

Extracting Significant Patterns for Oral Cancer Detection Using Apriori Algorithm

Neha Sharma¹, Hari Om²

¹Padmashree Dr. D. Y. Patil Institute of Master of Computer Applications, Pune, India

²Indian School of Mines, Dhanbad, India

Email: nvsharma@rediffmail.com, hariom4india@gmail.com

Received 13 January 2014; revised 12 February 2014; accepted 10 March 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Presently, no effective tool exists for early diagnosis and treatment of oral cancer. Here, we describe an approach for cancer detection and prevention based on analysis using association rule mining. The data analyzed are pertaining to clinical symptoms, history of addiction, co-morbid condition and survivability of the cancer patients. The extracted rules are useful in taking clinical judgments and making right decisions related to the disease. The results shown here are promising and show the potential use of this approach toward eventual development of diagnostic assay and treatment with sufficient support and confidence suitable for detection of early-stage oral cancer.

Keywords

Data Mining; Association Rule Mining; Apriori; Oral Cancer; WEKA

1. Introduction

In this paper, we have adopted Fayyad *et al.*'s definition of knowledge discovery and data mining. Knowledge discovery is the “non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data” [1]. Data mining is one of the steps in the process of knowledge discovery, consists of applying data analysis and discovery (learning) algorithms that produce a particular enumeration of patterns (or models) over the data. [2] [3]. Data mining could also be characterized as the procedure of finding useful patterns or meaning in raw data which subsequently can be used to develop a predictive model [3]-[5]. It is variously been called as KDD (knowledge discovery in databases), knowledge discovery, knowledge extraction, information discovery, information harvesting, data archeology and data pattern processing [6]. Knowledge discovery involves the additional steps of target data set selection, data preprocessing, and data reduction (reducing the

number of variables), which occur prior to data mining. It also involves the additional steps of information interpretation and consolidation of the information extracted during the data mining process. These extracted patterns will provide useful knowledge to decision makers [7].

Application of data mining are many—it has been utilized seriously and also widely by marketers, for direct marketing and cross-selling or up-selling; by financial institutions, for credit scoring and fraud detection; by manufacturers, for quality control and maintenance scheduling and by retailers, for market segmentation and store layout [8]. Data mining is becoming increasingly popular, if not increasingly essential in healthcare industry as well. It is so because modern medicine and various healthcare transactions generate almost daily, huge amounts of heterogeneous data that is too complex and voluminous to be processed and analyzed by traditional methods. For example, medical data may contain SPECT images, signals like ECG, clinical information like temperature, cholesterol levels, etc., as well as the physician's interpretation. Those who deal with such data understand that there is a widening gap between data collection and data comprehension. Computerized techniques are needed to help humans address this problem [9]. Data mining provides the methodology and technology to transform these mounds of data into useful information for decision making, with the intention of offer valuable quality services at reasonable costs, which is a main concern envisage by the healthcare organizations (hospitals, medical centers). Data mining applications can incredibly profit all stake holders of the healthcare industry such as hospitals, clinics, physicians, and patients, for example, by identifying effective treatments and best practices.

This paper presents an application of data mining in early detection and prevention of oral cancer and discusses how the generated patterns can be effectively used by physicians. The World Health Organization's Global Burden of Disease statistics distinguished malignancy or cancer as the second largest global cause of death, after cardiovascular disease [10]. Cancer is the fastest growing segment of the disease burden; global cancer deaths are anticipated to increase from 7.1 million in 2002 to 11.5 million in 2030 [11]. Advances in prevention, diagnostics and treatment of cancer have contributed to the improved prognosis for cancer patients: one third of cancers are preventable and another third are curable through early detection and effective therapy [12].

The objective of this article is to explore relevant literature and present the same in Section 2, cover the information about oral cancer in Section 3, examine the data mining methodology and then appropriate algorithm to implement in Section 4. Experimental results are presented in Section 5 and finally Section 6 highlights the conclusions and offers some future directions. At the end, acknowledgement and references are mentioned.

2. Literature Review

Kaladhar *et al.* [13] predict oral cancer survivability using the CART, Random Forest, LMT, and Naïve Bayesian classification algorithms, which classify the cancer survival using 10 fold cross validation and training dataset. Among these algorithms, the Random Forest technique classifies more accurately the cancer survival dataset as compared to other methods. Singh *et al.* [14] have applied the apriori algorithm with transaction reduction on the data of cancer symptoms by considering five different types of cancer to find the symptoms that help the cancer to spread and also the cancer type that spreads faster. Srikant *et al.* [15] have considered the problem of integrating constraints in the form of boolean expression that appoint the presence or absence of items in rules. Nahar *et al.* [16] discuss the significant prevention factors for a particular type of cancer. To find out the prevention factors, they have first constructed a prevention factor dataset through an extensive literature review and then three association rule mining algorithms: Apriori, Predictive apriori, and Tertius algorithms have been applied on that data to discover most of the significant prevention factors against a specific type of cancer. Experimental results illustrate that the Apriori is the most useful association rule-mining algorithm for discovery of the prevention factors.

Swami *et al.* [17] discuss the multidimensional association rules and the model for smoking habits in order to take some preventive measures to reduce the various habits of smoking in youths. Milovic *et al.* [18] discuss the applicability of data mining in healthcare and explain how the patterns can be used by physicians to determine diagnoses, prognoses, and apply for patients in healthcare organizations. A detailed survey on various methods adopted by the researchers for identification and classification of oral cancer detection at an earlier stage has been given in [19]. Chuang *et al.* [20] consider DNA repair genes by choosing a single nucleotide polymorphisms (SNPs) dataset with 238 samples of oral cancer and control patients for disease prediction. They report that the performance of the holdout cross validation is much better than cross validation and the best classifica-

tion accuracy is 64.2%. Gadewal *et al.* [21] have enlarged the oral cancer gene database to 374 genes by adding 132 gene entries to enable fast retrieval of updated information.

3. Oral Cancer

Oral malignancy is a heterogeneous assembly of tumors rolling out from diverse parts of the oral cavity, with distinctive predisposing factors, prevalence, and treatment outcomes. Oral tumor is one of the ten most incessant diseases worldwide with a yearly occurrence of over 300,000 cases, of which 62% arise in developing nations [22] [23]. There is a huge contrast in the rate of oral tumor in diverse regions of the worlds. The age-adjusted rates of oral tumor differ from over 20 for every 100,000 population in India, to 10 for every 100,000 population in the U.S., and less than 2 for every 100,000 population in the Middle East [24] [25]. In comparison with the U.S. population, where oral cavity malignancy represents only about of 3% of malignancies, it accounts for over 30% of all growths in India. It has been estimated that 83,000 new oral cancer cases occur every year iIn India [26] [27]. The variation in incidence and pattern of oral cancer is due to regional differences in the prevalence of risk factors. But as oral cancer has well-defined risk factors, these may be modified—giving real hope for primary prevention.

The clinician's issue is separating malignant lesions from a nearly infinite amount of other poorly characterized, questionable, and crudely comprehended lesions that also occur in the oral cavity. Most oral lesions are benign, yet many have a manifestation that may be effectively befuddled with threatening lesions and some are now considered pre-malignant because they have been statistically correlated with subsequently cancerous changes [28]. On the other hand, some malignant lesions seen in an early stage may mistaken for a benign. Early carcinomas are presumably asymptotic and ensuing signs are regularly misjudged in light of the fact that they imitate numerous benevolent lesions and the distress is negligible. Professional consultation is thus often delayed, increasing the chance for local spread and regional metastases. Stress must be placed on gaining access to high risk individuals for periodic oral examinations and educational efforts to increase the skill of primary health care providers in recognizing this problem. Squamous cell carcinoma accounts for 90% of the total number of malignant oral lesions. Therefore, the problem of oral cancer is primarily that of pathogenesis, diagnosis, and management of squamous cell carcinoma originating from oral muscular surface [29]. The aim of this work is to apply the association rule mining on the data pertaining to clinical symptoms, history of addiction, co-morbid condition and survivability in order to evaluate the clinical features, diagnosis, and treatment of oral cancer patients.

4. Association Rule Mining

Data mining technique, association rule mining is applied to search the hidden relationships among the attributes. It identifies strong rules discovered in databases using different measures of interestingness. Thus, an association rule is a pattern that states when X occurs, Y occurs with certain probability. In this paper, we adopt the standard definition of association rules [30]-[36].

Apriori Algorithm

The apriori is a classic algorithm for frequent item set mining and association rule learning over the transactional databases [37]. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by a apriori can be used to determine association rules, which highlight general trends in the database [38]. Association rules mining using apriori algorithm uses a “bottom up” approach, breadth-first search and a hash tree structure to count the candidate item sets efficiently. A two-step apriori algorithm is explained with the help of flowchart as shown in **Figure 1** and the algorithm is mentioned below:

Apriori algorithm: Candidate Generation and Test Approach

Step 1: Initially, scan database (DB) once to get frequent 1-itemset.

Step 2: Generate length $(k + 1)$ candidate item sets from length k frequent item sets.

Step 3: Test candidates against DB.

Step 4: Terminate, if no frequent or candidate set can be Generated.

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.

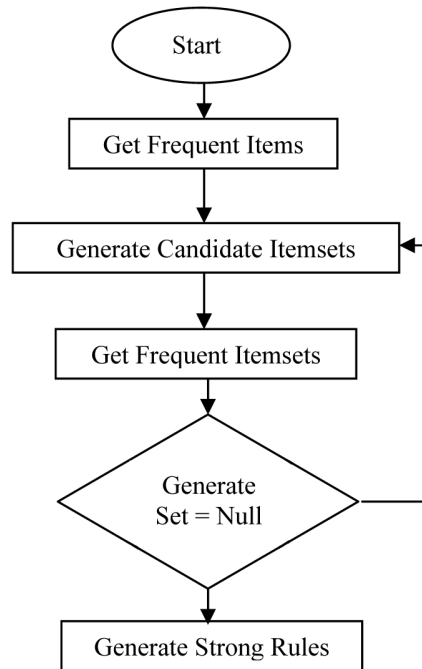


Figure 1. Flowchart of apriori algorithm.

Support: The rule holds with support $supp$ in T (the transaction data set) if $supp$ % of transactions contain $X \cup Y$ [39].

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

Confidence: The rule holds with confidence $conf$ in T if $conf$ % of transactions that contain X also contain Y [39] [40].

$$Conf(X \rightarrow Y) = P(Y|X) = Supp(X \cup Y) / Supp(X) = P(X \text{ and } Y) / P(X)$$

Lift: It is the probability of the observed support to that expected, if X and Y were independent [41].

$$Lift(X \rightarrow Y) = Supp(X \cup Y) / Supp(X) \times Supp(Y) = P(X \text{ and } Y) / P(X)P(Y)$$

Leverage: It measures the difference of X and Y appearing together in the dataset and what would be expected if X and Y were statistically dependent [42].

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

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) [43].

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

5. Experimental Results

The database for this work is created by collecting data through a retrospective chart review and the entire process is presented in [44]. There are total 33 variables and 1025 records of patients were created for the analysis. A data mining tool—WEKA 3.7.9 [45] has been used to explore the behaviour of the apriori algorithm for extracting the significant patterns for early detection of oral cancer. 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. Minimum support defined by the tool for the generated rule is 0.1 (103 instances) and minimum confidence is 0.9. Association rules for early detection of the oral cancer patients on the basis clinical symptoms, history of addiction and co-morbid condition are mentioned below and the same is presented in the graphical form in Figures 2-4:

Rule 1. Clinical-Symptom = Ulcer (577) ==> Survival = Dead (498) < conf: (0.98) > lift: (2.31) lev: (0.24) [246] conv:(28.27).

Rule 2. Clinical-Symptom = Burning-Sensation (287) ==> Survival = alive (286) <conf: (1) > lift: (2.27) lev: (0.16) [160] conv: (80.64).

Rule Details: Rule 1 and Rule 2 suggest ulcer as the clinical symptom may indicate oral cancer with more certainty in comparison to other clinical symptoms like burning sensation, loosening of tooth and mass, which subsequently lead to high mortality.

Rule 3. History-of-Addiction = Tobacco-Smoking and History-of-Addiction1 = Alcohol (131) ==> Survival = Dead (131) < conf: (1) > lift: (2.6) lev: (0.08) [80] conv: (80.64).

Rule Details: Rule 3 proposes that history of addiction like tobacco-smoking or tobacco-chewing and alcohol accounts for most oral cancers. Heavy smokers who use tobacco for a long time are most at risk. The risk is even

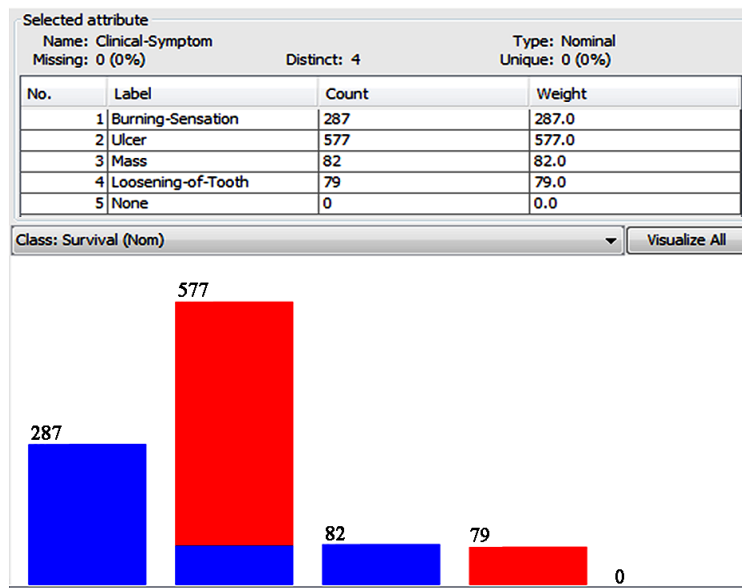


Figure 2. Clinical symptoms and survivability.

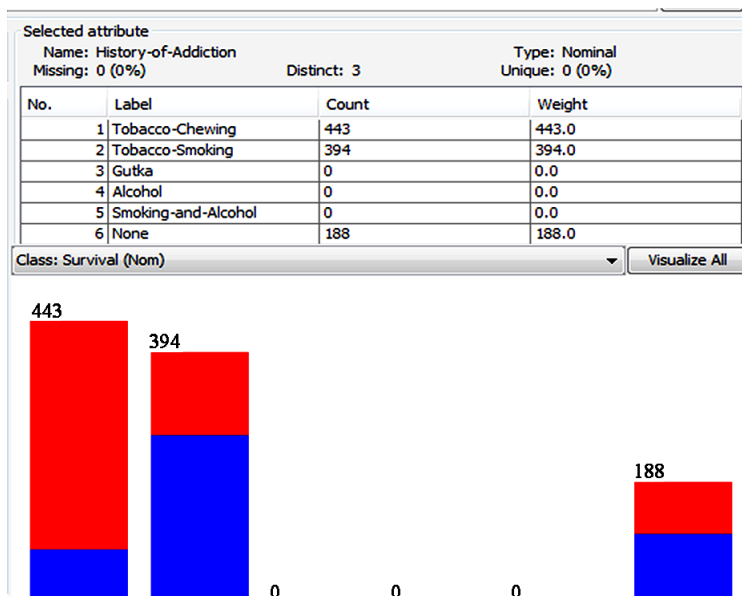


Figure 3. History of addiction and survivability.

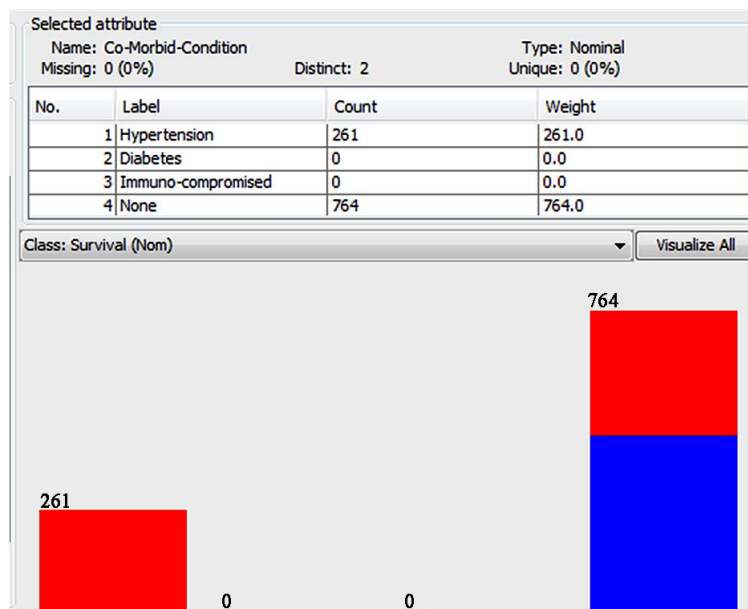


Figure 4. Co-morbid condition and survivability.

higher for tobacco users who drink alcohol heavily. In fact, three out of four oral cancers occur in people who use alcohol, tobacco, or both alcohol and tobacco.

Rule 4. Co-Morbid-Condition = Hypertension (261) \implies Survival = Dead (261) < conf: (1) > lift: (1.78) lev: (0.11) [114] conv: (114.33).

Rule Details: Rule 4 puts forward that co-morbid condition like hypertension may also be a reason for oral cancer and subsequently for high mortality.

Rule 5. Clinical-Symptom = Ulcer and History-of-Addiction = Tobacco-Smoking (131) \implies Survival = Dead (131) < conf: (1) > lift: (1.78) lev: (0.06) [57] conv: (57.38).

Rule Details: Rule 5 suggests that if clinical-Symptom is ulcer and history of addiction is tobacco-smoking, chances of patients suffering from oral cancer is high which may lead to mortality.

Rule 6. History-of-Addiction = Tobacco-Chewing and Co-Morbid-Condition = Hypertension (130) \implies Survival = Dead (130) < conf: (1) > lift: (1.78) lev: (0.06) [56] conv: (56.95).

Rule 7. History-of-Addiction = Tobacco-Smoking and Co-Morbid-Condition = Hypertension (131) \implies Survival = Dead (131) < conf: (1) > lift: (1.78) lev: (0.06) [57] conv: (57.38).

Rule Details: Rule 6 and Rule 7 hint that history of addiction like tobacco-smoking and tobacco-chewing along with co-morbid condition like hypertension increases the probability of oral cancer which may be the reason for high mortality.

The significant patterns generated using apriori algorithm can be summarized as follows:

If Clinical Symptoms = Ulcer.

If History of Addiction = Tobacco-Chewing/Smoking or Alcohol.

If Co-Morbid Condition = Hypertension.

Then,

Oral Cancer is suspected which has to be confirmed through biopsy and other diagnostic procedure.

6. Conclusion and Future Work

The data mining technique which is adopted for the research work is association rule mining. The algorithm used for implementing association rule mining is a popular apriori algorithm. Apriori has been used to extract the association among various valuable data pertaining to clinical symptoms, history of addiction and co-morbid condition. The rules generated would certainly assist the practitioners in early discovery of oral cancer and consequently help in prevention of the disease. The experimental results demonstrate that all the generated rules hold the highest confidence level, thereby, making them very useful for early detection and prevention of oral

cancer. In future, we intend to extend this research work by attempting to extract significant patterns and useful rules through the association rule mining algorithm using various attributes like predisposing factors, gross examination, tumor site, tumor size, neck node, etc. and use it more effectively for early detection and prevention of oral cancer.

Acknowledgements

The authors would like to thank Dr. Vijay Sharma, MS, ENT, for his valuable contribution in understanding the occurrence and diagnosis of Oral Cancer. The authors devote their sincere thanks to the management and staff of Indian School of Mines, for their constant support and motivation.

References

- [1] Fayyad, U. M., Piatetsky-Shapiro, G. and Smyth, P. (1996) Data Mining to Knowledge Discovery in Databases. *AI Magazine*, **17**, 37-54.
- [2] Han, J., Kamber, M. and Pei, J. (2011) Data Mining: Concepts and Techniques. 3rd Edition, Morgan Kaufmann Publishers, Burlington.
- [3] Khosla, R. and Dillon, T. (1997) Knowledge Discovery, Data Mining and Hybrid Systems. Engineering Intelligent Hybrid Multi-Agent Systems, Kluwer Academic Publishers, Norwell, 143-177.
- [4] Kincade, K. (1998) Data Mining: Digging for Healthcare Gold. *Insurance & Technology*, **23**, IM2-IM7.
- [5] Milley, A. (2000) Healthcare and Data Mining. *Health Management Technology*, **21**, 44-47.
- [6] Fayyad, U.M., Piatetsky-Shapiro, G. and Smyth, P. (1996) Data Mining to Knowledge Discovery: An Overview. Advances in Knowledge Discovery and Data Mining, AAAI Press/MIT Press, 1-36.
- [7] Houston, A.L., Chen, H., Hubbard, S.M., Schatz, B.R., Ng, T.D., Sewell, R.R. and Tolle, K.M. (1999) Medical Data Mining on the Internet: Research on a Cancer Information System. *Artificial Intelligence Review*, **13**, 437-466. <http://dx.doi.org/10.1023/A:1006548623067>
- [8] Koh, H.C. and Tan, G. (2005) Data Mining Applications in Healthcare. *Journal of Healthcare Information Management*, **19**, 64-72.
- [9] Cios, K.J. (2001) Medical Data Mining and Knowledge Discovery. *Studies in Fuzziness and Soft Computing*, **60**, 502.
- [10] Mathers, C.D. and Loncar, D. (2006) Projections of Global Mortality and Burden of Disease. *PLOS Medicine*, **3**, 2002-2030.
- [11] World Health Organization (2007) Department of Measurement and Health Information Systems: World Health Statistics 2007, World Health Organization, Geneva.
- [12] World Health Organization (2002) Department of Management of Noncommunicable Diseases: National Cancer Control Programmes, World Health Organization, Geneva.
- [13] Kaladhar, D.S.V.G.K., Chandana, B. and Kumar, P.B. (2011) Predicting Cancer Survivability Using Classification Algorithms. *International Journal of Research and Reviews in Computer Science (IJRRCS)*, **2**, 340-343.
- [14] Singh, S., Yadav, M. and Gupta, H. (2012) Finding the Chances and Prediction of Cancer through Apriori Algorithm with Transaction Reduction. *International Journal of Advanced Computer Research*, **2**, 23-28.
- [15] Srikant, R., Vu, Q. and Agrawal, R. (1997) Mining Association Rules with Item Constraints. *Proceeding KDD9*, New Port Beach, 14-17 August 1997, 67-73.
- [16] Nahar, J., Kevin, S.T., Ali, A.B.M.S. and Chen, Y.P. (2009) Significant Cancer Prevention Factor Extraction: An Association Rule Discovery Approach. *Journal of Medical Systems*, **35**, 353-367. DOI 10.1007/s10916-009-9372-8
- [17] Swami, S., Thakur, R.S. and Chandel, R.S. (2011) Multi-Dimensional Association Rules Extraction in Smoking Habits Database. *International Journal of Advanced Networking and Applications*, **3**, 1176-1179.
- [18] Milovic, B. and Milovic, M. (2012) Prediction and Decision Making in Health Care Using Data Mining. *International Journal of Public Health Science*, **1**, 69-78.
- [19] Anuradha, K. and Sankaranarayanan, K. (2012) Identification of Suspicious Regions to Detect Oral Cancers at an Earlier Stage—A Literature SURVEY. *International Journal of Advances in Engineering & Technology*, **3**, 84-91.
- [20] Chuang, L.Y., Wu, K.C., Chang, H.W. and Yang, C.H. (2011) Support Vector Machine-Based Prediction for Oral Cancer Using Four SNPs in DNA Repair Genes. *Proceedings of the International Multi Conference of Engineers and Computer Scientists*, Hong Kong, 16-18 March 2011, 16-18.
- [21] Gadewal, N.S. and Zingde, S.M. (2011) Database and Interaction Network of Genes Involved in Oral Cancer: Version II. *Bioinformatics*, **6**, 169-170. <http://dx.doi.org/10.6026/97320630006169>

- [22] Perkin, D.M. and Lara, E. (1980) Estimates of the World Wide Frequency of Sixteen Major Cancers. *International Journal of Cancer*, **41**, 184-197.
- [23] 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.
- [24] Sankaranarayan, R., Masuyer, E., Swaminathan, R., Ferley, J. and Whelan, S. (1998) Head and Neck Cancer: A Global Perspective on Epidemiology and Prognosis. *Anticancer Research*, **18**, 4779-4786.
- [25] Sankaranarayanan, R., Ramadas, K. and Thomas, K. (2005) Effect of Screening on Oral Cancer Mortality in Kerala, India: A Cluster-Randomised Controlled Trial. *The Lancet*, **365**, 1927-1933. [http://dx.doi.org/10.1016/S0140-6736\(05\)66658-5](http://dx.doi.org/10.1016/S0140-6736(05)66658-5)
- [26] Manoharan, N., Tyagi, B.B. and Raina, V. (2004) Cancer incidences in rural Delhi. *Asian Pacific Journal of Cancer Prevention*, **11**, 73-78.
- [27] Agrawal, M., Pandey, S., Jain, S. and Maitin, S. (2012) Oral Cancer Awareness of the General Public in Gorakhpur City, India. *Asian Pacific Journal of Cancer Prevention*, **13**, 5195-5199. <http://dx.doi.org/10.7314/APJCP.2012.13.10.5195>
- [28] American Cancer Society (1996) Cancer Facts and figures, Atlanta (GA), The Society.
- [29] Cancer Research Capign (1990) Oral Cancer. *Fact Sheet*, **14**.
- [30] Agrawal, R., Imielinski, T. and Swami, A. (1993) Mining Association Rules between Sets of Items in Large Databases. Proceedings of the 1993 *ACM SIGMOD International Conference on Management of Data*, Washington DC, May 1993, 207-216.
- [31] An, J., Chen, Y.P.P. and Chen, H. (2005) DDR: An Index Method for Large Time Series Datasets. *Information Systems*, **30**, 333-348. <http://dx.doi.org/10.1016/j.is.2004.05.001>
- [32] Chen, Y.P.P. and Chen, F. (2008) Targets for Drug Discovery Using Bioinformatics. *Expert Opinion on Therapeutic Targets*, **12**, 383-389. <http://dx.doi.org/10.1517/14728222.12.4.383>
- [33] Lau, R.Y.K., Tang, M., Wong, O., Milliner, S.W. and Chen, Y.P.P. (2006) An Evolutionary Learning Approach for Adaptive Negotiation Agents. *International Journal of Intelligent Systems*, **21**, 41-72. <http://dx.doi.org/10.1002/int.20120>
- [34] Ordonez, C. (2006) Association Rule Discovery with the Train and Test Approach for Heart Disease Prediction. *IEEE Transaction on Information Technology. Biomed*, **10**, 334-343. <http://dx.doi.org/10.1109/TITB.2006.864475>
- [35] Ordonez, C. and Omiecinski, E. (1999) Discovering Association Rules Based on Image Content. *IEEE Advances in Digital Libraries Conference (ADL'99)*, Baltimore, 19-21 May 1999, 38-49.
- [36] Ordonez, C., Santana, C.A. and Braal, L. (2000) Discovering Interesting Association Rules in Medical Data. *ACM DMKD Workshop*, Dallas, 14 May 2000, 78-85.
- [37] Agrawal, R. and Srikant, R. (1994) Fast Algorithms for Mining Association Rules in Large Databases. *Proceedings of the 20th International Conference on Very Large Data Bases, VLDB*, Santiago de Chile, 12-15 September 1994, 487-499.
- [38] Zaki, M.J. (2004) Mining Non-Redundant Association Rules. *Data Mining and Knowledge Discovery*, **9**, 223-248. <http://dx.doi.org/10.1023/B:DAMI.0000040429.96086.c7>
- [39] Agrawal, R., Imielinski, T. and Swami, A. (1993) Mining Association Rules between Sets of Items in Large Databases. *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, Washington DC, 26-28 May 1993, 207-216.
- [40] Hipp, J., Güntzer, U. and Nakhaeizadeh, G. (2000) Algorithms for Association Rule Mining—A General Survey and Comparison. *ACM SIGKDD Explorations Newsletter*, **2**, 58-64. <http://dx.doi.org/10.1145/360402.360421>
- [41] Brin, S., Motwani, R., Ullman, J.D. and Tsur, S. (1997) Dynamic Itemset Counting and Implication Rules for Market Basket Data. *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD 1997)*, Tucson, 13-15 May 1997, 255-264.
- [42] Piatetsky-Shapiro, G. (1991) Discovery, Analysis, and Presentation of Strong Rules. *Knowledge Discovery in Databases*. AAAI/MIT Press, Cambridge, 248, 255-264.
- [43] Brin, S., Motwani, R., Ullman, J.D. and Tsur, S. (1997) Dynamic Itemset Counting and Implication Rules for Market Basket Data. *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD 1997)*, Tucson, 13-15 May 1997, 265-276.
- [44] Sharma, N. and Om, H. (2012) Framework for Early Detection and Prevention of Oral Cancer Using Data Mining. *International Journal of Advances in Engineering & Technology*, **4**, 302-310.
- [45] Witten, I.H. and Frank, E. (2005) *Data Mining: Practical Machine Learning Tool and Techniques*. 2nd Edition, Elsevier, Amsterdam.