large amount of different data types collect and store in diverse medical databases. Data analysis with artificial intelligent techniques is main tools that can provide access to hidden knowledge in large database of clinical and administrative data in hospitals. Intelligent data analysis methods are necessary for effective usage of

stored data and provide useful knowledge for health care personnel to improve decision support, increase quality of preventive; treatment and diagnosis planning. However intelligent data analysis faced with some challenges. Standardization of data collection, creating and using standard terminologies and databases consistent with these terminologies help to meet the challenges related to data collection from various places and application data in intelligent data analysis. Also Explain benefit of intelligent system to oncologist, emphasizing the helpful role of these types of systems, report generated results and feedbacks to them can decrease physician resistance and improve their attitudes and trust.

References

[1] Sankaranarayanan R, Ramadas K, Qiao Y. Managing the changing burden of cancer in Asia. *BMC Medicine* 2014; **12**:3. Doi: 10.1186/1741-7015-12-3.

[2] Shubair M. Cancer Control in Developing Countries: Need for Epidemiological Surveillance based on Health Information Systems and Health Services Research. *Journal of Health Informatics in Developing Countries* 2012; **16**(1): 356- 360.

[3] Lim Kim J, Cho K, Park E, Hyun Cho W. A Single Measure of Cancer Burden Combining Incidence with Mortality Rates for Worldwide Application. *Asian Pac J Cancer Prev* 2014; **15** (1): 433-439.

[4] Mohammadzadeh N, Safdari R, Rahimi A. Multi-Agent Systems: Effective Approach for Cancer Care Information Management. *Asian Pac J Cancer Prev* 2013; **14** (12): 7757-7759.

[5] Holmes JH, Peek K. Intelligent data analysis in biomedicine. *Journal of Biomedical Informatics* 2007; **40** : 605–608.

[6] Doods J, Botteri F, Dugas M, Fritz F, and on behalf of EHR4CR WP7. A European inventory of common electronic health record data elements for clinical trial feasibility. *Trials* 2014; **15**:18. doi:10.1186/1745-6215-15-18.

[7] Stacey M, McGregor C. Temporal abstraction in intelligent clinical data analysis: *A survey. Artificial Intelligence in Medicine* 2007; **39**: 1-24.

[8] Shulman N L., Willett W, Sievers A, Knaul M F. Breast Cancer in Developing Countries:Opportunities for Improved Survival . *Journal of Oncology* 2010; **2010** (2010) 1-6. doi:10.1155/2010/595167

[9] Fayyad U, Piatetsky-Shapiro G, Smyth P. From Data Mining to Knowledge Discovery in Databases. *AI Magazine* 1996; **17** (3):37-54.

[10] Ivan Lee S, Ghasemzadeh H, Mortazavi B, Lan M, Alshurafa N, Ong M, Sarrafzadeh M. Remote Patient Monitoring: What Impact Can Data Analytics Have on Cost?.*Wireless Health'13 Proceedings of the 4th Conference on Wireless Health*. 2013.DOI>10.1145/2534088.2534108

[11] Rusu M, Saplacan G, Sebestyen G, Cenan C, Krucz L, Nicoara T, Todor N. Distributed e-Health System with Smart Self-care Units. 2009. *IEEE 5th International Conference on Intelligent Computer Communication and Processing, 2009*. ICCP. available from : http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5284744&url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel5%2F5278388%2F5284721%2F05284744.pdf%3Farnumber%3D5284744 (access on 2 September 2012)

[12] Wint Khaing H. Data Mining based Fragmentation and Prediction of Medical Data. *IEEE.Computer Research and Development (ICCRD), 2011 3rd International Conference* on (Volume: 2). DOI: 10.1109/ICCRD.2011.5764179

[13] Kumar D.S, Sathyadevi G, Sivanesh S. Decision Support System for Medical Diagnosis Using Data Mining. *IJCSI International Journal of Computer Science* 2011; 8, **3**(1):147-153

[14] Coiera E W. AI in medicine: overview and challenges. 1996 *The institution of Electrical Engineers*. Printed and published by the IEE. Savoy Place, London WCPR OBL, UK. available from: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=578416&url=http%3A%2F%2 Fieeexplore.ieee.org%2Fiel3%2F4425%2F12561%2F00578416.pdf%3Farnumber%3D578416 (access on 31 august 2012).

[15] Kharya S. Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease. *International Journal of Computer Science, Engineering and Information Technology (IJCSEIT)* 2012; **2**(2):55-66. DOI : 10.5121/ijcseit.2012.2206.

[16] Saleema J S, Shenoy P D, Venugopal K R, Patnaik L M. Cancer Prognosis Prediction Model using Data Mining Techniques. *CiiT International Journal of Data Mining and Knowledge Engineering* 2014; **6**(1):21-29.
[17] Krishnaiah V, Narsimha G, Subhash Chandra N. Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques. *International Journal of Computer Science and Information Technologies* 2013; **4** (1): 39 – 45.
[18] Li L, Tang H, Wu Z, Gong J, Gruidl M, Zou J, Tockman M,Clark R A. Data mining techniques for cancer detection using serum proteomic profiling. *Artificial Intelligence in Medicine* 2004; **32**: 71-83.
[19] Peng Y, Yao B, Jiang J. Knowledge-discovery incorporated evolutionary search for micro calcification detection in breast cancer diagnosis. *Artificial Intelligence in Medicine* 2006; **37**: 43-53.
[20] Keles A, Hasiloglu A. S, Keles A, Aksoy Y. Neuro-fuzzy classification of prostate cancer using NEFCLASS-J . *Computers in Biology and Medicine* 2007; **37**: 1617 – 1628.
[21] Ramani R G, Gracia Jacob S. Improved Classification of Lung Cancer Tumors Based on Structural and Physicochemical Properties of Proteins Using Data Mining Models. *PLoS One* 2013; **8**(3): e58772. DOI: 10.1371/journal.pone.0058772
[22] Kumar Mohanty A, Ranjan Senapati M, Kumar Lenka S. An improved data mining technique for classification and detection of breast cancer from mammograms. *Neural Computing and Applications* 2013; **22**(1): 303-310.
[23] Jerez-Aragonés J, Gómez-Ruiz J A, Ramos-Jiménez G, Muñoz-Pérez J,Alba-Conejo E. A combined neural network and decision trees model for prognosis of breast cancer relapse. *Artificial Intelligence in Medicine* 2003; **27**:45-63.
[24] Jajroudi M, Baniasadi T, Kamkar L, Arbabi F, Sanei M, Ahmadzade M. Prediction of Survival in Thyroid Cancer Using Data Mining Technique Technol Cancer Res Treat., 2013 November 4th, Epub ahead of print DOI: 10.7785/tcrt.2012.500384
[25] Wan M, Fang L, Yu M, Cheng W, Wang P. Predicting the Survival Status of Cancer Patients with Traditional Chinese Medicine Symptom Variation Using Logistic Regression Model. *Advanced Data Mining and Applications Lecture Notes in Computer Science* Volume **8347**, 2013. pp 211-218. DOI: 10.1007/978-3-642-53917-6_19
[26] Chaurasia V, Pal S. Data Mining Techniques: To Predict and Resolve Breast Cancer Survivability. *International Journal of Computer Science and Mobile Computing* 2014; **3** (1): 10-22.
[27] Lakshmi K R, Kumar S P. Utilization of Data Mining Techniques for Prediction and Diagnosis of Major Life Threatening Diseases Survivability-Review. *International Journal of Scientific & Engineering Research* 2013; **4**(6):923-932.
[28] Costa C, Luís Oliveira J. Telecardiology through ubiquitous Internet services. *International journal of medical informatics* 2012; **81**:612-621.
[29] Huang J. Remote health monitoring adoption model based on artificial neural networks. *Expert Systems with Applications* 2010; **37**: 307-314.
[30] Hao J, Jayachandran M, Ni N, Phua J, Liew H, Wai Aung Aung P, Biswas J, Foo S, Low J, Yap L K P. An Intelligent Elderly Healthcare Monitoring System Using Fiber-Based Sensors. *Journal of the Chinese Institute of Engineers* 2010; **33**(5): 653-660.
[31] Ivančević V, Knežević M, Simić M, Luković I, Mandić D. Dr Warehouse - An Intelligent Software System for Epidemiological Monitoring, Prediction, and Research. *ThinkMind // DBKDA 2013, The Fifth International Conference on Advances in Databases, Knowledge, and Data Applications* // View article dbkda_2013_8_20_30101. p: 204 – 210.
[32] Chen C. Web-based remote human pulse monitoring system with intelligent data analysis for home health care. *Expert Systems with Applications* 2011; **38**: 2011–2019.
[33] Nouira K, Trabelsi A. Intelligent Monitoring System for Intensive Care Units. *J Med Syst* 2012; **36**:2309-2318.
[34] Li s. A multi-agent-based, semantic-driven system for decision support in epidemic management. *Health Informatics Journal* 2014; 21. DOI: 10.1177/1460458213517704
[35] Lewis F I, Ward M P. Improving epidemiologic data analyses through multivariate regression modeling. *Emerging Themes in Epidemiology* 2013; **10**:4. DOI:10.1186/1742-7622-10-4
[36] He D, Mathews SC, Kalloo AN, Hutfless S. Mining high-dimensional administrative claims data to predict early hospital readmissions. *J Am Med Inform Assoc* 2014; **21**(2):272-9. DOI: 10.1136/amiajnl-2013-002151.
[37] Zhang X, Zhang Z, Huang D. Prediction of Length of ICU Stay Using Data-mining Techniques: an Example of Old Critically Ill Postoperative Gastric Cancer Patients. *Asian Pacific J Cancer Prev* 2012; **13**: 97-101.

[38] Khalilia M, Chakraborty S, Popescu M. Predicting disease risks from highly imbalanced data using random forest. *BMC Medical Informatics and Decision Making* 2011; **11**:51. DOI:10.1186/1472-6947-11-51.

[39] Mathias JS1, Agrawal A, Feinglass J, Cooper AJ, Baker DW, Choudhary A. Development of a 5 year life expectancy index in older adults using predictive mining of electronic health record data. *J Am Med Inform Assoc* 2013; **20**(e1):e118-24. DOI: 10.1136/amiajnl-2012-001360.

[40] Rathi M, Chaussalet T. Risk prediction model using fuzzy regression method for predicting unplanned hospital admissions. *Systems Conference (SysCon), 2013 IEEE International*; P: 595 – 598. DOI: 10.1109/SysCon.2013.6549943.

[41] Mohammadzadeh N. designing an intelligent data analysis model in patients with chronic heart failure (CHF) based on multi-agent systems. *PhD thesis of health information management*. Tehran University of Medical Sciences.2014.

[42] Perner P. intelligent data analysis in medicine-Recent advances. *Artificial Intelligence in Medicine* 2006; **37**: 1-5.

[43] Nayak R. Intelligent Data Analysis: Issues and Challenges. July 14-18 2002. In: *Proc. of the 6th. World Multi Conferences on Systematics, Cybernetics and Informatics, Florida USA*. Available at: http://eprints.qut.edu.au/1479/1/ida-issues.pdf-

[44] Han J, Kamber M, Pei J. *Data Mining Concepts and Techniques*, third edition. Elsevier: USA. 2012. p : 22,24,32,34.