FRAMEWORK FOR EARLY DETECTION AND PREVENTION OF ORAL CANCER USING DATA MINING

Neha Sharma[1] and Hari Om[2]
[1] Dr. D.Y. Patil Institute of Master of Computer Applications, University of Pune, India
[2]Computer Science and Engineering Department, Indian School of Mines, Dhanbad, India

ABSTRACT

India has the dubious distinction of harbouring the world's largest number of oral cancer patient with an annual age standardized incidence of 12.5 per 100,000. The treatment is successful only if the lesion is diagnosed early. This paper proposes ED&P framework which is used to develop a data mining model for Early Detection and Prevention of malignancy of Oral Cavity. The database of 1025 patients has been created and the required information stored in the form of 36 attributes.

KEYWORDS: *Oral Cancer, Data Mining, Early Detection, Prevention, Lesion, Malignancy*

I. INTRODUCTION

Globally, about, 5,75,000 new cases and 3,20,000 deaths occur every year from oral cancer [1]. Head and neck cancer is the sixth common malignancy and is the major cause of cancer morbidity and mortality worldwide. In India, cancers of Head and Neck comprise ~24.1 % of total cancers seen at Tata Memorial Centre, Mumbai and of these ~13.2 % are from the oral cavity [2]. Although the Oral Cavity lesions constitute only a small minority of pathological conditions, they are of great significance, as they have potential to jeopardize the health and longevity of the patient [3]. Most oral cancers in India present in advanced stage of malignancy. It is essential to establish an accurate diagnosis, to initiate optimal therapy for oral cavity lesions [4]. One of the main barriers to treatment and control of oral cancer is the identification and risk assessment of early disease in the community in a cost-effective fashion. The Oral healthcare industry generates a great deal of information regarding oral disease data that can be collected and "analysed or mined" to determine the hidden patterns and trends that will help healthcare practitioners for effective decision making.

Data Mining & Statistical Analysis is the search for valuable information in large volumes of data. It is now widely used in health care industry. Finding useful patterns or meaning in raw data has variously been called KDD (knowledge discovery in databases), data mining, knowledge discovery, knowledge extraction, information discovery, information harvesting, data archaeology and data pattern processing [5]. In this paper, we will use Fayyad et al.'s (1996b) definitions 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" [6]. Data mining, 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." Data mining is typically a bottom-up knowledge engineering strategy [7]. Knowledge discovery involves the additional steps of target data set selection, data pre-processing, 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. The medical data is difficult to analyse because of its characteristics such as huge volume and heterogeneity, temporality of the data, and high frequency of null values.

The objective of this study is to propose the ED&P framework which is used to develop a data mining model for Early Detection and Prevention of malignancy of Oral Cavity. This paper is organized as follows. The next section reviews related work. Section 3 covers the information about malignancy in oral cavity. Section 4 gives the methodology used to collect the data from various sources and prepare it for mining process. Framework to develop a model to explore the hidden pattern is covered in section 5. Conclusion and future work are given in the section 6 and Acknowledge in the last section.

II. RELATED WORK

Nikhil Sureshkumar Gadewal et. al. [8] have compiled and enlarged the oral cancer gene database to include 374 genes by adding 132 gene entries to enable fast retrieval of updated information.

Jin Oh Kang et. al. [9] have developed four data mining models, including two artificial neural network (ANN) models and two classification and regression tree (CART) models, to predict both the total amount of hospital charges and the amount paid by the insurance of cancer patients and compared their efficacies. The ANN models showed better prediction accuracy than CART models. However, the CART models, which serve different information from ANN model, can be used to allocate limited medical resources effectively and efficiently. For the purpose of establishing medical policies and strategies, using those models together is warranted.

Adbelghani Bellaachia et. al. [10] have presented an analysis of the prediction of survivability rate of breast cancer patient. They have investigated three data mining techniques: The Naïve Bayes, the back propagated neural network, and the C4.5 decision tree algorithms, followed by Cross-Validation with 10 folds. Finally, they concluded that C4.5 algorithm has a much better performance than other two techniques.

Delen et. al. [11] has employed the SEER dataset and compared a few models in their study. Delen et al. reported that the Decision Trees algorithm had a much better performance than the other two algorithms, Artificial Neural Network and Logistic Regression model. In their study, the Logistic Regression model had the worst accuracy. In Delen et al.'s study, 46% survival and 54% death patients were occupied in their dataset. Decision Trees showed the best performance for accuracy.

Arihito Endo et. al. [12] present optimal models to predict the survival rate of breast cancer patients in five years. In this study, Logistic Regression model showed the highest accuracy. The J48 had the highest sensitivity and the ANN had the highest specificity. The Decision Trees models tend to show high sensitivity. And the Bayesian models were apt to show the accuracy going up. They found that the optimal algorithm might be different by the predicted objects and dataset.

Shantakumar B. Patil and Dr. Y.S. Kumaraswamy [13] used the K-means clustering algorithm and MAFIA algorithm on Heart Disease Data Warehouse to mine the frequent patterns relevant to heart disease. Then the significant weightage of the frequent patterns are calculated. Further, the patterns significant to heart attack prediction are chosen based on the calculated significant weightage. These significant patterns can be used in the development of heart attack prediction system.

Shantakumar B. Patil and Dr. Y.S. Kumaraswamy [14] did the extension of their previous work by training the neural network with the selected significant patterns in order to predict heart attack in an efficient manner. They have employed the Multi-layer Perceptron neural network for the design of prediction system with Back-propagation as training algorithm. The experimental results have illustrated the efficacy of the designed prediction system in predicting the heart attack.

DSVGK Kaladhar et.al. [15] have predicted oral cancer survivability using Classification algorithms like CART, Random Forest, LMT, and Naïve Bayesian. The algorithms classify the cancer survival using 10 fold cross validation and training data set. The Random Forest technique correctly classified the cancer survival data set. The absolute relative error is less when compared to other methods.

III. ORAL CANCER

Oral cancer is a subtype of head and neck cancer and is any cancerous growth located in any subsites of the oral cavity [16]. 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 [17]. 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).
Prognosis of oral cancer depends on early diagnosis. Despite advanced surgical techniques and other treatment modalities, the 5-year survival rate remains ~40-50% [18][19]. 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 [20].
The symptoms for an oral cancer at an earlier stage [21] 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 [22]. 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 [23].

IV. RESEARCH METHODOLOGY

The proposed framework portrays the methodology for developing a model for Early Detection and Prediction of Oral Cancer using data mining. The framework is referred as ED&P framework as shown in figure1. The following section describes in detail the methodology that was adopted for the creation of database and a framework that will be adopted to develop a model for early detection and prevention of oral cancer.

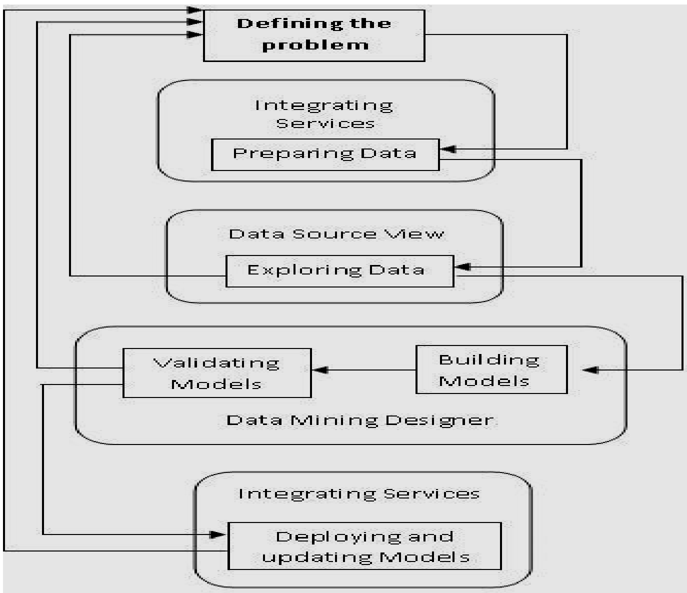

Figure 1. ED&D Framework

4.1. Reason to create the database

So far no proper epidemiological data on oral cancer disease is available in India. Information currently available is mostly on the basis of crude incidence rate available from three metropolitan cities covered under National cancer registry project [24].

4.2. Data Collection

Data collection is done by a retrospective chart review of data 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. A total of 1025 records of patients were used for the analysis.

4.3. Data Preparation

Data preparation is the process to consolidate and clean the data that was identified and collected in the previous step. Data was scattered across multiple hospitals and stored in different formats or contained inconsistencies such as incorrect or missing entries. Data preparation is not just about removing bad data, but about finding hidden correlations in the data, identifying sources of data that are the most accurate, and determining which columns are the most appropriate for use in analysis. Collection of data from various sources followed by appropriate correction and refinement of the data to make it appropriate for analysis was carried out is shown in Figure 2.

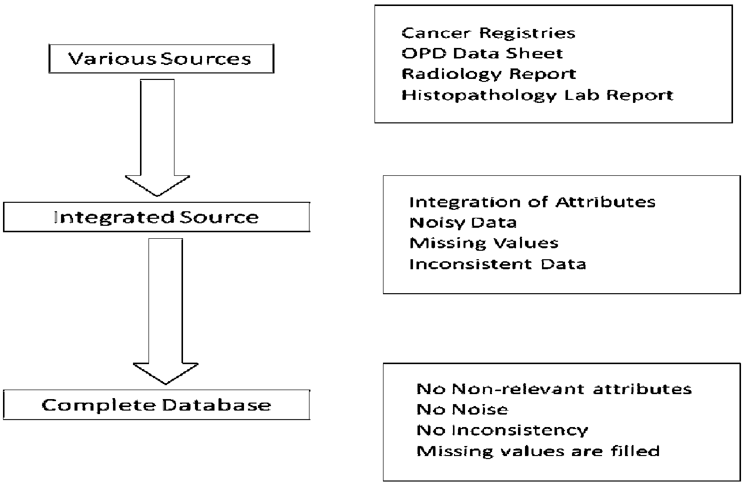

Figure 2. Process of Data Collection and Data Preparation

a) Data Integration

Data integration is the process of merging of data from multiple data sources into one coherent data store so that the encoding structure, unit of measurement, physical storage structure, naming convention and datatype formats of attributes are same. For example patient_id in one source must be matched up with patient_number in another source. Also, the attributes in the final database was not available at one source and are integrated from various sources. Careful integration of the data from multiple sources helped in creating the database that has all the attributes required to carry out the research work. Integration also helped in reducing and avoiding redundancies and inconsistencies in the resulting data set. This will help in improving the accuracy and speed of the subsequent mining process.

The final database includes all the fields that are required for the research, pertaining to Oral Cancer from the records of the Cancer Registries of Tertiary Care Hospitals, OPD data sheet, radiology report and histopathology lab report of the hospitals. Initially 40 attributes were identified, out of which 4 attributes (Name, Address, Date of Presentation, Occupation) were discarded as the attributes were not relevant to the research in hand. Finally, the dataset is prepared with 36 attributes. There are only two Continuous or Numerical Variables, which are CaseId and Age. Remaining 34 variables are Discrete or Categorical variables. [Table 1] shows all the discrete variable along with its domain, i.e. the various values it can hold.

Table 1. Discrete or Categorical Variables and its Domain

No	Variable	Domain
1	Sex	Male, Female
2	Socioeconomic Status	Good, Average, Poor, Above Average, Below Average
3	Clinical Symptoms	Burning Sensation, Ulcer, Loosening of tooth, None
4	History of Addiction	Tobacco-Chewing, Tobacco-Smoking, Gutka, Alcohol, Smoking and Alcohol, None
5	History of Addiction1	Tobacco-Chewing, Tobacco-Smoking, Gutka, Alcohol, Smoking and Alcohol, None
6	Co-Morbid Condition	Hypertension, Diabetes, Immuno-compromised, None
7	Co-Morbid Condition1	Hypertension, Diabetes, Immuno-compromised, None
8	Gross Examination	Ulcero-proliferative, Infiltrative, Verrucous, Plaque Like, Polypoidal
9	Site	BM, LA, RMT, LIP, Tongue, UA, Palate
10	Predisposing factor	Leukoplakia, Erythoplakia, Submucous Fibrosis, Linchen Planus, None
11	Tumour Size	<2cm, 2 cm to 4 cm, >4 cm
12	Neck Node	Present, Absent
13	LFT	Normal, Deranged
14	FNAC of Neck Node	Yes, No
15	Diagnostic Biopsy	Squamous Cell Carcinoma, Variant of SCC, Benign
16	USG	Yes, No
17	CT Scan / MRI	Bony Involvement, Normal
18	Diagnosis	SCC, Verrucous, Benign, Plaque Like, Sarcomatoid, Acantholytic, Adenoca, Lymphoepithelioma Like
19	Staging	I, II, III, IV
20	Surgery	Y, N
21	Radiotherapy	Y, N
22	Chemotherapy	Y, N
23	Histopathology	Variant of SSC-Verrucous, Adenoca, Basaloid, Plaque Like, Sarcomatoid, acantholytic, Lymphoepithelioma like
24	1st Followup Symptoms	Difficulty in Swallowing, Swelling
25	1st Followup Examination	Ulcero-proliferative, Infiltrative, Verrucous, Plaque Like, Polypoidal
26	2nd Followup Symptoms	Difficulty in Swallowing, Swelling
27	2nd Followup Examination	Ulcero-proliferative, Infiltrative, Verrucous, Plaque Like, Polypoidal
28	3rd Followup Symptoms	Difficulty in Swallowing, Swelling
29	3rd Followup Examination	Ulcero-proliferative, Infiltrative, Verrucous, Plaque Like, Polypoidal
30	4th Followup Symptoms	Difficulty in Swallowing, Swelling
31	4th Followup Examination	Ulcero-proliferative, Infiltrative, Verrucous, Plaque Like, Polypoidal
32	5th Followup Symptoms	Difficulty in Swallowing, Swelling
33	5th Followup Examination	Ulcero-proliferative, Infiltrative, Verrucous, Plaque Like, Polypoidal
34	Survival	Y, N

b) Data Cleaning

Real world data tends to be incomplete, noisy and inconsistent. Data Cleaning routine attempt to fill in the missing values, smooth out noise while identifying outliers, and correct inconsistencies in the data.

i) Missing Values

Various strategies were applied to generate the missing values depending on the importance of the missing value and its relation to the search domain. Either fill in the missing value manually or use a global constant to fill in the missing value.

Table 2. Missing Values

Diagnosis	Staging	Surgery	Radiotherapy	Chemotherapy	Histopathology
SCC	IV	Y	Y	Y	SCC
Verrucous	II	Y			Verrucous
Benign	No	Y	N	N	Schwanoma
Benign	No	Y	N	N	
SCC	IV	Y	Y	Y	SCC

ii) Noisy Data

Noise is a random error or variance in a measured variable. Many techniques were used to remove the noise and smooth out the data. Outliers were detected by clustering, where similar values are organized into groups, or clusters, values that fall outside of the set of clusters may be considered outliers. Combined approach of computer and human inspection- Using clustering techniques and constructing groups of data sets, human can then sort through the patterns in the list to identify the actual garbage ones. This is much faster than manually search through the entire database.

Table 3: Outliers

Tumor Size (in cm)	Neck Nodes	LFT	FNAC of Neck Node	Diagnostic Biopsy	USG	CTScan/MRI	Diagnosis
> 4	Abs	Normal	No	Benign	No	Normal	Benign
< 2	Abs	Normal	No	Benign	No	Normal	Benign
> 4	Abs	Normal	No	Benign	No	Normal	Verrucous
< 2	Abs	Normal	No	Benign	No	Normal	Benign
> 4	Abs	Normal	No	Benign	No	Normal	Benign

iii) Inconsistent Data

There may be inconsistencies in the data recorded for some transactions. Some data inconsistency may be corrected manually using external references. Other forms of inconsistency are due to data integration, where a given attribute may have different names in different databases or the same data value is represented by different names.

Table 4: Data Inconsistency (Tobacco-Smoking and Smoking represent the same value)

Clinical Symptom	History of Addiction	History of Addiction1	Co Morbid Condition	Co Morbid Condition1	Gross Examination
Burning Sensation	Smoking	Alcohol	None	None	Plaque-Like
Burning Sensation	Tobacco-Smoking	Alcohol	None	None	Plaque-Like
Burning Sensation	None	None	None	None	Polypoidal
Burning Sensation	Smoking	Alcohol	None	None	Plaque-Like
Mass	None	None	None	None	Polypoidal
Burning Sensation	Tobacco-Smoking	Alcohol	None	None	Plaque-Like

4.4 Exploring Data

Exploration techniques include calculating the minimum and maximum values, calculating mean and standard deviations, and looking at the distribution of the data. The technique ensures that the data is absolutely balanced and would lead to stable and accurate result.

a) Selection of the data-mining task

Data Mining tasks chosen are Association rule and Classification. In data mining, association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. Piatetsky-Shapiro[25] describes analyzing and presenting strong rules discovered in databases using different measures of interestingness learning. Classification is the task of generalizing known structure to apply to new data. It is the problem of identifying which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known. The individual observations are analyzed into a set of quantifiable properties, known as various explanatory variables, features, etc.

b) Selection of the data-mining algorithm

A data mining algorithm is a set of heuristics and calculations that creates a data mining model from data. Choosing the best algorithm to use for a specific analytical task is challenge. While we can use different algorithms to perform the same task, each algorithm produces a different result, and some algorithms can produce more than one type of result. For example, we can use the Decision Trees algorithm not only for prediction, but also as a way to reduce the number of columns in a dataset, because the decision tree can identify columns that do not affect the final mining model.
We intend to apply algorithms like Decision Trees Algorithm, Naive Bayes Algorithm, Clustering Algorithm, Neural Network Algorithm, Association Algorithm.

4.5 Build a Model

To create a model, the algorithm first analyzes the data we provide, looking for specific types of patterns or trends. The algorithm uses the results of this analysis to define the optimal parameters for creating the mining model. These parameters are then applied across the entire data. Typically multiple models are created with different configurations and all the models are tested to see which yields the best results for the research problem and data collected.

4.6 Validate the Model

It is important to validate the performance of mining models before deploying them into a production environment. The data would be separated into training and testing sets to test the accuracy of predictions. There are many validation approaches like lift and gain, cross-validation, classification matrices, scatter plots and profit charts for assessing the quality and characteristics of a data mining model.

4.7 Deploying and Updating Models

Model will to be deployed so that it can perform many tasks depending on need. However, Prediction, Detection and Prevention process has to be upgraded so that they become the next set of inputs to the system. Each time the system will generate the refined and accurate outputs.

V. CONCLUSION AND FUTURE WORK

An ED&P Framework which is an information system is presented that will deliver the necessary information to clinical, administrative, and policy researchers and analysts in an effective and efficient manner. The system will deliver the technology and knowledge that users need to readily: (1) organize relevant data, (2) detect cancer patterns (3) formulate models that explain the patterns, and (4) evaluate the efficacy of specified treatments and interventions with the formulations. In this paper we have discussed how the database of 1025 cases with 36 attributes has been created and prepared for data mining task. Our future work shall involve applying data mining algorithms using the data set and identifying the useful patterns.

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] Langdon JD, Russel RC, Williams NS, Bulstrode CJK Arnold, Oral and Oropharyngeal cancer practice of surgery, London: Hodder Headline Group; 2000.

[2] K.A. Dinshaw and B. Ganesh, Annual Report 2002-2005, Hospital based cancer registry, Tata Memorial Hospital, 2008

[3] Willium GS, Hine MK, Levy BM, A text book of oral Pathol. 4th ed. Philadelphia: W.B. Saunders Company

[4] Warren S, Biopsy in relation to tumour diagnosis, 4th ed. Boston: American Cancer Society, Mass Deviation; 1968.

[5] Fayyad, U. M., Piatetsky-Shapiro, G. & Smyth, P. (1996a). From Data Mining to Knowledge Discovery: An Overview. In Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. & Uthurusamy, R. (eds.) Advances in Knowledge Discovery and Data Mining, 1–36. AAAI Press/MIT Press

[6] Fayyad, U. M., Piatetsky-Shapiro, G. & Smyth, P. (1996b). From Data Mining to Knowledge Discovery in Databases. AI Magazine 17(3): 37–54

[7] Khosla, R. & Dillon, T. (1997). Knowledge Discovery, Data Mining and Hybrid Systems. In Engineering Intelligent Hybrid Multi-Agent Systems, 143–177. Kluwer Academic Publishers.

[8] Nikhil Sureshkumar Gadewal, Surekha Mahesh Zingde, Database and interaction network of genes involved in oral cancer: Version II, Bioinformation 6(4): 169-170 (2011)

[9] Jin Oh Kang, Suk-Hoon Chung, Yong-Moo Suh et. al. Prediction of Hospital Charges for Cancer Patients with Data Mining Technique. J Kor Soc Med Informatics 2009;15(1):13-23

[10] Bellaachia and Erhan Guven, Predicting Breast Cancer Survivability using Data Mining Techniques. Ninth Workshop on Mining Scientific and Engineering Datasets in conjunction with the Sixth SIAM International Conference on Data Mining (SDM 2006), Saturday, April 22, 2006.

[11] Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. Artificial Intelligence in Medicine. 2005 Jun; 34(2):113-27.

[12] Arihito Endo et. al. [15] Comparison of Seven Algorithms to Predict Breast Cancer Survival. Biomedical Soft Computing and Human Sciences 2008 ; 13(2) : 11- 16

[13] Shantakumar B. Patil and Dr. Y.S. Kumaraswamy. Extraction of Significant Patterns from Heart Disease Warehouse for Heart Attack Prediction. IJCSNS International Journal of Computer Science and Network Security February 2009 ; VOL.9 No.2

[14] Shantakumar B. Patil and Dr. Y.S. Kumaraswamy [2009] Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network, European Journal of Scientific Research 2009; ISSN 1450-216X Vol.31 No.4 : 642-656

[15] DSVGK Kaladhar, B. Chandana and P. Bharath Kumar, “Predicting cancer survivability using Classification algorithms”, International Journal of Research and Reviews in Computer Science (IJRRCS) Vol.2, No.2, pp 340 – 343, April 2011.

[16] Werning, John W (May 16, 2007). Oral cancer: diagnosis, management, and rehabilitation. p. 1. ISBN 978-1588903099.

[17] http://www.oralcancerfoundation.org/facts/index.htm

[18] Jemal A, Thimas A, Murray T, Thun M. Cancer statistics, 2002. CA Cancer J Clin 2002;52:181-2.

[19] Woolgar JA, Scott J, Vaughan ED, Brown JS, West CR, Rogers S. Survival, metastasis and recurrence of oral cancer in relation to pathological features. Ann R Coll Surg Engl 1995;77:325-31.

[20] 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

[21] 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, No.4, pp 199 – 209, August 2008.

[22] www.yourtotalhealth.ivillage.com

[23] www.oralcancerawareness.org

[24] Mehta F.S.; Hamner J.E.:Tobacco related oral mucosal lesions and conditions in India.Tata Inst. of Fundamental Research.1993

[25] Piatetsky-Shapiro, Gregory (1991), Discovery, analysis, and presentation of strong rules, in Piatetsky-Shapiro, Gregory; and Frawley, William J.; eds., Knowledge Discovery in Databases, AAAI/MIT Press, Cambridge, MA.

Authors

Neha Sharma, completed her MCA from College of Engineering, Bhubaneshwar under Orissa University of Agriculture and Technology in the year 1999. She is currently pursuing her Ph.D. (Computer Science) at Indian School of Mines, Dhanbad. She has presented 5 papers in International Conference which are available on ACM Digital Library and IEEE Explore. She has also published 3 papers in International Journals. She has got 9 years of teaching experience and 4 years of industrial experience. Her area of interest includes Data Mining, Database Design, Analysis and Design, Software Engineering and Metadata.

Presently, **Hari Om**, is working as Assistant Professor in the department of Computer Science & Engineering at Indian School of Mines, Dhanbad, India. He earned his Ph.D. in Computer Science from Jawaharlal Nehru University, New Delhi, India. He has published several research papers in International and National Journals including various Trans. of IEEE, and presented many at International and National conferences of high repute. His research areas are Video-on-Demand, Cryptography, Network security, Data Mining, Image Processing, and Signal Processing.