Early Detection and Prevention of Oral Cancer: Association Rule Mining on Investigations

NEHA SHARMA[1], HARI OM[2]
[1]Padmashree Dr. D.Y.Patil Institute of MCA, [2]Computer Science and Engineering,
[1]University of Pune, [2]Indian School of Mines
[1]Sector-24, Nigdi Pradhikaran, Pune, Maharashtra; [2]Indian School of Mines, Dhanbad, Jharkhand
[1]INDIA; [2]INDIA
[1]nvsharma@rediffmail.com; [2]hariom4india@gmail.com

Abstract: - Early detection and prevention of oral cancer is critical, as it can increase the survival chances considerably, allow for simpler treatment and result in a better quality of life for survivors. In this research paper, the popular association rule mining algorithm, apriori is used to find the spread of cancer with the help of various investigations and then assess the chance of survival of the patient. This is achieved by extracting a set of significant rules among various laboratory tests and investigations like FNAC of neck node, LFT, Biopsy, USG, CT scan-MRI and survivability of the oral cancer patients. The rules clearly show that if FNAC of neck node, USG and CT scan/ MRI is positive then chance of survival is reduced. However, if LFT is normal, probability of survival is high. If diagnostic-biopsy results in squamous-cell-carcinoma then it clearly indicate oral cancer, which may lead to high mortality if appropriate treatment is not initiated. The experimental results demonstrate that all the generated rules hold the highest confidence level, thereby, making investigations very essential to understand the spread of cancer after clinical examination for early detection and prevention of oral cancer.

Key-Words: - Data Mining, Association Rule Mining, Apriori, Oral Cancer, Weka, Investigations

1 Introduction

Medicine is an information-intensive profession. Developments in cancer medicine have customarily hailed from meticulous comprehension of biological processes, later translated into therapeutic interventions, whose competence is built by thorough analysis of clinical trials. Throughout the most recent two decades the expanding throughput of microarray screening, spectral imaging and longitudinal studies are transforming the comprehension of cancer pathology into as much a data-based as a biologically and clinically driven science, with potential to impact more strongly on evidence-based decision support moving towards personalized medicine [1]. Apart from microarray screening and spectral imaging, the data are also available in the form of patient charts and histories, research literature and evidence [2]. The large quantity of data definitely surpasses the abilities of humans for efficient usage without specialized tools for analysis [3, 4, 5]. The situation is described as rich in data, but poor in information. In order to fill this growing gap, different approaches from the field of Data Mining are applied. Ideally, data mining integrates all available medical information, epidemiological data, and pertinent information collated from across time zones, systems and databases, which is subsequently analysed.

Data mining, also referred to as knowledge discovery, is the science of extracting critical information from large amount of existing raw data and deploying that information across the organization [3, 5, 6]. It is the process of automatic search to discover patterns and trends that go beyond simple analysis and can answer questions that cannot be addressed through simple query and reporting techniques. But data mining does not work by itself. Data mining uses sophisticated mathematical algorithms to segment the data and evaluate the probability of future events. Automatic discovery of patterns, prediction of likely outcomes, creation of actionable information and focus on large data sets and databases are the key properties of data mining [7, 8, 9]. It does not eliminate the need to know the business, to understand the data, or to understand analytical methods. Data mining discovers hidden information in data, however it cannot tell the value of the information to your organization. Important patterns might already be known as a result of acquaintance to business domain and working with data over time. However, data mining can confirm or qualify such empirical observations in addition to finding new patterns that may not be immediately discernible through simple observation.

Data mining tasks to uncover the different patterns include characterization, discrimination, association analysis, classification and regression, cluster analysis, outlier analysis and evolution analysis. In this paper, we intend to apply association rule learning, which is a popular and well researched method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using different measures of interestingness [10]. These rules have many applications in areas ranging from e-commerce to sports to census analysis to medical diagnosis.

We have undertaken the study of oral cancer which is of significant public health importance in India. Public health officials, private hospitals, and academic medical centres within India have recognized oral cancer as a grave problem [11]. Possible signs and symptoms of oral cancer when a patients may report include: a lump or thickening in the oral soft tissues, soreness or a feeling that something is caught in the throat, difficulty chewing or swallowing, ear pain, difficulty moving the jaw or tongue, hoarseness, numbness of the tongue or other areas of the mouth, or swelling of the jaw that causes dentures to fit poorly or become uncomfortable. The above mentioned signs and symptoms observed in the patients on their visit to OPD, must be used efficiently for early detection and treatment, as they are critical and can increase the survival chances considerably, allow for simpler treatment and result in a better quality of life for survivors. In the previous paper "Significant Pattern for Oral Cancer Detection: An Association Rule Mining" written by authors, many significant rules among various valuable information pertaining to clinical examination, history and survivability of the cancer patients were extracted to assist the practitioners in early detection of the disease and prediction of distribution of cancer in the oral cavity, and consequently help in prevention of the oral cancer. After successfully generating vital rules from the details of clinical examination, we now extend our work by attempting to find the spread of cancer with the help of various investigations and then assess the chance of survival of the patient.. In this research paper, the popular association rule mining algorithm, apriori is used to extract a set of significant rules among various laboratory tests and investigations like FNAC of neck node, LFT, Biopsy, USG, CT scan-MRI and survivability of the oral cancer patients.

This paper is organized as follows: The next section reviews various related literature, section 3 covers the information about oral cancer and various investigating techniques; section 4 gives the brief about Association Rule Mining; section 5 presents the experimental results; whereas section 6 covers the conclusions and future work and at the end acknowledgement and references made are mentioned.

2 Literature Review

RuthRamya et al. [12], apply association rules into classification to improve the accuracy and obtain the valuable rules and information in the case of chest pain. The class label has taken good advantage in the rule mining step so as to cut down the search space. The proposed algorithm also synchronize the rule generation and classifier building phases, shrinking the rule mining space when building the classifier to help speed up the rule generation. Swami S. et al. [13], aimed at generating multidimensional association rule and its model of smoking habits in order to take some preventive measures to reduce the various habits of smoking in youths. Ha and Joo [14], builds hybrid method combining association rule and classification trees which aims at helping physician to make fast and accurate classification of chest pain disease. Singh S. et. al. [15] applied the apriori algorithm with transaction reduction on cancer symptoms. They considered five different types of cancer and tried to find the symptoms helping the cancer to spread and also the type of cancer that spreads faster. Srikant et al. [16] considered the problem of integrating constraints in the form of boolean expression that appoint the presence or absence of items in rules. Few authors in their research papers focused on mining frequent item sets and association rules from the viewpoint of the user's interaction with the system [17, 18, 19, 20, 21].

Milovic B. et al.[22] present the applicability of data mining in healthcare and explains how these patterns can be used by physicians to determine diagnoses, prognoses and apply treatments for patients in healthcare organizations. Anuradha K. et al. [23] have done a detailed survey on various methods adopted by the researchers for oral cancer detection at an earlier stage. A comparison is made among the various methods for identification and classification of cancers. Nahar J. et al. [24] extract the significant prevention factors for particular type of cancer. To find out the prevention factors, they have first constructed a prevention factor dataset through an extensive literature review. Subsequently, three association rule mining algorithms: Apriori, Predictive apriori, and Tertius algorithms have been employed in order to discover most of the significant prevention factors against a specific type of cancer. Experimental results illustrate that Apriori is the most useful association rule-mining algorithm for discovery of prevention factors.

Chuang et al. [25] consider DNA repair genes. They chose single nucleotide polymorphisms (SNPs) dataset with 238 samples of oral cancer and control patients for disease prediction. All prediction experiments were conducted using the support vector machine and they reported that the performances of the holdout cross validation was superior to 10-fold cross validation, and the best classification accuracy was 64.2%. Gadewal et al. [26] compiled and enlarged the oral cancer gene database to include 374 genes by adding 132 gene entries to enable fast retrieval of updated information. Kaladhar et al. [27] predict oral cancer survivability using classification algorithms that include CART, Random Forest, LMT and Naïve Bayesian classification algorithms. These algorithms classify the cancer survival using 10 fold cross validation and training data set. The Random Forest classification technique correctly classifies the cancer survival dataset leading to absolute relative error relatively less as compared to other methods.

3 Oral Cancer and Investigating Techniques

Oral cancer is a subtype of head and neck cancer and is any cancerous growth located in any sub sites of the oral cavity [28]. It may arise as a primary lesion originating in any of the oral tissues, by metastasis from a distant site of origin, or by extension from a neighbouring anatomic structure, such as the nasal cavity. Also, the Oral cancers may originate in any of the tissues of the mouth, and may be of varied histologic types: SCC, teratoma, adenocarcinoma derived from a major or minor salivary gland, lymphoma from tonsillar or other lymphoid tissue, or melanoma from the pigment-producing cells of the oral mucosa. However the commonest histologic type is squamous cell carcinomas accounting for 90% of cancer [29]. Oral or mouth cancer most commonly involves the tongue. It may also occur on the floor of the mouth, cheek, gingiva (gums), lips, or palate (roof of the mouth).

The symptoms for an oral cancer at an earlier stage [30] are: 1) Patches inside the mouth or on lips that are white, red or mixture of white and red, 2) Bleeding in the mouth 3) Difficulty or pain when swallowing, 4) A lump in the neck. These symptoms should raise the suspicion of cancer and needs proper treatment. Treatments for Oral Cancer include surgery, radiation therapy and chemotherapy [31]. But even this is not always successful as 70% of the cases relapses and the results in death. The treatment is successful only if the lesion is diagnosed early, but sadly many times, it is ignored and the patient reports late when the lesion is untreatable. The cost of the treatment runs in lakhs and in spite of this there is no guarantee of cure. The surgery is morbid, often disfiguring the face [32].

Prognosis of oral cancer depends on early diagnosis. Despite advanced surgical techniques and other treatment modalities, the 5-year survival rate remains ~40-50% [33, 34]. Unfortunately, oral cancer is usually detected when it becomes symptomatic. An early disease is difficult to be differentiated from benign lesions. Therefore recognizing point of high risk of developing oral cancer is of importance [35].

Various laboratory tests and investigating techniques that can be used for assessing the extent of oral cancer in patient's body are as follows:

3.1 Liver Function Tests (LFT), are groups of clinical biochemistry laboratory blood assays designed to give information about the state of a patient's liver. Most liver diseases cause only mild symptoms initially, but it is vital that these diseases be detected early.

3.2 Fine Needle Aspiration Cytology (FNAC) is a type of biopsy procedure in which a thin needle is inserted into an area of abnormal-appearing tissue or body fluid to collect the sample.

3.3 Biopsy - If any abnormalities are discovered during the exam, a small tissue sample, or biopsy, usually is taken. This biopsy is important, as it is the only sure way to know if the abnormal area is cancer.

3.4 Ultra-SonoGraphy (USG) is an ultrasound based diagnostic imaging technique used for visualizing intra-abdominal structures.

3.5 A CT Scan and an MRI operate differently and are better suited for assessing the loco-regional spread of cancer. An MRI suited for examining soft tissue (ligament and tendon injury, spinal cord injury, brain tumors etc.) where as a CT scan is better suited for bone injuries, lung and chest imaging, and detecting cancers.

4 Association Rule Mining

In this paper we adopt the standard definition of association rules [36, 37, 38, 39, 40, 41, 42]. Association rule mining is being applied to search for hidden relationships among the attribute. It is intended to identify strong rules discovered in databases using different measures of interestingness. Therefore, an association rule is a pattern that states when X occurs, Y occurs with certain probability.

4.1 Apriori Algorithm

Association rules mining using apriori algorithm uses a "bottom up" approach, where frequent item sets are extended one item at a time (a step known as candidate generation), and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found. The algorithm uses breadth-first search and a hash tree structure to count candidate item sets efficiently. Apriori is a two-step process, where in the first step, frequent test-sets are discovered and in the second step, association rules are derived from the frequent test-sets [43], algorithm of the same is mentioned below:

4.1.1 Apriori algorithm: Candidate Generation and Test Approach

Step1: Initially, scan Database once to get frequent 1-itemset.
Step2: Generate length (k+1) candidate itemsets from length k frequent itemsets.
Step3:Test the candidates against Database.
Step4: Terminate when no frequent or candidate set can be generated.

4.1.2 Pseudo-code for the Apriori algorithm is as follows:

```
L1 = {frequent items};
for (k = 1; Lk !=∅; k++) do begin
     Ck+1 = candidates generated from Lk;
     for each transaction T in database do
               increment the count of all candidates
               in Ck+1that are contained in T
     Lk+1  = candidates in Ck+1 with min_support
    end
return ∪k Lk;
```

Where, Ck: Candidate itemset of size k
Lk: frequent itemset of size k

4.2 Rule Measures

To select interesting rules from the set of all possible rules generated, constraints on various measures of significance and interest can be used. The best-known constraints are minimum thresholds on support and confidence.

4.2.1 Support: The rule holds with support supp in T (the transaction data set) if sup% of transactions contain $X \cup Y$ [36].

$$Supp(X \rightarrow Y) = P(X \cup Y).$$

4.2.2 Confidence: The rule holds with confidence conf in T if conf % of transactions that contain X also contain Y [36, 44].

$$\begin{aligned} Conf\,(X \rightarrow Y) &= P(Y \mid X) \\ &= Supp(X \cup Y) / Supp(X) \\ &= P(X \text{ and } Y) / P(X) \end{aligned}$$

4.2.3 Lift: It is the probability of the observed support to that expected if X and Y were independent [45].

$$\begin{aligned} Lift(X \rightarrow Y) &= Supp(X \cup Y) / Supp(X) \text{ x } Supp(Y) \\ &= P(X \text{ and } Y) / P(X)P(Y) \end{aligned}$$

4.2.4 Leverage: It measures the difference of X and Y appearing together in the data set and what would be expected if X and Y where statistically dependent [10].

$$Lev(X \rightarrow Y) = P(X \text{ and } Y) - (P(X)\,P(Y))$$

4.2.5 Conviction: It is the probability of the expected frequency that X occurs without Y (that is to say, the frequency that the rule makes an incorrect prediction) [46].

$$\begin{aligned} Conv(X \rightarrow Y) &= 1 - Supp(Y) / 1 - conf\,(X \rightarrow Y) \\ &= P(X)\,P(\text{not } Y) / P(X \text{ and not } Y) \end{aligned}$$

5 Experimental Results

The database for the current research work is created by collecting data through a retrospective chart review from ENT and Head and Neck Department related to Oral Cancer from the records of the Cancer Registries of Tertiary Care Hospitals, OPD data sheet and archives of departments of Histopathology, Surgery and Radiology. Clinical details, personal history and habits were collected manually from the records to complete the datasheet of the patients. The data collection was done in non-randomized or non-probabilistic method, as all the data in the registries for the period of five years was considered. The dataset is based on the records of all the patients who have been reported with a lesion and treated at the centre from Jan 2004 and June 2009. The complete process of data preparation, data integration and data cleaning (ie. removing missing values, noisy data and inconsistent data) to create the database of oral cancer patients was presented in the previous paper “Framework for early detection and prevention of oral cancer using data mining” published by the authors [47]. There are total 33 variables and 1025 records of patients were created for the analysis.

A data mining tool, Weka 3.7.9 is used for implementation of the association rule mining. It is a collection of open source of many data mining and machine learning algorithms, including pre-processing on data, classification, clustering and association rule extraction. It is Java based open source tool created by researchers at the University of Waikato in New Zealand [48]. The oral cancer data is initially stored in MS Excel sheet, then converted into comma separated values (.csv file) and subsequently to attribute relation file format (.arff file), which is the acceptable format to weka tool.

Apriori algorithm has been implemented for finding the association rules among various investigations (like FNAC of neck node, LFT, Biopsy, USG, CT scan-MRI), diagnosis and survivability of the cancer patients to find the importance of laboratory tests and investigations to obtain a extent of oral cancer. Following rules are generated with minimum support defined by the tool as 0.1 (103 instances) and minimum confidence as 0.9:

Rule1:LFT=Normal(449) ==> Survival=Dead (449)
conf:(1) lift:(1) lev:(0) conv:(0)

Rule Explanation: Rule 1 recommends that if LFT is normal, probability of survival is high. Fig. 1 reflects the rule in graphical form.

Rule 2: Neck-Nodes=Present (131) ==> FNAC-of-Neck-Node = Positive (131)
conf:(1) lift:(1.79) lev:(0.06) conv:(57.77)

Rule 3: Survival=alive (449) ==> FNAC-of-Neck-Node = Negative (449)
conf:(1) lift:(1.15) lev:(0.06) conv:(57.38)

Rule Explanation: Rule 2 and Rule 3 advocate that if FNAC of neck node is positive then the chance of survival is reduced. Fig. 2 reflects the rule in graphical form.

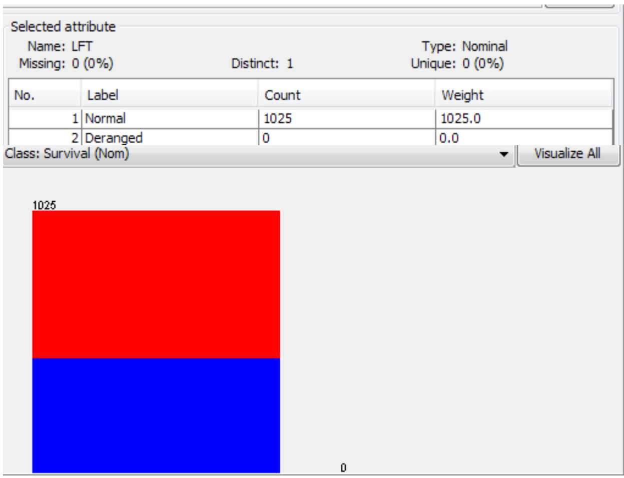

Fig. 1 LFT and Survivability

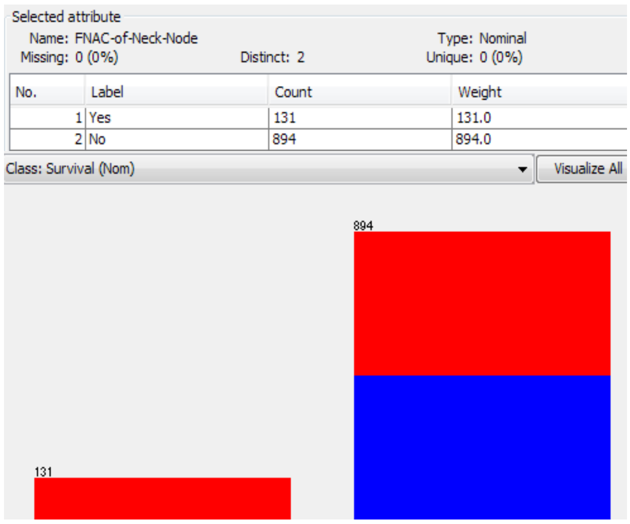

Fig. 2 FNAC of Neck Node and Survivability

Rule 4: Diagnostic-Biopsy=Squamous-Cell-Carcinoma(576) ==> Survival=Dead (576)
conf:(1) lift:(1.78) lev:(0.25) conv:(252.32)

Rule Explanation: Rule 4 proposes that if diagnostic-biopsy results in squamous-cell-carcinoma then it clearly indicate oral cancer. This may lead to high mortality if appropriate treatment is not initiated. Fig. 3 reflects the rule in graphical form.

Rule5:USG=Positive(132)==>Survival=Dead (132)
conf:(1) lift:(1.78) lev:(0.06) conv:(57.82)

Rule Explanation: Rule 5 suggests that if USG is positive then it detects the spread of oral cancer in patient's body and also the chance of survival is meagre. Fig. 4 reflects the rule in graphical form.

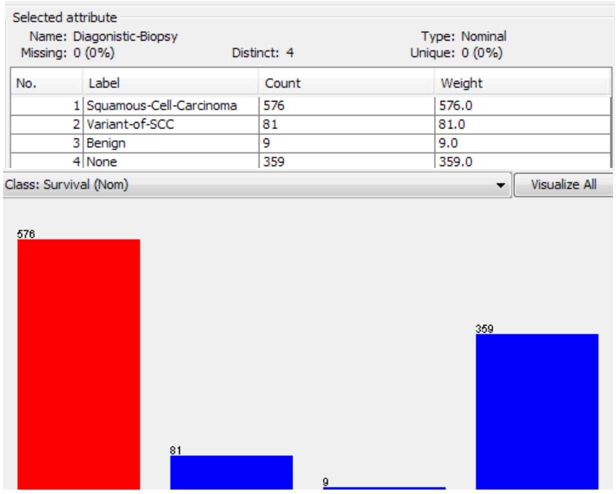

Fig. 3 Diagnostic Biopsy and Survivability

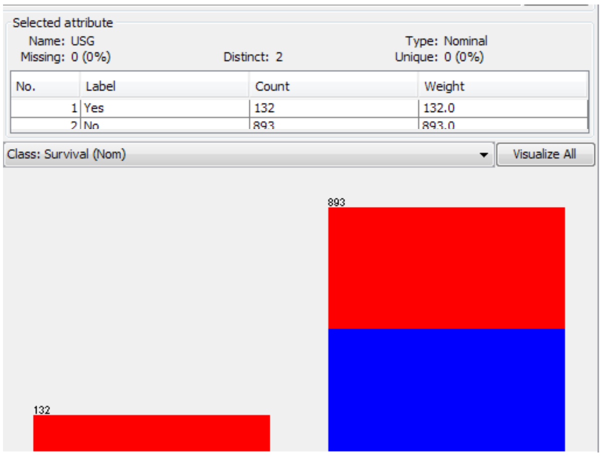

Fig. 4 USG and Survivability

Rule 6: CTScan-MRI=Bony-Involvement (130)
==> Survival=Dead (129)
conf:(0.99)lift:(1.77) lev:(0.05) conv:(28.47)

Rule Explanation: Rule 6 puts forward that if CT scan/ MRI detect the bony involvement, then the chances of survival is less. Fig. 5 reflects the rule in graphical form.

Rule7: Diagnosis=SCC(576)=>Survival=Dead (576)
conf:(1) lift:(1.78) lev:(0.25) conv:(252.32)

Rule Explanation: Rule 7 suggests that if diagnosis indicates SCC (Squamous-Cell-Carcinoma) which is a type of oral cancer, then it may lead to high mortality if appropriate treatment is not initiated.

Rule 8: Site=Tongue/Palate/BM &
Diagnostic–Biopsy=Squamous-Cell-Carcinoma==>Diagnosis=SCC
conf:(1) lift:(1.78) lev:(0.13) conv:(137.99)

Rule Explanation: Rule 8 suggests that if site is either tongue, plate or BM and diagnostic–biopsy is Squamous–Cell-Carcinoma then diagnosis usually is SCC which is a type of oral cancer and needs proper treatment.

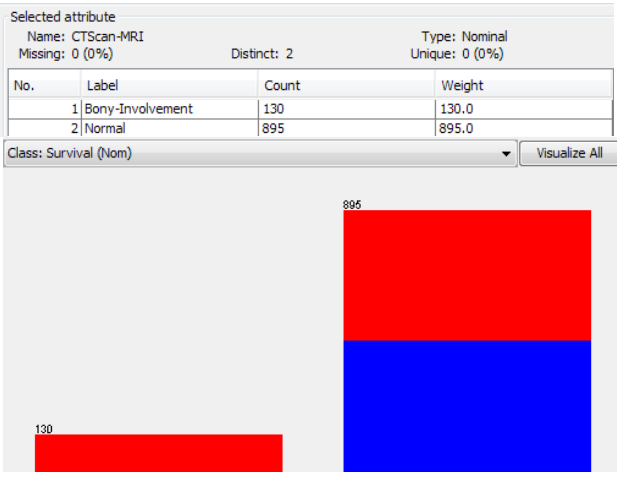

Fig. 5 CT Scan-MRI and Survivability

6 Conclusion

The association rule mining algorithm, apriori has established the importance of investigations and laboratory tests in assessing the degree and extent of oral cancer. The experimental results demonstrate that all the generated rules hold the highest confidence level, thereby, making them very useful for early detection and subsequently prevention of oral cancer. The rules clearly show that if FNAC of neck node, USG and CT scan/ MRI is positive then chance of survival is reduced. However, if LFT is normal, probability of survival is high. If site is either tongue, plate or BM and diagnostic–biopsy is Squamous–Cell-Carcinoma then diagnosis usually is SCC then it clearly indicate oral cancer, which may lead to high mortality if appropriate treatment is not initiated. In future, we intend to extend this research by attempting to extract significant patterns and useful rules through the association rule mining algorithm from extracting most effective course of treatment.

References:

[1] P. J. G. Lisboa, A. F. G. Taktak, "The use of artificial neural networks in decision support in cancer: A systematic review," Neural Netw., vol. 19, pp. 408–415, 2006.

[2] L. H. Andrea, C. Hsinchun, M. H. Susan, R.S. Bruce, D. N. Tobun, R. S. Robin and M. T. Kristin, “Medical Data Mining on the Internet: Research on a Cancer Information System,” Artificial Intelligence Review, 13.437–466, 1999.
[3] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, Third Edison. Morgan Kaufmann Publishers. ISBN 978-0123814791, July 2011.
[4] Hen, Lai Ee 2008. Performance Analysis of Data Mining Tools Cumulating with a Proposed Data Mining Middleware. Journal of Computer Science.
[5] U. M. Fayyad, G. Piatetsky-Shapiro and P. Smyth, “From Data Mining to Knowledge Discovery in Databases,” American Association for Artificial Intelligence (AAAI-AI Magazine), 37–54, 1996
[6] U. M. Fayyad, G. Piatetsky-Shapiro and P. Smyth, “From Data Mining to Knowledge Discovery: An Overview,” Advances in Knowledge Discovery and Data Mining (AAAI Press/MIT Press), 1–36, 1996
[7] Data Mining Curriculum. ACM SIGKDD. 2006-04-30.
[8] C. Clifton, Encyclopædia Britannica: Definition of Data Mining, 2010
[9] T. Hastie, R.Tibshirani, and J. Friedman, “The Elements of Statistical Learning: Data Mining, Inference, and Prediction,” 2009.
[10] G. Piatetsky-Shapiro, “Discovery, analysis, and presentation of strong rules,” Knowledge Discovery in Databases, AAAI/MIT Press, Cambridge, MA. 229-248, 1991
[11] K.R.Coelho, “Challenges in Oral Cancer Burden in India,” Journal of Cancer Epidemiology, vol. 2012, Article ID 701932, 17 pages.
[12] K. RuthRamya et al, “A Class Based Approach for Medical Classification of Chest Pain”, International Journal of Engineering Trends and Technology, Vol. 3, Issue. 2, pp.89-93, 2012
[13] Swami S. et al, “ Multidimentional Association Rules Extraction in Smoking Habit Database”, International Journal of Advanced Networking and Applications, vol: 03, issue: 03, pp. 1176-1179 (2011)
[14] Sung Ho Ha and Seong Hyeon Joo, “ A Hybrid Data Mining Method for Medical Classification of Chest Pain”, World Academy of Science, Engineering and Technology, 37, pp: 608-613, 2010.
[15] S. Singh, M. Yadav and H. Gupta, “Finding the Chances and Prediction of Cancer through Apriori Algorithm with Transaction Reduction,” International Journal of Advanced Computer Research. (ISSN (print): 2249-7277 ISSN (online): 2277-7970, vol. o2, issue 02, pp: 23-28, June 2012.
[16] R. Srikant, Q. Vu, and R. Agrawal, “Mining association rules with item constraints,” Proceeding KDD9, pp: 67-73, 1997
[17] T. N. Anh, D. V. Hai, T. C. Tin and L. H. Bac, “Efficient Algorithms for Mining Frequent Itemsets with Constraint,” Proceedings of the third international conference on knowledge and systems engineering, 2011.
[18] R. J. Bayardo, R. Agrawal and D. Gunopulos, “Constraint-Based Rule Mining in Large, Dense Databases,” Data Mining and Knowledge Discovery, Kluwer Academic Publication, vol. 4, issue 2/3, pp. 217–240, 2000.
[19] G. Cong, and B. Liu, “Speed-up Iterative Frequent Itemset Mining with Constraint Changes,” ICDM, pp: 107- 114, 2002.
[20] A. J. Lee, W. C. Lin, and C. S. Wang, “Mining Association rule with multi-dimensional constraints,” Journal of Systems and Software, pp: 79-92, 2006.
[21] R. T. Nguyen, V. S. Lakshman, J. Han and A. Pang, “Exploratory Mining and Pruning Optimizations of Constrained Association Rules,” International Conference on Management of Data, ACM-SIG-MOD. 13-24, 1998.
[22] B. Milovic and M. Milovic, “ Prediction and decision making in health care using data mining. International Journal of Public Health Science. 01, 02 (December 2012), 69-78, 2012.
[23] K. Anuradha, and K. Sankaranarayanan, “Identification of suspicious regions to detect Oral cancers at an earlier stage– a literature Survey,” International Journal of Advances in Engineering & Technology, vol. 03, issue 01, pp: 84-91, March 2012.
[24] J. Nahar, S. T. Kevin, A. B. M. S. Ali and Y. P. Chen, “Significant cancer prevention factor extraction: An association rule discovery approach,” J Med Syst, Springer, October 2009, DOI 10.1007/s10916-009-9372-8
[25] L. Y. Chuang, K. C. Wu, H. W. Chang and C. H. Yang, “Support Vector Machine-based Prediction for Oral Cancer Using Four snps in DNA Repair Genes,” Proceedings of the International MultiConference of Engineers and Computer scientists, March 16-18 2011.

[26] N. S. Gadewal and S. M. Zingde, " Database and interaction network of genes involved in oral cancer: Version II," Bioinformation, vol. 06, issue 04, pp. 169-170, 2011.
[27] DSVGK Kaladhar, B. Chandana and P. B. Kumar, "Predicting cancer survivability using Classification algorithms," International Journal of Research and Reviews in Computer Science (IJRRCS), vol. 02, issue 02, pp. 340-343, April 2011.
[28] Werning, John W, Oral cancer: diagnosis, management, and rehabilitation, p. 1. ISBN 978-1588903099, May 16, 2007.
[29] http://www.oralcancerfoundation.org/facts/index.htm
[30] Crispian Scully, Jose.V. Bagan, Colin Hopper, Joel. B. Epstein, "Oral Cancer: Current and future diagnostics techniques – A review article", American Journal of Dentistry, vol. 21, issue 4, pp 199 – 209, August 2008.
[31] www.yourtotalhealth.ivillage.com
[32] www.oralcancerawareness.org
[33] A. Jemal, A. Thimas, T. Murray and M. Thun, Cancer Statistics, 2002, CA Cancer J Clin 2002; 52:181-2.
[34] J. A. Woolgar, J Scott, E. D. Vaughan, J. S. Brow, C. R. West, S. Rogers, Survival, metastasis and recurrence of oral cancer in relation to pathological features, Ann R Coll Surg Engl 1995;77:325-31.
[35] E. Allegra, N. Lombardo, L. Puzzo1, A. Garozzo, The usefulness of toluidine staining as a diagnostic tool for precancerous and cancerous oropharyngeal and oral cavity lesions, ACTA oto rhinolaryngologica italica 2009;29:187-190
[36] R. Agrawal, T. Imielinski and A. Swami, A, "Mining association rules between sets of items in large databases," Proceedings of the 1993 ACM SIGMOD international conference on Management of data, pp. 207–216, 1993.
[37] J. An, Y. P. P. Chen and H. Chen, "DDR: An Index method for large time series datasets," Information System, vol. 30 pp. 333–348, 2005.
[38] Y. P. P Chen and F. Chen, "Targets for drug discovery using Bioinformatics," Expert opinion ther. Targets, vol. 12, issue 04, pp. 383–389, 2008.
[39] R. Y. K Lau, M. Tang, O. Wong, S. W. Milliner and Y. P. P Chen,"An evolutionary learning approach for adaptive negotiation agents," International Journal of Intelligent Systems, vol. 21, issue 01, pp. 41–72, 2006.
[40] C. Ordonez, "Association rule discovery with the train and test approach for heart disease prediction," IEEE Transaction on Information Technology. Biomed, vol. 10, issue 02, pp. 334–343, 2006.
[41] C. Ordonez and E. Omiecinski, "Discovering association rules based on image content" IEEE Advances in Digital Libraries Conference (ADL'99), pp. 38–49, 1999.
[42] C. Ordonez, C. A. Santana and L. Braal, "Discovering interesting association rules in medical data," ACM DMKD Workshop, pp. 78–85, 2000.
[43] M. J. Zaki, "Mining non-redundant association rules," Data Mining and Knowledge Discovery, 09, 223-248, 2004.
[44] J. Hipp, U. Güntzer and G. Nakhaeizadeh, "Algorithms for association rule mining - a general survey and comparison," ACM SIGKDD Explorations, Newsletter 2: 58, 2000, doi:10.1145/360402.360421
[45] S. Brin, R. Motwani, J. U. Ullman and S. Tsur, "Dynamic itemset counting and implication rules for market basket data," Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD 1997), Tucson, Arizona, USA, pp. 265-276, May 1997.
[46] Brin, S., Motwani, R., Ullman, J.D. and Tsur, "Dynamic itemset counting and implication rules for market basket data," Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD 1997), Tucson, Arizona, USA, pp. 255-264, May 1997.
[47] N. Sharma and HariOm, "Framework for early detection and prevention of oral cancer using data mining," International Journal of Advances in Engineering and Technology, vol. 4, issue 2, pp. 302-310, Sep. 2012
[48] I. H. Witten and E. Frank, Data Mining: Practical Machine Learning Tool and Techniques, Second Edition, Elsevier.