IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 2, Ver. I (Mar.-Apr. 2017), PP 33-38
www.iosrjournals.org

Probability of Predicting Cancer in Patient by Analyzing Risk Factors

Pallavi Mirajkar[1], Dr. G. Prasanna Lakshmi[2]
[1]*Research Scholar,Faculty of Science,Pacific University, Udaipur*
[2]*Guide (WOS-A) Andhra University*

Abstract: *Recently analytical hierarchy process is applied in different fields like selection of business plan, banking industries, technology selection, public transport, and decision in health care. One of the applications of analytical hierarchy process is medical diagnosis which is mostly used in research area. Many researchers are focusing on medical field. This study introduces an approach to find out risk level of the cancer based on analytical hierarchy process (AHP). Cancer is the most frequent cause of death. Health of the patients are affected by the cancer risk factors such as biological, genetically, environmental and lifestyle. Analytical hierarchy process approach incorporate four risk factors that are considered to identify cancer risk level. This approach helps to make decision about risk level of cancer.*

Keywords: *Cancer, Multi Criteria Decision Making, Analytical Hierarchy Process, pair comparison*

I. Introduction

Cancer refers to any one of a large number of diseases characterized by the development of abnormal cells that divide uncontrollably and have the ability to infiltrate and destroy normal body tissue. Cancer often has the ability to spread throughout your body. Research shows that certain risk factors increase the chance that a person will develop cancer. A risk factor is anything that increases a person's chance of developing cancer. These are the most common risk factors for cancer like age, tobacco, sunlight, smoking, alcohol and so on. Cancer treatment may include chemotherapy, radiation, and/or surgery. Therefore if earlier cancer is diagnosed and treated, better the chance of being cured.

Nowadays multi criteria decision making techniques have been developed and their usage are increasing tremendously to achieve desired goal. Multiple criteria decision making has a more standardizing approach based on a different paradigm that assumes the existence of something that allows the decision maker to find the best choices. This is done through a metric or using mechanisms based on the comparison of alternatives. The aim of these approaches is to observe the behavior of decision makers, help them to understand the decision problem, take account of all factors that influence the decision and prescribe the set of preferred solutions.

The Analytical Hierarchy Process (AHP) was developed by Thomas L. Satty. The analytical hierarchy process is a decision support tool which can be used to solve complex decision problems. It uses a multi-level hierarchical structure of objectives, criteria, sub criteria, and alternatives. Analytical hierarchy process helps decision makers to find one that best suits their goal and their understanding of the problem. It provides a comprehensive and rational framework for structuring a decision problem, for representing and quantifying its elements, for relating those elements to overall goals, and for evaluating alternative solution.

II. Related Work

T. Sowmiya, M. Gopi, M. New Begin, L.Thomas Robinson (2014) used several aspects of data mining procedures which are used for lung cancer prediction for the patients and also reviewed the aspects of ant colony optimization (ACO) technique in data mining. Ant colony optimization assists in increasing or decreasing prediction value of the diseases. This case study assorted data mining and ant colony optimization techniques for appropriate rule generation and classifications on diseases, which pilot to exact Lung cancer classifications. P.Ramachandran N.Girija, T. Bhuvaneswari (2014) approached k-means clustering algorithm for partitioning the data into cancer and non cancer clusters, where the initial cluster centers is represented by the mean value of the weightage of significant patterns. The objective of the clustering is that the data object is assigned to unknown classes that has a unique feature and hence maximize the intraclass similarity and minimize the interclass similarity and applied Decision tree algorithm to mine frequent patterns from the data set and patterns that are mined by the decision tree are well defined and distinguished to be separated as cancer and non cancer. The aim of cancer recurrence prediction is to predict, given a set of gene expression data, whether or not a particular cancer will recur within a particular time frame.

To summarize the gap between above studies K. Arutchelvan, Dr. R. Periyasamy (2015) introduced an algorithm using data mining technique and decision tree ,the user has to enter into the cancer prediction system,

they need to answer the queries, related to genetic and non genetic factors. The prediction system assigns the risk value to each question based on the user responses. Once the risk value is predicted, the range of the risk can be determined by the prediction system. It has four levels of risk like low level, intermediate level, high level and very high level. Based on the predicted risk values the range of risk will be assigned.

Shoon Lei Win, Zaw Zaw Htike, Faridah Yusof, Ibrahim A. Noorbatcha (2014) proposed a three-layered framework that consists of entropy-based gene selection, entropy minimization discretization and prediction. In the case of genetic data classification, not all the genes in a genetic sequence might be responsible for predicting cancer recurrence. They propose to employ a gene selection process to select relevant prognostic genes in an unsupervised manner and an entropy-based discretization process to discretize the gene expression levels. V.Krishnaiah et al (2013) applied data mining classification techniques to develop a model to predict patients with Lung cancer disease appears to be Naive Bayes followed by IF-THEN rule, Decision Trees and Neural Network. For Diagnosis of Lung Cancer Disease Naïve Bayes observes better results and fared better than Decision Trees.

Dash B, Mishra D, Rath A, Acharya M (2010) applied a hybridized K-means algorithm which combines the steps of dimensionality reduction through PCA, Using the proposed algorithm a given data set was partitioned in to k clusters. The experimental results show that the proposed algorithm provides better efficiency and accuracy comparison to original k-means algorithm with reduced time. Logically data mining algorithms CART (Classification and Regression Tree), ID3 (Iterative Dichotomized 3) and decision table (DT) Hnin Wint Khaing(2011). These classification algorithms are selected because they are very often used for research purposes and have potential to yield good results. Moreover, they use different approaches for generating the classification models, which increases the chances for finding a prediction model with high classification accuracy.

Williams, Kehinde et al (2015) introduced Naive Bayes Classifier is a probabilistic model based on Baye's theorem. It is one of the frequently used methods for supervised learning. It provides an efficient way of handling any number of attributes or classes which is purely based on probabilistic theory. Bayesian classification provides practical learning algorithms and prior knowledge on observed data. They presented that J48 decision trees classifier is a simple decision learning algorithm, J48 decision trees is a better model for the prediction of breast cancer risks for the values of accuracy, recall, precision and error rates recorded for both models.

III. Multi-Criteria Decision Making

Mark Velasquez1 and Patrick T. Hester (2013) compared different MCDM methods such as MAVT, MAUT, ELECTRE and analytical hierarchy process (AHP). Based on the literature reviewed, the observed advantages and disadvantages, as well as areas of application for each method, are summarized This research could lead to a survey of users to assess which advantages and disadvantages are more prevalent for each method.

Recently Praveen Thokala, Nancy Devlin, Kevin Marsh, Rob Baltussen, et al (2016) analyzed multi-criteria decision analysis (MCDA) is intended to serve as a tool to help decision makers reach a decision—their decision, not the tool's decision. The decision makers can deliberate on which is the most appropriate evidence and thus, the most appropriate score and the most appropriate "total value" before making their final decision.

Multi Criteria Decision Analysis (MCDA) is a well accepted framework that can simultaneously assess multiple criteria for priority setting of interventions Baltussen R, Niessen L(2006). Different approaches of multi-criteria decision analysis (MCDA) are proposed but contain at least the following elements: 1) selection of relevant interventions, 2) selection of criteria for priority setting, 3) collecting evidence and rating the performance of interventions on selected criteria, 4) deliberation on the evidence and performance of interventions with the aim to select the best interventions for implementation Dolan JG(2010).

Kristie Venhorst, Sten G Zelle1, Noor Tromp and Jeremy A Lauer (2014) proposed Delphi study, questionnaires were used to discuss a final list of criteria with clear definitions and potential scoring scales. For this Delphi study, multiple breast cancer policy and priority-setting experts from different low and middle income countries (LMICs) were selected and invited by the World Health Organization. Fifteen international experts were participated in all three Delphi rounds to assess and evaluate each criterion. This study resulted in a preliminary rating tool for assessing breast cancer interventions in low and middle income countries. The tool consists of 10 carefully crafted criteria such as effectiveness, quality of the evidence, magnitude of individual health impact, acceptability, cost-effectiveness, affordability, safety, geographical coverage, and accessibility, with clear definitions and potential scoring scales.

Claudio Diaz-Ledezma, Paul M. Lichstein, James G. Dolan, and Javad Parvizi (2014) conducted the multi-criteria decision analysis(MCDA) through the analytic hierarchy process, comparing the diagnostic strategies in terms of benefits, opportunities, costs, and risks. The analytical hierarchy process (AHP) method of

multi-criteria decision analysis (MCDA) for medical decision support which allows the decision maker to design a hierarchical structure and evaluate the trade-offs between decision criteria and alternatives.
C. Chandrasekar, P.S. Meena(2012) introduced Extreme Learning Algorithm (ELM) algorithm for classification in which the extreme learning algorithm (ELM) is trained using Levenberg Marquardt algorithm for training. Analytic Network Process (ANP) is used to improve the performance of ELM. This approach also makes use of ANOVA statistical ranking approach. Nguyen T, Khosravi A, Creighton D, Nahavandi S (2015) introduced a novel approach to gene selection based on a substantial modification of analytic hierarchy process (AHP). They stated that the modified analytic hierarchy process (AHP) systematically integrates outcomes of individual filter methods to select the most informative genes for microarray classification. This paper also proposed fuzzy standard additive model (FSAM) for cancer classification.

IV. Proposed Work

Introduction

Here, analytical hierarchy process method of multi-criteria decision making (MCDM) is used to develop a rational and objective finding. The analytical hierarchy process first decomposes the decision problem into a hierarchy of sub problems. Then the decision-maker evaluates the relative importance of its various elements by pair wise comparisons. The analytical hierarchy process converts these evaluations to numerical values (weights or priorities), which are used to calculate a score for each alternative (Saaty, 1970). The advantage of the analytical hierarchy process is simplicity and consistency as compared to other decision support methods.

A consistency index measures the extent to which the decision-maker has been consistent in response. Considering all relevant criteria along with their importance are discussed. Suppose that the cancer risk factors are identified and the decision-makers are responsible for identifying the risk of cancer of patients under each of the N criteria.
Cancer risk factors are categorized as following:

Table1: Cancer risk factors

Category	**Attributes**
Environmental	• Chemical Exposure • Sunlight Exposure • Radiation
Lifestyle	• Smoking • Tobacco • Alcohol • Diet and exercise
Biological	• Age • Infectious agent • Weight loss • Obesity
Genetically	• Family History • Inheritance • Hormones

Prof. Saaty has proposed Random Consistency Index ( ) as given below:

n	**RI**
1	0
2	0
3	0.58
4	0.9
5	1.12
6	1.24
7	1.32
8	1.41
9	1.45
10	1.49

Table2: Random consistency index

Preference scales of paired comparisons:

Numerical Judgment	Judgment of preference
1	Equally Preferred
2	Equally to moderately preferred
3	Moderately preferred
4	Moderately to strongly preferred
5	Strongly preferred
6	Strongly to very Strongly preferred
7	Very strongly preferred
8	Very strongly to extremely preferred
9	Extremely preferred

Table3: Preference Scale

The structure of hierarchy can be drawn as the following:

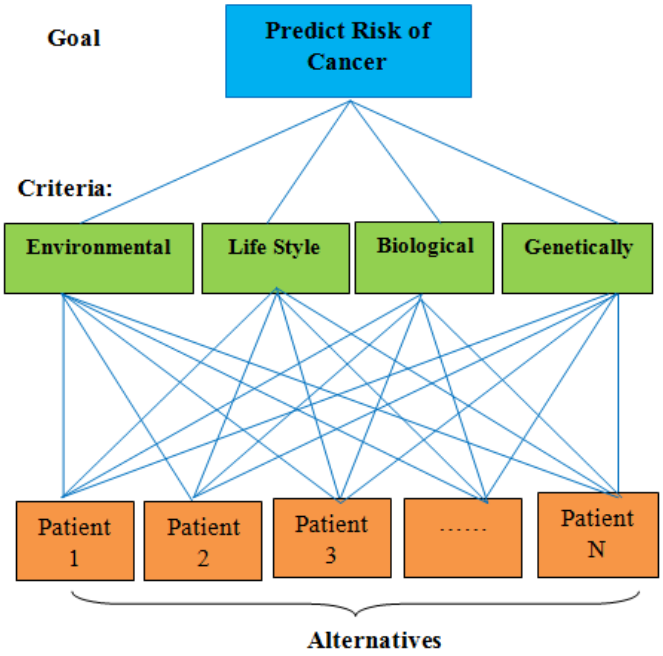

Fig. 1 Hierarchy Diagram

The steps of proposed method are depicted as given below:

1. Define the Criteria =Risk Factors {Environmental, Lifestyle, Biological, Genetically} and Alternatives = {Patient1, Patient2, Patient3… Patient N}

2. Compute comparison matrix for level1 with respect to the goal. Make the Upper triangular**,** Diagonal= 1 and Lower triangular = $A_{ij} = 1/A_{ij}$

Criteria	Risk Factor				Priority vector
	F1	F2	F3	F4	
Risk Factor 1	1				
Risk Factor 2		1			
Risk Factor 3			1		
Risk Factor 4				1	
Sum of Col					

Table 4: Comparison matrix with respect to the goal.

3. Divide each element of the matrix with the sum of its column, is called as priority vector.

4. Calculate Eigenvector of matrix (λ_{max}) by summation of product between each element of priority vector and the sum of columns of reciprocal, consistency index (CR) and consistency ratio (CR)

$$CI = \frac{\lambda \max - n}{n-1} \quad \text{and} \quad CR = \frac{CI}{RI}$$

5. Repeat step 2 and 3 with respect to the each risk factor as declared in step 1

Criteria No.1 Risk Factor1	Patients				Priority vector
	P1	P2	P3	PN	
P1	1				
P2		1			
P3			1		
P N				1	
Sum of Col					

Table 5: Comparison Matrix with respect to the each risk factor

6. Calculate adjusted weight for each risk factor.
7. Compute the overall composite weight of each patient from patient1 to patient N
Each alternative = (weight * priority vector1) + (weight * priority vector2) + … + (weight * priority vector n)
8. Arrange the composite weight in ascending order. Patient with highest value has HIGH RISK of cancer.

V. Conclusion

In the proposed paper, explored work can help in the medical field to identify risk level of the cancer such as Low Risk, Medium Risk or High Risk and can also help in decision making process at early diagnosis. The main objective of the study is to provide earlier intimation about the risk of cancer to the patient that can save the time and cost of the treatment.

In future, we planned to work on the technique that analyzes the food that causes cancer.

References

[1]. https://www.cancer.gov
[2]. K.Arutchelvan, Dr.R.Periyasamy "Cancer Prediction System using Data mining Technique" International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 02 Issue: 08 | Nov-2015 www.irjet.net p-ISSN: 2395-0072
[3]. T. Sowmiya, M. Gopi, M. New Begin L.Thomas Robinson "Optimization of Lung Cancer using Modern data mining techniques." International Journal of Engineering Research ISSN:2319- 6890)(online),2347-5013(print)VolumeNo.3,Issue No.5, pp : 309-3149(2014)
[4]. V.Krishnaiah "Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques" International Journal of Computer Science and Information Technologies, Vol. 4 (1) 2013, 39 – 45 www.ijcsit.Com ISSN: 0975-9646
[5]. Cancer Prevention and Control Retrieved from http://www.cdc.gov/cancer/dcpc/resources/features/worldcancerday/ Retrieved on: 15 November 2013.
[6]. Shoon Lei Win, Zaw Zaw Htike, Faridah Yusof, Ibrahim A. Noorbatcha "Cancer Recurrence Prediction Using Machine Learning" International Journal of Computational Science and Information Technology (IJCSITY) Vol.2, No.2, May 2014.
[7]. Hnin Wint Khaing, "Data Mining based Fragmentation and Prediction of Medical Data", International Conference on Computer Research and Development, ISBN: 978-1-61284-840-2,2011
[8]. Delen, Dursun, Glenn Walker, and Amit Kadam. "Predicting breast cancer survivability: a comparison of three data mining methods." Artificial intelligence in medicine 34.2 (2005): 113-127.
[9]. Priyanga, A., and S. Prakasam. "Effectiveness of Data Mining-based Cancer Prediction System (DMBCPS)." International Journal of Computer Applications 83.10 (2013).
[10]. P.Ramachandran N.Girija,T.Bhuvaneswari " Early Detection and Prevention of Cancer using Data Mining Techniques" in International Journal of Computer Applications (0975 – 8887) Volume 97– No.13, July 2014
[11]. Li, Lihua, et al. "Data mining techniques for cancer detection using serum proteomic profiling." Artificial intelligence in medicine 32.2 (2004): 71-83.
[12]. Williams, Kehinde, et al. "Breast cancer risk prediction using data mining classification techniques." Transactions on Networks and Communications3.2 (2015): 01.
[13]. Dash B, Mishra D, Rath A, Acharya M. A hybridized K-means clustering approach for high dimensional dataset. International Journal of Engineering, Science and Technology. 2010;2(2):59-66.
[14]. Mark Velasquez and Patrick T. Hester "An Analysis of Multi-Criteria Decision Making Methods" International Journal of Operations Research Vol. 10, No. 2, 56□66 (2013)
[15]. Gilberto Montibeller and Alberto Franco " Handbook of Multicriteria Analysis", Applied 25 Optimization 103, DOI 10.1007/978-3-540-92828-7_2, © Springer-Verlag Berlin Heidelberg 2010
[16]. http://eprints.lse.ac.uk/12761/1/Multi-criteria_Analysis.pdf
[17]. Praveen Thokala, Nancy Devlin, Kevin Marsh, Rob Baltussen, Meindert Boysen, Zoltan Kalo, Thomas Longren, Filip Mussen, Stuart Peacock, John Watkins, Maarten Ijzerman "Multiple Criteria Decision Analysis for Health Care Decision Making—An Introduction: Report 1 of the ISPOR
[18]. Baltussen R, Niessen L: Priority setting of health interventions: the need for multi-criteria decision analysis. Cost Eff Resour Alloc 2006, 4:14
[19]. Kristie Venhorst, Sten G Zelle1 , Noor Tromp and Jeremy A Lauer "Multi-criteria decision analysis of breast cancer control in low- and middle- income countries: development of a rating tool for policy makers" Venhorst et al. Cost Effectiveness and Resource Allocation 2014, 12:13
[20]. Dolan JG: Multi-criteria clinical decision support: a primer on the use of multiple criteria decision making methods to promote evidence-based, patient-centered healthcare. Patient 2010, 3(Suppl 4):229–248
[21]. Generous N, Margevicius KJ, Taylor-McCabe KJ, Brown M, Daniel WB, Castro L, et al.(2014) Selecting Essential Information for Biosurveillance—A Multi-Criteria Decision Analysis.PLoSONE(1):e86601.doi:10.1371/journal.pone.0086601 January 29, 2014

[22]. C. Chandrasekar, P.S. Meena "Microarray Gene Expression for Cancer Classification using Fast Extreme Learning Machine with ANP" International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue 2,Mar-Apr 2012, pp.229-235.
[23]. Nguyen T, Khosravi A, Creighton D, Nahavandi S (2015) Hierarchical Gene Selection and Genetic Fuzzy System for Cancer Microarray Data Classification. PLos ONE 10(3): e0120364 doi: 10.1371/ journal. pone.0120364

Biography:

Mrs. Pallavi Mirajkar passed M.Sc. (Computer Science) from North Maharashtra University, Jalgaon. Currently she is pursuing her Ph.D. from Pacific University, Udaipur.

Dr. G.Prasanna Lakshmi, Guide. Women Scientist (WOS-A) Andhra University.