International Journal of Engineering & Technology, 7 (2.7) (2018) 729-732

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

A Literature Survey on Data Mining Approach to Effectively Handle Cancer Treatment

Sri Hari Nallamala[1]*, Siva Kumar Pathuri[2], Dr. Suvarna Vani Koneru[3]

[1]Research Scholar, CSE Dept.,
K L E F (Deemed to be University), Vaddeswaram, Guntur, A.P.
[2]Associate Professor, CSE Dept.,
K L E F (Deemed to be University), Vaddeswaram, Guntur, A.P.
[3]Professor, CSE Dept.,
V.R. Siddhartha Engineering College, Vijayawada, Krishna Dist., A.P.
**Corresponding author E-mail: nallamala.srihari@gmail.com*

Abstract

The effective treatment of cancer is not very easy since diagnosis of cancer involves many stages of treatment with gradually changing lifestyles. Physicians play vital role in identifying the correct cause and feel ambiguity for making perfect decisions about hundreds of data available from the internet resource. IDA (Intelligent Data Analysis) which is a part from Data Mining techniques is quiet useful to most of the physicians for decision making about types of cancers. IDA facilitates physicians to classify, detect and analyze the cancer outcome to patients. Healthcare Management System also aids the practitioners to practically search, analyze and compare the result analysis of the patient with existing data in the HMS and guide proper treatment to the cancer affected patient. Health care data analysis comprises enormous data with diversity of health information. One among the most important points that pull down the practitioner's confidence is that utility of latest software and most sophisticated computing machines. This put them in to the state of confusion for proper and elegant decision making for treating the cancer affected patients. Problems in user interaction, lack of awareness in data mining, improper knowledge in electronic guidelines makes physicians to work with old methods of treatment. Traditional medical practicing and modern methods of computing do not match either because of ignorance. IDA and HMS have significant impact for cancer treatment with speedy diagnosis and faster recovery. This also shows great impact on costs, clinical outcomes and proper guidelines for clinical approach. The prime motto of this survey article is to analyze the survey application, bring out the importance of comparison strategies of IDA to improve decision making for medical practitioner for effective cancer treatment.

Keywords: *Cancer treatment; Decision making; Healthcare Management Systems (HMS); Intelligent Data Analysis; Medical practitioners.*

1. Introduction

The nightmare of cancer is a very serious health hazard to most of the public in day to day life. It has a great impact on socio-economic factors in most parts of the geographical globe [1]. Developing nations are mostly prone to new cancers compared to developed and underdeveloped nations [2]. Therefore, early detection with further proper treatment should enrich affected people to enhance longer life span [1]. As per the current survey from the advanced research laboratories, it is expected the exponential growth of cancer problems are due to increase in the population, food habits, living style and improper utility of technology. Therefore, prioritization and Awareness on cancer prevention programs have been initiated by most of the nations. HMS and IDA are also part of these programs [3]. An elaborated, detailed and structured cancer data would be specifically useful for treatment of cancer with an effective manner [4].

2. Data mining in the Hospital Management System

Data Mining especially in the medical streams has a great advantage and plays most prominent issues to human being in this fast growing technology. It has its wings spread over for various Data Mining Techniques and KDD (Knowledge Discovery). Medical Data Mining competes with most of the challenges issues like large complex data structures which are heterogeneous and non hierarchical with unknown time series and invariable time quality. The already existing health care systems are subjected to collection of data chunks in invariable size, distributed and scattered in nature. These are probably integrated for decision making which has become a great challenging issue. It has been becoming critical aspect between the Data miner and the specialist for establishing bridge for these two persons. The common barrier prevailing for these two persons are legal, ethical and social aspects [5]. Some of the fever cancer diagnostic data mining techniques rarely used was cancer Prognostics and Diagnostics. Other techniques such as Confirmative and Explanatory are also frequently used in medical mining approach. Some of the serious challenges faced by

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

the data miners especially in the medical data mining are as follows:

Conglomeration of Data

Medical data scattered in the internet resource are heterogeneous. This heterogeneous information is collected from previous history of cancer affected patients, their images from scanned copies of their reports, notes from the physicians and their respective interpretations [6]. Such scattered heterogeneous information is considered to be the most important element for data mining and structuring this information is a vital tool for decision making for medical practitioners in the upcoming days.

Legitimate and social Dependencies

As Human body is most complicated structure in the universe, the information and data pertaining to this body after treating cancer affected patients is ware housed and safeguarded for future reference. In this regard, Legitimate, social dependencies and other ethical roles gear up to play a wide variety of roles. Safety measures and safe guarding metrics are to be calculated to prevent the information from misusing of data and abusing of patients. Medical information for humans is predominantly resourced for data ware housing. Data ownership needs to put great efforts to retrieve the available information for creating links among the datasets. Querying ownership of patient's data is an illegal act which encounters highly recurrent congressional enquiries and other law suits. There is another important and similar characteristic relation in data mining is safe guarding the security issues and protecting privacy factors. There shall be a break and misconception in patient's confidentiality and chance of taking necessary legal action. The patient strictly believes his physician about his medical issues and security about his personal records. Both ethically and legally practitioner cannot perform data analysis. The data Is to be protected under high privilege and needs to be posted over internet resource for some extend hiding the personal information. The information posted must hire some useful predictions.

3. Data Mining in Cancer Research

The works involved in the area of Data Mining are completely related to identify Cancer. "Mangasarian ez al" expressed his views regarding the concept and structure of linear programming and analyzed to the issue that has occurred due to clinical classifications of patients suffering from breast cancer. On the contrary, Bellazzi et al [7] revealed that the data mining tools and other techniques of mining can be adequately and productively used to extract Prognostic model and propagate the results pertaining to Reseczablehepazo cellular carcinoma. On the other hand, Land ez al [8] expressed about a new neural diagnosis of breast cancer with the help of mammogram verdict. Neural Networks paved path for both Walzer and Mohan [9] to classify the rules to project an algorithm that describe the diagnosis and causes of breast cancer. Caswel, in addition to that mentioned the state of art methodologies like preprocessing approach to mine the information using data mining techniques. Zupan et.al [10] made core study for prostate cancer and surprisingly disclosed that the machine learning classifiers can be magnificently used and implemented in overall methods of classification. Zhang and Zhang [11] developed and checked Prost Assure –They derived a comprehensive algorithm that detects the presence of prostate cancer from the concepts of neural networks. Their study with the help of algorithmic approach also helped to diagnose the serum tumor with single value index. They further stated that the complete process is carried out with the clusters of various patterns derived. These methods are currently in great demand for cancer detection. Most of the researchers stated that emerging pattern clustering uses data sets for cancer detection. Emerging patterns and projected clustering techniques are strong in different domains where they are used individually to solve different types of problems. Larry et al [12], one among the researchers has given a proposal stating integration of projected clustering techniques and emerging patterns to get effective mappings of DNA representations. The clusters resulted from DNA representations are used to distinguish and classify cancer related problems. With the help of micro array technology huge amount of data is categorized as collective information which can be further used for better analysis and treatment for cancer. Machine learning technique is useful for DNA Micro-Array methodology for identifying the cancer infection. It uses gene expression data, trains classifier classification. Sung–bae et al. [13] worked and achieved to represent the DNA expressions. DNA Micro Arrays are aptly implemented and used to completely derive the DNA expressions. The predictions which are part of the diagnosis to treat cancer involve two specific stages like classifiers of innumerous patterns and units of feature extractions. This enhances the easy and early prediction of cancer. Tumors are differentiated with the help of micro-array data using various artificial networks. Blair et al [14] developed a new alternative method called PAM (Plazforma) to analyze the problems in a huge variety. He used the technique of "nearest shrunken centroids" to analyze the problems. The patients referred in urology clinic undergo prostate specific antigen (PSA) concentration level using digital rectal exam. This is made possible with findings of data mining techniques. PSA velocity is used to measure the change in prostate specific antigen in a single patient over elapsed time.

Bagirov et al. [15] supported feature treatment decision making by illustrating the utilization of optimization based clustering techniques through clustering the prostate cancer patients. The clusters are produced by PSA (Prostate Specific Antigen) levels in the blood. These clusters had an impact on the patient's health and their lifespan in future. The data mining techniques which are used to build the classification models include discriminate analysis and logistic regression [16]. The linear discriminate analysis (LDA) is mostly criticized because of its belief on categorical behavior and misinterpretations of covariance of metrics at different levels of diagnosis. The Neural network analysis submitted its report stating that the methodology is much better compared to LDA and logistic regression on the classification capability. The limitations high lightened in this approach are prolonged staffing with enhanced training and further earmarked as mandate. It is also very typical to recognize the comparative need of potential input variables. Therefore, it is restricted to useless in various applications in handling the classification problems [17].

Delen[18] along with team members surveyed on the breast cancer patients and recognized around 202,932 records and further divided them in to two categories namely "survived (93,273"and "not survived (109,659). The outcome of this survey gave them 93% accuracy in survived category.

Tan Ac's utitla [19] with the assistance of c4.5 decision tree, bagged decision tree utilized and used on seven most commonly found in public cancers data array and compared with the predicted methods used before. By using c5 algorithm with bagging Liuqin's [20] experimental on breast cancer data available and predicted the surveillance of breast cancer people. Similarly, using c4.5 with and without bagging Jinyan Li Huiqing Liu's [21] conducted various experiments on the ovarian tumor data to recognize the cancer. Dong-Sheng cao's [22] stated that decision trees depending on ensemble method which is a combination of feature selection method backward elimination strategy with bagging in order to identify the relationship of structure activity which is concerned in area of chemo metrics associated to pharmaceutical industry. My ChauTu's [23] put forward a suggestion to the use of bagging with c4.5 algorithm.

Using Naïve bayes algorithm, identification of the symptoms of the heart disease of a patientis is typical. MY ChauTu's [24] with the help of bagging algorithm [25] recognized the symptoms of

heart disease in patients and estimated the contrast outcomes of decision tree induction method. Classifiers neural network SVM's and decision trees Tsirogiannis used bagging algorithm on the medical data bases. The outcome of the results showed improved accuracy of bagging compared without bagging. Pan wen[26] performed several experiments on ECG data to recognize abnormal high frequency electrocardiograph by the help of decision tree algorithm c4.5 with bagging karcochinporn c's [27] suggested a new classification algorithm TBWC which includes both decision tree with bagging and clustering. This algorithm is used on an experimental basis on two medical data sets namely cardiocography1 and cardiocography2 while other data sets are not concerned with medical domain. Chaurasia and pal [28] examined study on estimation of heart attack risk levels from the database of heart disease by using data mining techniques like Naïve Bayes, Ju8 decision tree and bagging approaches[29].

The final result revealed that the bagging techniques are more efficient compared to Bayesian classification and J48.

According to the medical research body at NewDelhi, India consists of around 17.3 lakh new cases of cancer and around 8.8 lakh cases of cancer affected patients with cervix, lung and breast may be victims of death by 2020. It is very clear indication that Indian ranks top in the list.

The Indian council of Medical Research (ICMR) revealed and forecasted that around 14.5 lakh cases of cancer infected patients may exist by 2016 and this gradually increase to 7.3 lakh cases by 2020. According to the data, only 12.5 % of the patients will arrive at early states of the disease. They proposed that Breast cancer is ranked first and most commonly found in females and mouth cancer in males are most frequently seen. The north east stated that the report of highest number of cancer cases in both males and females are around Aizwal district in Mizoram district. The pampumpare district in Arunchal Pradesh recorded the highest number of females are affected with cancer.

The lung cancer is one with a prediction of around 1.14 lakh (83000 in males and 31000 in females) are recorded during the year 2016. Cervix cancer is the 3rd most ranking cancer and predicted over one lakh new cases during 2016. Most of the cancers (30% in males and females) are caused due to the consumption of tobacco- which was stated by ICMR. The ICMR also performed various survey activities and conducted a country wide survey from 2012-2014 and resulted population Base cancer Registeries (PBCR) and Hospital Based Registeries. It also stated that there is a significant increase in the cancers pertaining to Rectum and colon among males in the PBCR's at Bangalore, Chennai and Delhi. There is also a predominant increase in the cancer cases of colon, rectum and proctate found among women with a magnificent rate. Nanda kumar, the head of the National cancer Registry stated that the cancer among children was also identified in most parts of Delhi. In India "one in eight Indians" show the sign of cancer and can be observed in their lifetime.

On a survey, it is very clearly understood that the problems pertaining to the cancer treatment can be effectively answer with the approach of data mining techniques and present relevant solutions to the medical practitioners.

Acknowledgement

This research paper was supported by Dept. of CSE, K L E F (Deemed to be University). I thankful to my Internal Supervisor Sri. Siva Kumar Pathuri, Associate Professor, Department of CSE of K L E F (Deemed to be University) and External Supervisor Dr. Suvarna Vani Koneru, Professor, Department of CSE of V.R. Siddhartha Engineering College who provided their insight and expertise that greatly assisted the research and also their comments that greatly improved the manuscript.

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 Prev2014; 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 Prev2013; 14 (12): 7757-7759.
[5] Cios, K.J. and Moore, G.W., "Uniqueness of medical data mining" Artificial intelligence in medicine 2002; 26: 1-24.
[6] Mangasarian O.L., "Breast cancer diagnosis and prognosis via linear programming", Operations Research, 1995, 43(4): 570-577.
[7] Bellazzi, R., Azzini, I., Toffolo, G., Bacchetti, S. and Lise, M., "Mining data from a knowledge management perspective: an application to outcome prediction in patients with respectable hepatocellular carcinoma", In Proceedings of AIME, 2001, 40-49
[8] Land W.H., "New results in breast cancer classification obtained from an evolutionary computation/adaptive boosting hybrid using mammogram and history data" Proceedings of 2001 IEEE Mountain Workshop on Soft Computing in Industrial Applications. IEEE, 2001, 47-52.
[9] Walter D., Mohan C.K., "ClaDia: A fuzzy classifier system for disease diagnosis" In Proceedings of the 2000 Congress on Evolutionary Computation, IEEE Part Vol. 2, 2000, 1429-1435.
[10] Zupan B., Demsar J., "Machine learning for survival analysis: a case study on recurrence of prostate cancer" Artificial intelligence in medicine, 2000, 20(1): 59-75.
[11] Zhang Z. and Zhang H., "Development of a neural network derived index for early detection of prostate cancer" In Proceedings of IJCNN'99 – International Joint Conference on Neural Networks, IEEE, Vol. 5 1999, 3636-3641.
[12] Larry T.H., "Using Emerging Pattern Based Projected Clustering and Gene expression data for Cancer detection" 2nd Asia-Pacific Bioinformatics Conference, Vol. 29, 2004.
[13] Sung-Bae C., "Machine Learning in DNA Microarray Analysis for Cancer Classification" 1st Asia-Pacific Bioinformatics Conference, Adelaide, Australia, 19, 2003.
[14] Blair E., Tibshirani, R., "Machine Learning Methods Applied to DNA Microarray Data can Improve the Diagnosis of Cancer", SIGKDD Explorations, 2003, 5(2): 48-55.
[15] Churilo L., Bagirov A.M., K. Schwartz, Smith K., Dally M., "Improving Risk Grouping Rules for Prostate Cancer Patients with Optimization", Proceedings of the 37th Hawaii International Conference on System Sciences, 2004.
[16] Tigrani V., John G., "Data mining and statistics in medicine: an application in prostate cancer detection", Source: www.robotics.stanford.edu, Accessed: December 23, 2004.
[17] Shieu-Ming C. et al, Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines, Expert Systems with Applications, 27 (2004) 133-142.
[18] Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. Artificial Intelligence in Medicine. 2005 Jun; 34(2):113-27.
[19] Tan AC, Gilbert D. "Ensemble machine learning on gene expression data for cancer classification", Appl Bioinformatics. 2003;2(3 Suppl):S75-83.
[20] Liu Ya-Qin, Wang Cheng, Zhang Lu," Decision Tree Based Predictive Models for Breast
[21] JinyanLiHuiqing Liu, See-Kiong Ng and Limsoon Wong*," Discovery of significant rules for classifying cancer diagnosis data",* Bioinformatics 19(Suppl. 2)Oxford University Press 2003.
[22] Dong-Sheng Cao, Qing-Song Xu ,Yi-Zeng Liang, Xian Chen, "Automatic feature subset selection for decision tree-based ensemble methods in the prediction of bioactivity", Chemometrics and Intelligent Laboratory Systems.

[23] My ChauTu, Dongil Shin, Dongkyoo Shin , *"Effective Diagnosis of Heart Disease through Bagging Approach",* 2nd International Conference on Biomedical Engineering and Informatics,2009.

[24] My ChauTu, Dongil Shin, Dongkyoo Shin, "*A Comparative Study of Medical Data Classification Methods Based on Decision Tree and Bagging Algorithms***"** Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, 2009.

[25] Tsirogiannis, G.L, Frossyniotis, D, Stoitsis, J, Golemati, S, Stafylopatis, A Nikita,K.S,"*Classification of Medical Data with a Robust Multi-Level Combination scheme*", IEEE international joint Conference on Neural Networks.

[26] Pan Wen, *"Application of decision tree to identify a abnormal high frequency electrocardiograph"*, China National Knowledge Infrastructure Journal, 2000.

[27] Kaewchinporn .C, Vongsuchoto. N, Srisawat. A *" A Combination of Decision Tree Learning and Clustering for Data Classification"*, 2011 Eighth International Joint Conference on Computer Science and Software Engineering (JCSSE).

[28] V. Chauraisa and S. Pal, "Data Mining Approach to Detect Heart Diseases", International Journal of Advanced Computer Science and Information Technology (IJACSIT),Vol. 2, No. 4,2013, pp 56-66.

[29] RAHMAN, M.Z.U., KARTHIK, G.V.S., FATHIMA, S.Y. and LAY-EKUAKILLE, A., 2013. An efficient cardiac signal enhancement using time-frequency realization of leaky adaptive noise cancelers for remote health monitoring systems. Measurement: Journal of the International Measurement Confederation, 46(10), pp. 3815-3835.