A New Feature Selection Method for Oral Cancer Using Data Mining Techniques

Hemanth Kumar A M*, Manasa M, Rakshitha B H, Sanjay R, Sneha C R

Department of CSE, ATMECE, Mysuru, Karnataka, India

ABSTRACT

The word cancer is used basically for more than 1000 different diseases including malignant tumours of different sites. Common to all forms of the disease is the failure of the mechanisms that regulate normal cell growth, explosion and cell death. Ultimately, there is evolution of the resulting tumour from mild to severe abnormality, with incursion of adjoining tissues and, ultimately, spread to other areas of the body. The primary risk factor for evolving oral cancer is tobacco use. Smoking cigarettes, cigars, and pipes all increase risk of oral cancer. Smokeless tobacco, also called "dip" or "chew," also enhance the risk. Alcohol consumption is another habit that is strongly associated with the growth of oral cancer. This paper uses data mining technology such as classification and prediction to identify oral cancer. Apriori algorithm is the innovation algorithm of Boolean association rules of mining frequent item sets. The datamining methods and techniques will be discovered to identify the suitable methods and techniques for efficient classification of data. The data mining techniques are effectively used to extract meaningful relationships from the data. Genetic algorithm were applied to association and classification techniques.

Keywords. cancer, oral, genetic algorithm, data mining, apriori algorithm

I. INTRODUCTION

Worthy dental care is significant to maintaining healthy teeth, gums and mouth. Dental problems, with bad breath, dry mouth, tooth decay, or thrush are all treatable with proper treatment and care. Oral cancer can affect any area of the oral cavity including the lips, gum tissues, tongue, cheek lining and the palate. Possible signs and symptoms of oral cancer when patients may report include. a lump or thickening in the oral soft tissues, throat soreness or feeling that something is trapped in the throat, struggle in mastication or consuming, ear pain, difficulty moving the jaw or tongue, roughness, numbness of the tongue or other areas of the mouth, or swelling of the jaw that causes dentures to fit inadequately or become Uncomfortable. Other symptoms of oral cancer may include.

1. a sore in your mouth or on your lip that does not heal

2. wound on the tongue or tonsil

3. white and red patches in the mouth or lips that does not heal

4. bleeding from the mouth

5. Difficulty swallowing, chewing, speaking, or moving the tongue

6.

A. THE RISK FACTORS FOR ORAL CANCERS ARE

Tobacco and Alcohol The major risk factors related with oral cancer are tobacco use, in any existing forms, and heavy alcohol drinking (people who drink

five to eight drinks per day). The mutual effects of alcohol andtobacco smoking have been shown to be synergistic. Of attention, a recent study showed that consumption is inversely associated with oral cancer. The threat for emerging oral cancer is five to nine times greater for smokers than non-smokers. The mechanisms of oral carcinogenesis is made by the tobacco smoking components. In specific the polymorphic inconsistency in the enzymes involved in biotransformation of tobacco-related pro-carcinogens plays an significant role in controlling oral cancer vulnerability. Alcohol drinking is associated to oral cancer because many organichazards derived from alcohol deprivation are digested into active forms that have toxic effects on entities. Ethanol isdissolved to acetaldehyde, a supposedcancer-causing agent.

Age and family history of cancer Age indicates a sequentialmodule in the chemical and biophysical procedures of cells that allow malignant change or the decrease of the resistant system ability. Exactly, the durablecontact to risk factors may affect the gene products that control epithelial cell propagation and death resulting in an unrestrained malignant propagation of cells. Also, family history of oral cancer plays an important role and is careful a risk factor. However, more studies are necessary to clarify which fragments and genes are responsible for oral cancer vulnerability in families. Family history of oral cancer is mostly related with an onset of the disease at an early age (about 45 years old). Oral cancer is also seen in family members without practices such as tobacco chewing, smoking or alcohol consumption.

HPV infection While the relationship between HPV infection and oropharyngeal cancer is now well recognized, it is still indistinct whether HPV infection may lead to oral cancer as well. Numerous studies propose an association between human papillomavirus (HPV) infection and oropharyngeal cancers, mainly HPV 16 (90-95% of HPV-positive tumors). In the US there has been a recent increasing

occurrence of cancer of the oropharynx due to persistent HPV infection, particularly among young white men.

The role of the protective effect of HPV vaccines against oropharyngeal cancer remains unclear. However, a recent randomized precise trial has shown that the occurrence of oral HPV four years after immunization is lower when compared to women who did not accept the vaccine. This advises that the injection may have possibly important insinuations for anticipation of progressively common HPV.

Other Risk Factors

I. Viral infections
II. Immuno-deficiencies
III. Poor nutrition
IV. Exposure to ultraviolet light

Signs and Symptoms of Oral Cancer

✓ Sores in the mouth or on the lips heal.
✓ A white or red patch of skin in the mouth.
✓ Swallowing and chewing is difficult.

Broadcast for oral cancer

Early detection of oral cancer using visual examination of the mouth stands being considered in countries where occurrence is high, such as Bangladesh, India, Pakistan, and Sri Lanka. The oral cavity is easily reachable for routine examination, and nonmedical recruits can willinglynoticewounds that are the ancestors of carcinoma. Furthermore, there remainwarnings that precursor wounds may retreat if tobacco use ceases, and that surgical treatment of early oral cancer is very actual.

Association Rule Mining

Data mining technique, association rule mining is applied to find the unknown relationships among the attributes. It identifies strong rules firm in databases using dissimilarmethods of interestingness. Thus, an

association rule is a pattern that states when X occurs, Y occurs with certain probability. Association rule mining follow on two main steps. The first step is to find all item sets with acceptable supports and the second step is to generate association rules by merging these common or enormous item-sets. In the outdated association rules mining, minimum support threshold and minimum confidence threshold values are presumed to be available for mining frequent item sets, which is hard to be set without specific knowledge; users have difficulties in setting the support threshold to obtain their required results. To use association rule mining without support threshold, another constraint such as similarity or confidence clipping is usually introduced

II. HISTORY OF ORAL CANCER

Traditionally, it has been difficult to determine which irregular tissues in the mouth are worthy of apprehension. The detail is, the regular person regularly has situations current in their mouths that impersonator the presence of pre-cancerous changes, and very early cancers of the soft tissues. One study determined that the average dentist sees 3-5 patients a day who show soft tissue irregularities, most of which are benign in nature. Even the humblest things, such as pathos ulcers, herpes simplex, herpes labialise, the wound left by accidentally biting the inside of your cheek, or sore spots from a poorly fitting prosthetic appliance or denture, all at first examination, share similarities with dangerous lesions. Some of these conditions cause physical discomfort, others are painless. There has been a propensity to watch these areas over an extended period to determine if they are unsafe or not. Inappropriately, this thinking leads to a state in which a dangerous lesion may continue to prosper and grow into a future stage, hard to cure cancer. Any sore, discoloration, induration, prominent tissue, irritation, hoarseness, **which does not resolve within a two week period on its own, with or without treatment**, should be considered suspect and worthy of further inspection or appointment. Besides a dull

visit to the dental office for regular examinations, it is the patient's responsibility to be aware of changes in their oral environment. When these changes occur, they need to be carried to the consideration of a capable dental expert for examination. The dental expert needs to be present in the knowledge base essential to make a proper analysis, and be knowledgeable in the proper broadcastmeasures to identify oral cancer.

III. EXISTING SYSTEM

Oral cancer is one of the major disease and it needs to be identified in the early stages. Disease detection is also one of the significant areas of research in medical. There is no automation for Oral Cancer Prediction. In existing we have many health maintenance tools and software. All these tools are used to store the data and retrieve the same data but no analysis is done. Today's medical field requires extraction of useful information from the analyzed data. Here we are concentrating on Oral Cancer Prediction using data mining techniques.

Problem Statement

Oral cancer is one of the major disease and it needs to be identified in the early stages. There is no automation for Oral Cancer Prediction and discovering the relationship between oral cancer and its attributes is tedious task and challenging.

IV. PROPOSED SYSTEM

To overcome the different drawbacks and issues mentioned in the existing system we develop an automation for the "oral cancer prediction" and finding the association among different attributes. Proposed system is an health sector application where the major objective is to predict the oral cancer disease in early stages. Proposed system makes use of data mining techniques "classification rules" for oral cancer prediction and uses "association rules" to discover the relationship between the oral cancer attributes.
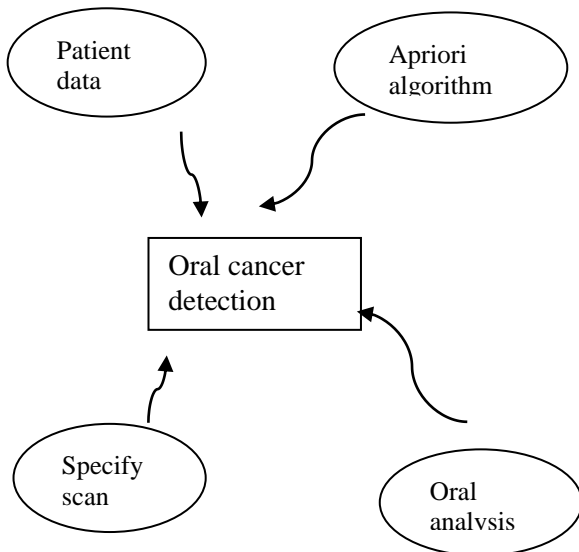
V. DATA FLOW DIAGRAM

Figure 2. dataflow diagram

VI. METHODOLOGY AND IMPLEMENTATION

Our paper mainly concentrates on two algorithm
1   Genetic algorithm
2   Apriori algorithm

1 Genetic Algorithm

These are search algorithms based on the mechanics of natural genetics. They combine a "Darwinian survival of the fittest" approach with a organized, yet randomized, information discussion. The advantage is that they can search complex and large amount of spaces efficiently and locate near best results quickly. The algorithm operates through a simple cycle.

1 Formation of a people of strings
2. Assessment of each string
3. Collection of the finest strings

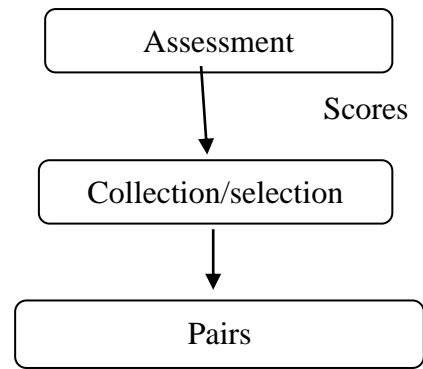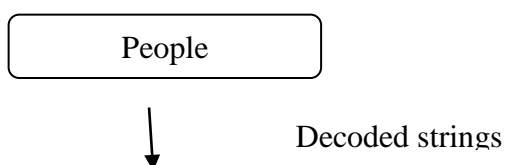The GA maps strings of numbers to each possible solution.

Figure 3. Genetic algorithm flow chart

2 Apriori Algorithm

The apriori is a typical algorithm for common item set mining and association rule learning over the transactional databases. It continues by identifying the frequent individual items in the database and spreading them to larger and larger item sets as long as those item sets appear adequately often in the database. The frequent item sets firm by a apriori can be used to control association rules, which highlight general tendency in the database. Association rules mining using apriori algorithm uses a "bottom up" method, breadth-first search and a hash tree structure to count the applicant item sets efficiently.

Apriori algorithm. Applicant Group and Test Approach

Step 1. Firstly, scan the database once to get common itemset.

Step 2. Make length (k + 1) applicant item sets from length k common item sets.

Step 3. Test applicants against DB.

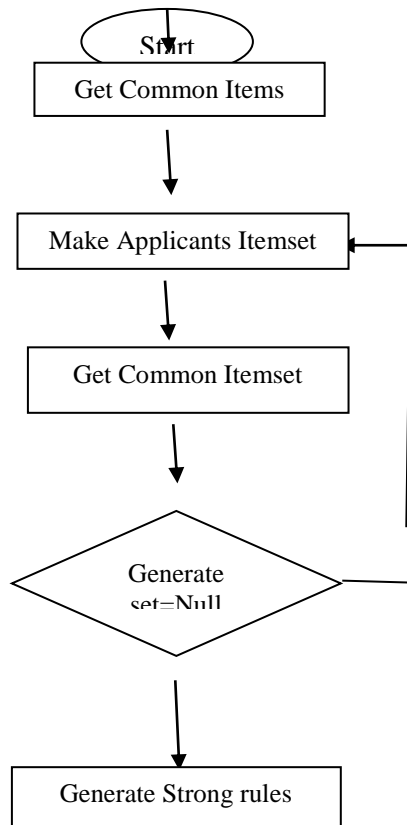Step 4. Terminate, if no common or applicant set can be generated.

Figure 4. Apriori algorithm flow chart

VII. CONCLUSION AND FUTURE WORK

Data mining is the non-trivial abstraction of hidden before unidentified and possibly useful information about data. Here, we use data mining techniques in oral cancer treatment. Data mining techniques have been broadly used for oral cancer analysis. In this paper we have deliberated the genuine methods that can be used for oral cancer classification.In future, we aim tospread this research by endeavoring to extractmomentousdesigns and useful rules through theassociation rule mining algorithm from extractingmost actual course of action.

VIII. REFERENCES

1. K Anuradha, Dr. K. Sankaranarayanan" International Journal of Advanced Research in Computer Science and Software Engineering"Volume 5, Issue 1, January 2015.

2. International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 10, October 2013.

3. Neha Sharma, Hari Om-" Extracting Significant Patterns for Oral Cancer Detection Using Apriori Algorithm".

4. S Warnakulasuriya, "Global epidemiology of oral and oropharyngealcancer", April-May 2009. [5] Hipp, Jochen ,Guntzer, Ullrich and Nakhaeizadeh, Gholamreza, "Algorithms for Association Rule Mining – A general Survey and Comparison". SIGKDD explorations, Vol 2, Issue – 1, pp 58 – 63, Mar – 2004.

5. Nikhil SureshkumarGadewal, Surekha Mahesh Zingde, "Database andinteraction network of genes involved in oral cancer", (2011).

6. Kaladhar, D.S.V.G.K., Chandana, B. and Kumar, P.B. (2011) Predicting Cancer Survivability Using ClassificationAlgorithms. International Journal of Research and Reviews in Computer Science (IJRRCS), 2, 340-343.

7. 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.

8. www.cancer.org/clinicaltrials.

9. http.//www.scribd.com/doc/28249613/Data-Mining-Tutorial