IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 7, Issue 8, August 2018

Paper on Genetic Algorithm for Detection of Oral Cancer

Nooreen Fatima[1], Mohammad sameer[2]

Computer Science Student, CS Engineering Department, Integral University, Lucknow, India[1]

Associate Professor, CS Engineering Department, Integral University, Lucknow, India[2]

Abstract: According to NCBI research journal , oral cancer ranks in the top three of all cancers in India, which accounts for over thirty per cent of all cancers reported in the country and oral cancer control is quickly becoming a global health priority. In these years recently India is a country having a majority of oral cancer patients in comparison to other countries. In a survey it was determined that 12.5 per 1,00,000 people suffer from oral cancer . The major causes of this disease spreading as an epidemic is not hereditary or in born, but the lifestyle that people live over here. Chewing and smoking tobacco and other intoxicant weeds containing harmful strains of carcinogenic pathogens is a hobby over here in India. Both the wealthy and poor class of people have easy access to these intoxicants, over which it seems to be obvious that the government has no strict quality control . Oral cancer and tuberculosis are the 2 major widespread diseases that are blooming over the entire Indian nation. Alcohol also has an involvement in oral cancer to some extent, but tobacco is the major trigger for it. Work is being done in order to find out methods of early detection and treatment of people through routine checkups and surveys organized both by government and non government organizations. Early detection is important because diagnosis of tongue and oral cavity in early stages can help preventing it easily. With the help of predictive models; such as: decision tree model, genetic algorithm, tree boost model and decision tree forest model; early detection and treatment of patients can be done.

Keywords: Data Mining, Oral Cancer, Genetic Algorithm

I. INTRODUCTION

India where tobacco use is exuberant, has made its government to make cigarette manufacture include a graphic warning on packets to support the protection of minority who are uneducated . Oral cancer rates are very high in India. In comparison with the U.S. population, where oral cavity cancer represents only about 3% of malignancies, it accounts for over 30% of all cancers in India.

Paan masala products are also dangerous, as they contain areca nuts, a potential cancer-causing agent. Areca nuts have addictive properties similar to caffeine, tobacco, and alcohol and can lead to a high number of cases of sub mucous fibrosis, which can become malignant. India has a high incidence of oral cancer due to tobacco use, and as said by the Global Adult Tobacco Survey, the age of initiation of tobacco habits in India is 17 years.
All forms of tobacco use and alcohol abuse are known risk factors for oropharyngeal cancer. 57% of all men and 11% of women between 15–49 years of age use some form of tobacco in India. Those who smoke bidis have a three-fold higher risk of oral cancer compared with non-smokers, and are also at increased risk of lung, stomach and esophageal cancer. As mentioned, not selling to minors is one way to slow down use. Also, large graphic warnings are said by some to be of benefit.

The Indian National Cancer Registry Programme report shows worrying rises in cancers of the upper aero-digestive tract (mouth, tongue, oro-pharynx, hypo pharynx, larynx and oesophagus) among both sexes as important sites for undertaking risk factor research and implementing early detection programmes. Improved public health education and promotion is vital, as are top down policy approaches such as those of the Framework Convention on Tobacco Control, extended to include all forms of smokeless tobacco.

II. METHODOLOGY

The genetic algorithm is a method for solving both constrained and unconstrained optimization problems that is based on natural selection, the process that drives biological evolution. The genetic algorithm repeatedly modifies a population of individual solutions. At each step, the genetic algorithm selects individuals at random from the current population to be parents and uses them to produce the children for the next generation. Over successive generations, the population "evolves" toward an optimal solution. You can apply the algorithm to solve a variety of optimization

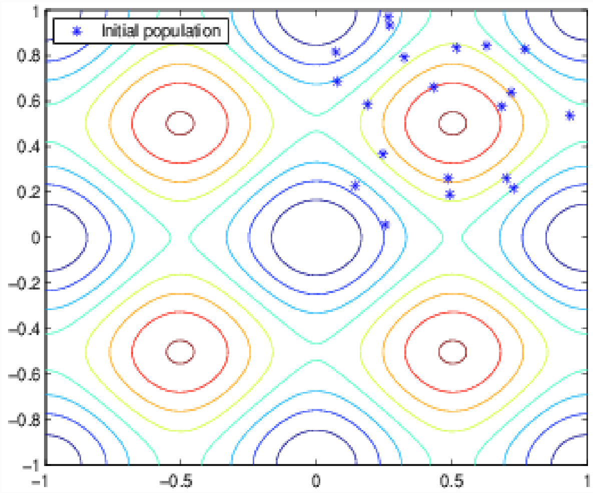
DOI 10.17148/IJARCCE.2018.7821

problems that are not well suited for standard optimization algorithms, including problems in which the objective function is discontinuous, non differentiable, stochastic, or highly nonlinear. The genetic algorithm can address problems of mixed integer programming, where some components are restricted to be integer-valued.

The genetic algorithm uses three main types of rules at each step to create the next generation from the current population:

- *Selection rules* select the individuals, called *parents*, which contribute to the population at the next generation.
- *Crossover rules* combine two parents to form children for the next generation.
- *Mutation rules* apply random changes to individual parents to form children.

Initial Population: The algorithm begins by creating a random initial population, as shown in the following figure.



In this example, the initial population contains 20 individuals, which is the default value of Population size in the Population options. Note that all the individuals in the initial population lie in the upper-right quadrant of the picture, that is, their coordinates lie between 0 and 1, because the default value of Initial range in the Population options is [0;1].

If you know approximately where the minimal point for a function lies, you should set Initial range so that the point lies near the middle of that range. For example, if you believe that the minimal point for Rastrigin's function is near the point [0 0], you could set Initial range to be [-1;1]. However, as this example shows, the genetic algorithm can find the minimum even with a less than optimal choice for Initial range.

Selection: The selection function chooses parents for the next generation based on their scaled values from the fitness scaling function. An individual can be selected more than once as a parent, in which case it contributes its genes to more than one child. The default selection option, Stochastic uniform, lays out a line in which each parent corresponds to a section of the line of length proportional to its scaled value. The algorithm moves along the line in steps of equal size. At each step, the algorithm allocates a parent from the section it lands on.

A more deterministic selection option is Remainder, which performs two steps:

- In the first step, the function selects parents deterministically according to the integer part of the scaled value for each individual. For example, if an individual's scaled value is 2.3, the function selects that individual twice as a parent.
- In the second step, the selection function selects additional parents using the fractional parts of the scaled values, as in stochastic uniform selection. The function lays out a line in sections, whose lengths are proportional to the fractional part of the scaled value of the individuals, and moves along the line in equal steps to select the parents.

Note that if the fractional parts of the scaled values all equal 0, as can occur using Top scaling, the selection is entirely deterministic.

III. REVIEW OF LITERATURE

Every year approximately 2,00,000 deaths worldwide and 46,000 deaths particularly in India account for oral cancer (Jemal et al. 2010). The study shows that the developing countries like Melanesia, South-Central Asia and Central and

Eastern Europe have the highest rate of oral cavity cancer, whereas the developed countries like Africa, Central America, and Eastern Asia have the lowest rate of oral cavity cancer, for both males and females (Ferlay et al. 2010).
Oral cancer, with its widely variable rate of occurrence, has one of the highest incidences in Indian subcontinent where it ranks among the top three types of cancer in the country (Elango et al. 2006). Age-adjusted rate of oral cancer in India is high, that is, 20 per 100,000 population that accounts for over 30 % of all cancerous persons in the country (Sankaranarayanan et al. 1927). It is of public health importance in India as it has been estimated that 83,000 new oral cancer cases occur here each year (Manoharan et al. 2004; Agrawal et al. 2012). The difficulty level is high because it is usually diagnosed at later stages which result in low treatment outcomes and considerable high cost to the patients who typically cannot afford this type of treatment (Khandekar et al. 2006). The prognosis for patients with oral cancer also remains poor in spite of advances in therapy of many other malignancies. Early diagnosis and treatment remains the key to improved patient survival. To achieve success in treatment, it is essential to determine the hidden patterns and trends in the oral disease data, which can be collected from oral healthcare industry and subsequently ‘‘analyzed or mined’’ to help healthcare practitioners for effective decision making.

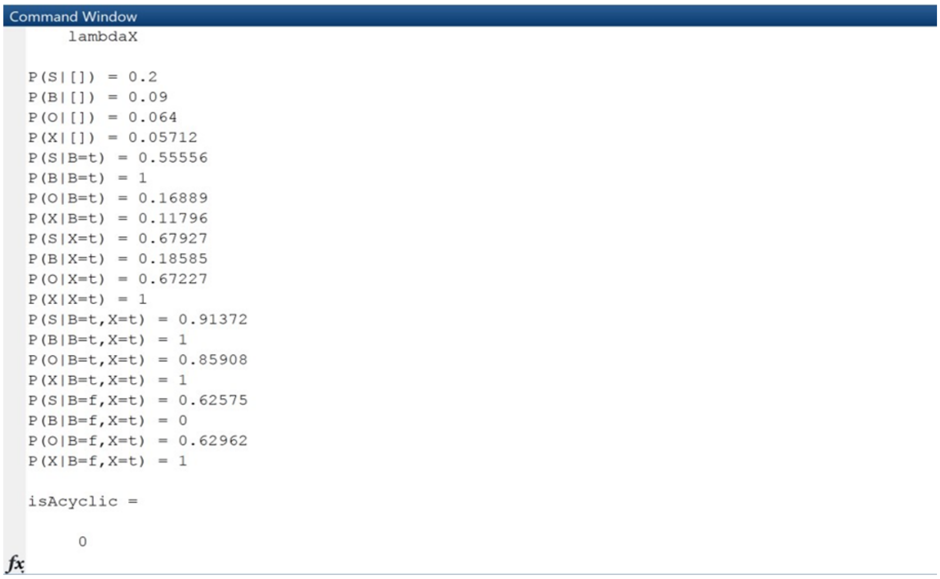
Manual extraction of patterns from the data has occurred for centuries. The Bayes’ theorem (1700s) and regression analysis (1800s) were considered to be early methods of identifying patterns in the data. The proliferation, ubiquity and increasing power of computer technology have dramatically increased the data collection, storage, and manipulation ability (Han and Kamber 2012). As datasets have grown in size and complexity, the direct ‘‘hands-on’’ data analysis has increasingly been augmented with indirect, automated data processing, aided by other discoveries in computer science, such as neural networks, cluster analysis, genetic algorithms (1950s), decision trees (1960s) and support vector machines (1990s). Data mining is the process of applying these methods with the intention of uncovering hidden patterns in large datasets (Kantardzic and Mehmed 2003).
Data mining (the analysis step of the ‘‘knowledge discovery in databases’’ process, or KDD), an interdisciplinary subfield of computer science, is a computational process of discovering patterns in large datasets (Fayyad et al. 1996; Data Mining Curriculum (2006); Clifton and Christopher 2010; Hastie et al. 2009). It bridges the gap from applied statistics and artificial intelligence (which usually provide the mathematical background) to database management by exploiting the way the data are stored and indexed in the databases to execute the algorithms more efficiently, allowing such methods to be applied to ever larger datasets. The overall goal of data mining process is to extract information from a dataset and transform it into an understandable structure for further use (Data Mining Curriculum 2006). Apart from the raw analysis, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization and online updating (Data Mining Curriculum 2006). There are various data mining tools available that can be used to predict behavior's and future trends, allowing businesses to make proactive, knowledge-driven decisions. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations. The information gained can be used to develop a model for prediction and classification of new data (Cunningham and Holmes 1999).

CONCLUSION

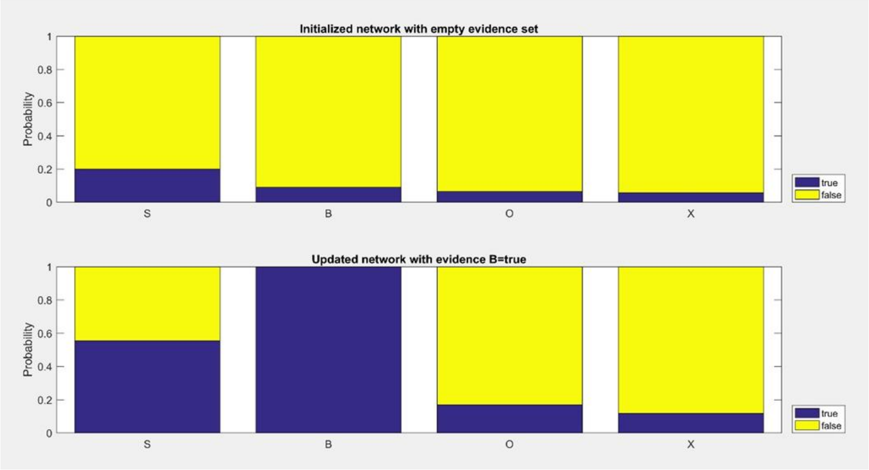
With the help of this study we can conclude that in spite of having a large amount of medical data, it lacks in the quality and the completeness of data because of which highly sophisticated data mining techniques are required to build up a efficient decision support system .Data mining techniques and methods applied in patient medical dataset has resulted in innovations, standards and decision support system that have significant success in improving the health of patients and the overall quality of medical services. In our work we have used genetic algorithm for prediction of oral cancer .By applying probability function on parameters and creating fitness function we have found the range of each data. So depending on different conditions there are different probability of having oral cancer. After applying genetic algorithm on data set we can conclude that probability of oral cancer is highest in case where probability of abnormal X-rays is 1 and probability of bronchitis is also 1 which can be concluded from the graph in previous chapters. Hence saving cost and time to undergo medical tests and checkups and ensuring that the patient can monitor his health on his own and plan preventive measures and treatment at the early stages of the diseases.

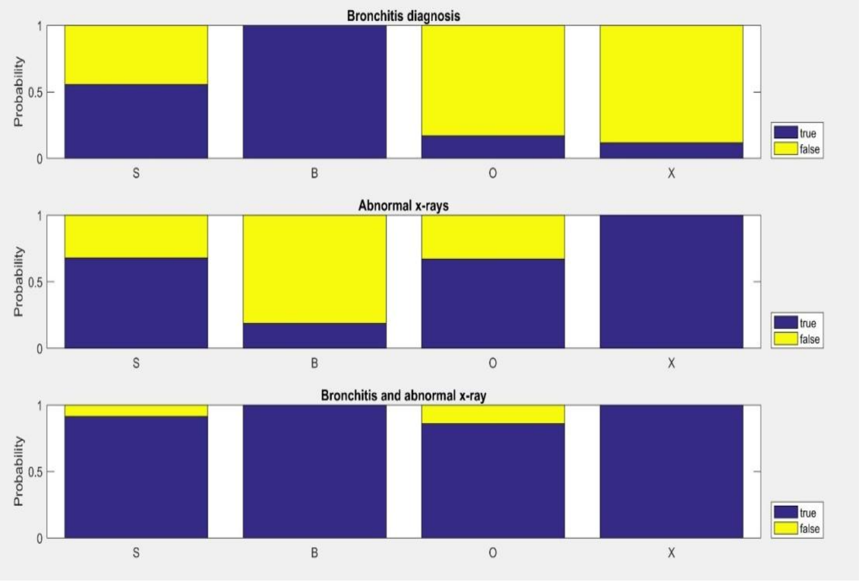
RESULT

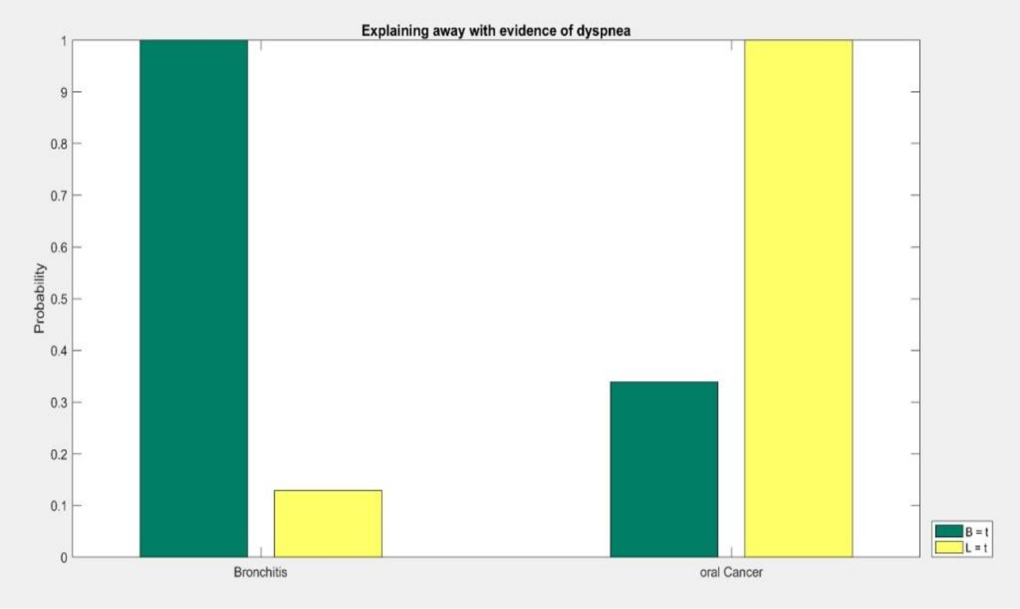
After using Genetic algorithm on the data set of oral cancer patients following results were obtained.

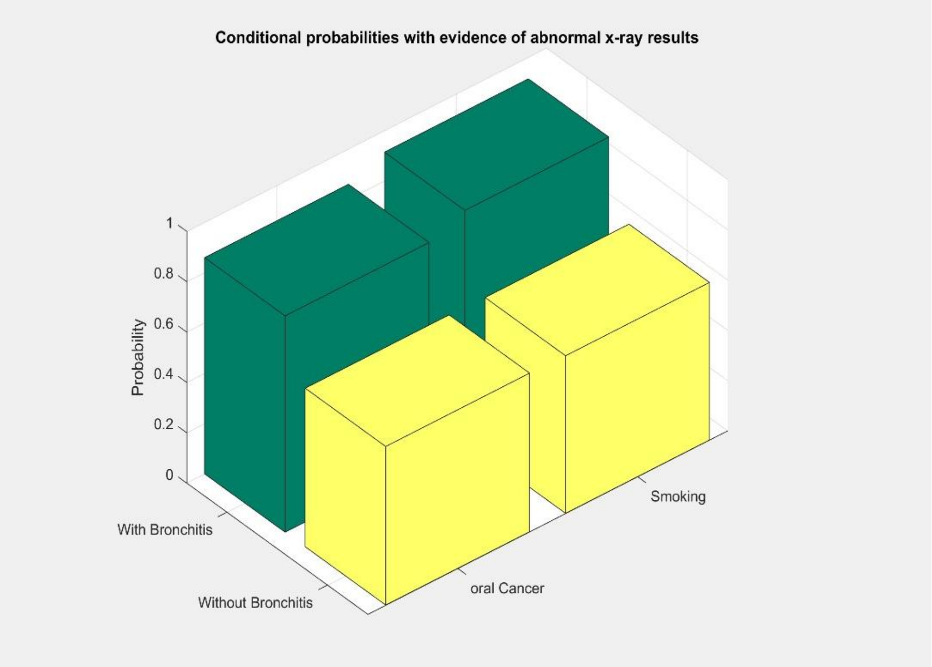


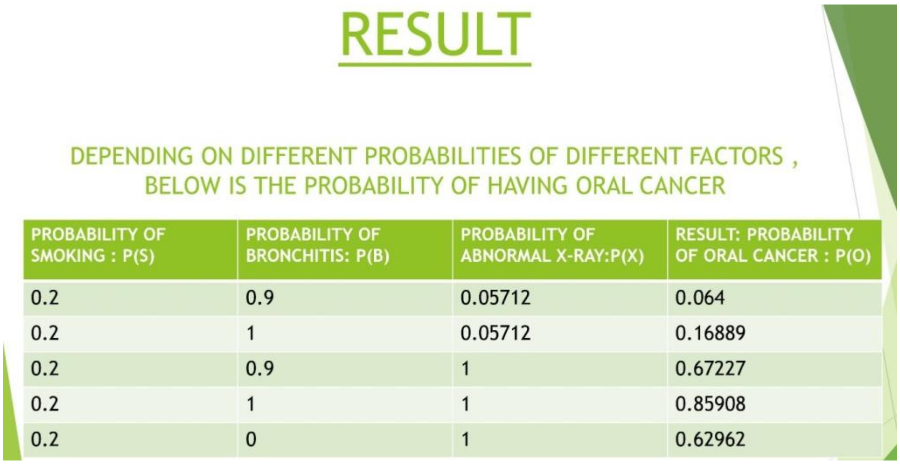
```
    lambdaX

P(S|[]) = 0.2
P(B|[]) = 0.09
P(O|[]) = 0.064
P(X|[]) = 0.05712
P(S|B=t) = 0.55556
P(B|B=t) = 1
P(O|B=t) = 0.16889
P(X|B=t) = 0.11796
P(S|X=t) = 0.67927
P(B|X=t) = 0.18585
P(O|X=t) = 0.67227
P(X|X=t) = 1
P(S|B=t,X=t) = 0.91372
P(B|B=t,X=t) = 1
P(O|B=t,X=t) = 0.85908
P(X|B=t,X=t) = 1
P(S|B=f,X=t) = 0.62575
P(B|B=f,X=t) = 0
P(O|B=f,X=t) = 0.62962
P(X|B=f,X=t) = 1

isAcyclic =

     0
```











RESULT

DEPENDING ON DIFFERENT PROBABILITIES OF DIFFERENT FACTORS , BELOW IS THE PROBABILITY OF HAVING ORAL CANCER

PROBABILITY OF SMOKING : P(S)	PROBABILITY OF BRONCHITIS: P(B)	PROBABILITY OF ABNORMAL X-RAY:P(X)	RESULT: PROBABILITY OF ORAL CANCER : P(O)
0.2	0.9	0.05712	0.064
0.2	1	0.05712	0.16889
0.2	0.9	1	0.67227
0.2	1	1	0.85908
0.2	0	1	0.62962

REFERENCES

[1]. http://www.wcrf.org/cancer_facts/5-most-common-cancers.php
[2]. http://www.cancer.gov/cancertopics/cancerlibrary/what-is-cancer
[3]. http://www.medicalnewstoday.com/info/cancer-oncology/
[4]. http://www.wcrf.org/cancer_facts/5-most-common-cancers.php
[5]. http://www.scribd.com/doc/9522057/an-overview-of-soft-computing-with-its-techniques-and-application
[6]. http://dl.acm.org/citation.cfm?id=1539227
[7]. http://www.iconip2012.org/s02.pdf
[8]. http://www.biomedcentral.com/1471-2105/6/148
[9]. http://news.harvard.edu/gazette/tag/cluster-of-cancer-cells/
[10]. http://www.healthcommunities.com/cancer-treatment-and-care/cancer-staging.shtml
[11]. http://en.wikipedia.org/wiki/Oncology#Progress_and_research
[12]. http://www.cancer.net/patient/All+About+Cancer/Cancer.Net+Feature+Articles/Treatments%2C+Tests%2C+and+Procedures/Understanding+Cancer+Research+Studies%2C+Part+II
[13]. K. Anuradha and K.Sankaranarayanan "International Journal of Advances in Engineering & Technology, March 2012" Vol. 3, Issue 1, pp. 84-91
[14]. M. E. Futschik, A. Reeve, and N. Kasabov, "Evolving connectionist systems for knowledge discovery from gene expression data of cancer tissue," Artif. Intell. Med., vol. 28, pp. 165–189, 2003
[15]. Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm by Latha Parthiban and R.Subramanian in International Journal of Biological and Life Sciences 3:3 2007
[16]. Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers by Jyoti Soni, Uzma Ansari, Dipesh Sharma in International Journal on Computer Science and Engineering (IJCSE)
[17]. Decision support system for heart disease diagnosis using neural network Delhi Business Review Vol. 8, No. 1 (January - June 2007)
[18]. Using neural networks to predict cardiac arrhythmias by Roland Adams,& Anthony Choi
[19]. Clinical Decision Support System : Risk level prediction of Heart disease using weighted fuzzy rules by P.K .Anuj in Journal of King Saud University – Computer and Information Sciences (2012) 24, 27–40
[20]. Modeling & Design of evolutionary network for heart disease detection by KS Kavitha and KV Ramakrishnan in IJCSI
[21]. A Data Mining Approach for Prediction of Heart Disease Using Neural Networks by Miss. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte published in IJCET in 2012
[22]. Microarray gene expression profile data mining model for clinical cancer research by Rui Xue ; Dept. of Inf. & Comput. Sci., Univ. of Hawaii, Honolulu, HI, USA ; Jianying Li ; Streveler, D.J. published in IEEE explore in 2004
[23]. Decision Support in Heart Disease Prediction System using Naive Bayes by Mrs.G.Subbalakshmi, Mr. K. Ramesh, Mr. M. Chinna Rao
[24]. Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction by Jyoti Soni, Ujma Ansari
[25]. Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network by Shantakumar B.Patil, Y.S.Kumaraswamy
[26]. Prediction of ECG Signals Missing Parts Using Artificial Neural Network by Najmeh Mohsenifar, Ali Sadr in Canadian Journal on Computing in Mathematics, Natural Sciences, Engineering and Medicine Vol. 2 No. 8, November 2011
[27]. "Identification of suspicious regions to detect oral cancers at an earlier stage– a literature survey"by K.Anuradha.
[28]. "Unification of heterogeneous data towards the prediction of oral cancer reoccurrence" by Konstantinos.
[29]. "Survey of Human Cancer Classification using Micro Array Data" by G. Sophia Reena
[30]. "Artificial Neural Networks Applied to Survival Prediction in Breast Cancer" by M. Lundina J. Lundina H.B. Burked S. Toikkanenb L
[31]. "An Extensive Survey on Artificial Neural Network Based Cancer Prediction Using Soft Computing Approach" written by Manaswini Pradhan, Dr. Ranjit Kumar Sahu.
[32]. Khosla, R. and Dillon, T. (1997) Knowledge Discovery, Data Mining and Hybrid Systems. Engineering, IntelligentHybrid MultiAgent Systems, Kluwer Academic Publishers, Norwell, 143-177.
[33]. Fayyad, U.M., Piatetsky-Shapiro, G. and Smyth, P. (1996) Data Mining to Knowledge Discovery: An Overview. Advancesin Knowledge Discovery and Data Mining, AAAI Press/MIT Press, 1- 36.
[34]. Cios, K.J. (2001) Medical Data Mining and Knowledge Discovery. Studies in Fuzziness and Soft Computing, 60, 502.
[35]. K.R.Coelho, "Challenges in Oral Cancer Burden in India," Journal of Cancer Epidemiology, vol. 2012, Article ID 701932, 17 pages.
[36]. Elango, J.K., Gangadharan, P., Sumithra, S. and Kuriakose, M.A. (2006) Trends of Head and Neck Cancers in Urban And Rural India. Asian Pacific Journal of Cancer Prevention, 7, 108-112.
[37]. American Cancer Society (2012) Cancer Facts and figures, Atlanta (GA), The Society
[38]. Changsheng Xiang, ZiYing Zhou, "A New Music Classification Method based on BP Neural Network", JDCTA, Vol. 5, No. 6, pp. 85 ~ 94, 2011.
[39]. Han Xiao, Yuanjiang Li, "A New Thought based on the Service Composition of Automatic Transmission Semantic Grid in Internet of Things", IJACT, Vol. 3, No. 7, pp. 10 ~ 16, 2011
[40]. Ren Fang, Ma Jian-Feng, "Attribute-Based Access Control Mechanism for Perceptive Layer of the Internet of Things", JDCTA, Vol. 5, No. 10, pp. 396 ~ 403, 2011.
[41]. Li xinwu, "A New Color Correction Model for based on BP Neural Network", AISS, Vol. 3, No. 5, pp. 72 ~ 78, 2011. Haykin, S. (1998). "Neural Networks: A Comprehensive Foundation", Prentice-Hall, Upper Saddle River, U.S.A.
[42]. Kecman, V. (2001). "Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models", The MIT Press, Cambridge, U.S.A. Kjartansson, E. (1979). "Constant *Q*-wave Propagation and Attenuation", Journal of Geophysical Research, Vol. 84, No. B9, pp. 4737–4748.
[43]. Ismail Taha and Joydeep Ghosh ,Wisconsin breast cancer database using a hybrid symbolic- connectionist system," university of Texas Austin 1996.
[44]. Esugasini Subramaniam, Tan Kuan Liung, Mohd. YusoffMashor,Nor Ashidi Mat Isa ,Breast Cancer