Proceedings of the 12th INDIACom; INDIACom-2018; IEEE Conference ID: 42835
2018 5th International Conference on "Computing for Sustainable Global Development", 14th - 16th March, 2018
Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA)

A REVIEW ON THROAT CANCER DETECTION USING DATA MINING

Anushka Bhardwaj
Dronacharya College of Engineering, Gurgaon
anushka.18895@ggnindia.dronacharya.info

Chesta Sagar
Dronacharya College of Engineering, Gurgaon
chesta.18898@ggnindia.dronacharya.info

Neha Verma
Dronacharya College of Engineering, Gurgaon
neha.verma@ggnindia.dronacharya.info

ABSTRACT: Cancer, from its very name suggests, is a huddle of diseases and is an illness that is spreading all over the world at a phenomenal and impressive pace; may it be of Stomach, Skin, Throat, Prostate, Colon, Leukemia, Blood, Lungs, Breast, Cervix, etc. It is a disease that involves abnormal magnification in the cells having the potential to spread and damage many other parts of the body and eventually causes death of a person at some later stages of life.It has been seen that it outlines neoplasms. Its early detection plays a vital rolesince it can prevent spreading of the disease and is useful mainly for the two types of cancers that is cervical and colorectal. One, who is being troubled with cancer should be aware of its early detection and prevention techniques thoroughly and completely under the guidance and recommendations given by a highly qualified doctor. Data mining is one of the latest techniques used for the same whenever a person is detected having this disease. It refers to pulling out the data from the huge pre-existing datasett and hence a new information is then generated from it. This new information generated from the pre-existing dataset is analyzed and scrutinize thoroughly.

Keywords: Data Mining, Biological, Potential, Throat Cancer, Prediction, Neoplasms, Disease, Classification, Detection,Prostate, Colon,Leukemia, Blood, Lungs, Breast, Cervix,Pre-Existing, Dataset, Techniques.

I. INTRODUCTION

Cancer that is caused in the human mouth is customarily recognized and acknowledged as a sort of head and neck cancer that is rising globally and growing condemningin many regions around the Globe. Throat cancer is a type of cancer caused in throat. Throat cancer uses data mining technique for its detection and prevention. The techniques and the procedure used in data mining poses a great challenge for the researchers. It uses the basic technologies like clustering, classification and prediction. These are used to discover potential cancer patients. Cluster mining refers to the grouping of similar data, Classification predicts the class of data which has a target variable and this variable has an unknown value, and finally Prediction is used to foretell the missing or unavailable numerical values rather than class labels. The data is first gathered, then processed, fed to database and finally arranged and categorized to pull in significant patterns using decision tree algorithms and further clustering of data is also done and hence a prediction system is developed to determine risk factors of this cancer. Data mining technique had a great improvement in the biological field and is still being used at a good pace.

It is mainly of two main types: pharyngeal cancer and laryngeal cancer. Laryngeal cancer affects the larynx which is basically the voice box having cartilage and muscles that further enables us to talk. Therefore, laryngeal cancer can damage the voice of the person whereas the pharyngeal cancer occurs in pharynx that is the zone in and around the neck and throat. This is divided into three segments:-

a) The situated above part constitutes the Nasopharynx.

b) The middle part that is situated below the top layer constitutes the Oropharynx

c) The bottom partthat is the lowest of all the layers constitutes theHypopharynx

The oropharynx is the area in the back of the human mouth, or may be even the back of the throat. The oropharynx is sometimes called as pharynx. The usual symptoms carried bythis type of cancer can be because of the following possible reasons:-. Constant pain in the mouth or any other area of face; Thickening and formatting of a lump in the neck which forms a compact mass clot;Red and White color patch formation on the tonsil or lining of the mouth, gums or even on the tongue; May cause problem in swallowing or chewing and even one may face problem in speaking; Voice may get changed; High weight loss; Denture of the teeth may get changed; Bleeding from the mouth may also occur; and Continuous bad breath.Cancer is a chronic disease which is suffered by the people in majority who smokes, consumes any other type of tobacco, great

 ISSN 0973-7529; ISBN 978-93-80544-28-1

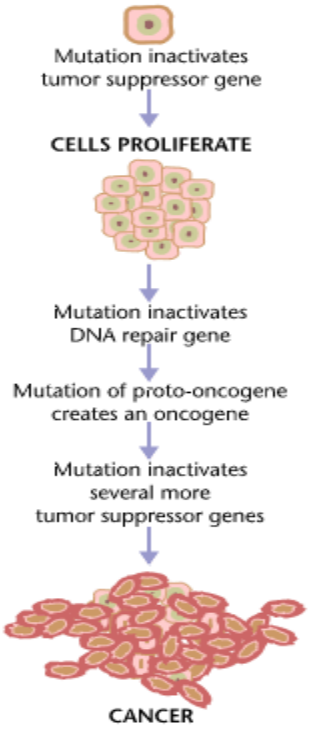

Fig. 1. Picture showing how cancer is formed

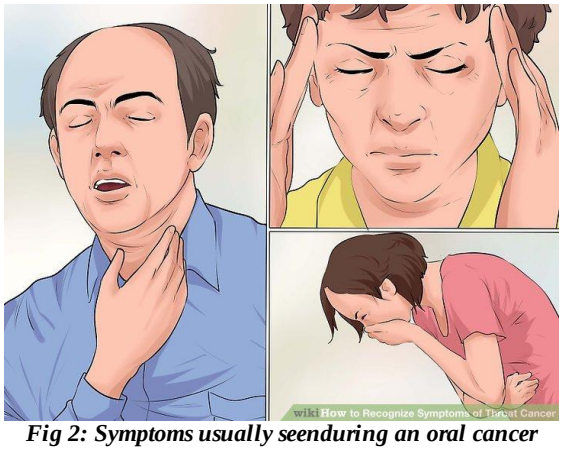

Fig 2: Symptoms usually seenduring an oral cancer

consumption of alcohol, excessive exposure to sun or may be inherited from the family.However, patients having oral cancer should be predicted by using various classification techniques.

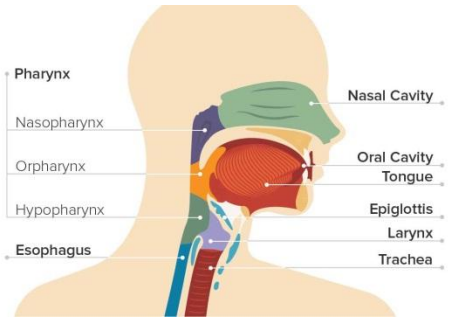

Fig 3:Diagram of human throat

The use and consumption of Tobacco by the humans causes approximately 22% of cancer deaths all over. Consumption of alcohol and maintaining poor health is also another factor that causes deaths majorly of about 10%.Certain and various types of infections, human exposure to radiations and many pollutants comes in the other categories leading to human deaths in this developing world. Due to Hepatitis B and Hepatitis C, the spread of cancer has been increased significantly. Mostly about 5-10% of the cancer patients are suffering from this disease because of the genetic reasons. They inherit this disease from either their mother or from their father or even from their grandparents. The people attacked by cancer can prevent their sufferings by keep good health, by making a habit of eating good, fresh and healthy food, fruits and vegetables, by avoid smoking, by getting vaccinated on regular basis against certain types of infections and also by maintaining a proper and healthy weight.

Our paper will be discussing the use of data mining technologies such as classification, clustering and prediction to reveal cancer in the cancer patients and differentiate them from the normal and non-cancerous ones. This technique is one such technique which allows implementing together to create a newness in the methods to diagnose the existence of cancer of a particular type in a particular patient. When ever initializing to work on Data Mining problems, it is essential to pull out all the data together into sets of instances. The data pulled must be properly assembled, integrated and cleaned. Here, in this proposed work, the dataset is fetched from various diagnostic centers which holds the cancer patients and also the patients' detailed particulars and the accumulated and assembled data is then pre-processed to remove duplicate and missing information and values.

II. LITERATURE REVIEW

V.Krishna[1] introduced a new system for the prediction of lung cancer that uses various Data Mining techniques. The most efficient and effective method to catch hold this disease appears to be IF-THEN rule and decision tree.*Singh.S*[2] used the Apriori algorithm for the reduction of cancer symptoms and thus having some chance to cure cancer. Five different types of cancer were taken into account and were determined for finding the cause of cancer and also the type of cancer that spreads faster. The symptoms of oral cancer can be as-change in the voice, heavy weight loss, swollen lymph

nodes,cough with blood,patches in the mouth which can be either red or white or even mixture of both red and white, pain in swallowing or chewing, etc. These symptoms may lead to higher stage of the cancer,if prolonged and not treated.The early detection of this cancer is usually a difficult task and is caught only when it is touching the last stage or may even at the last stage.It is ignored by the most of the patients or it also does not come in recognition of the patient. There are three treatments like,surgery, chemotherapy and radiation for reducing its effect.70percent of these treatments are usually not successful. But there are 40 to 50 percent chances of living up to 5 years even after taking these treatments and may sometimes end up dying.

III. RESEARCH METHODOLOGY

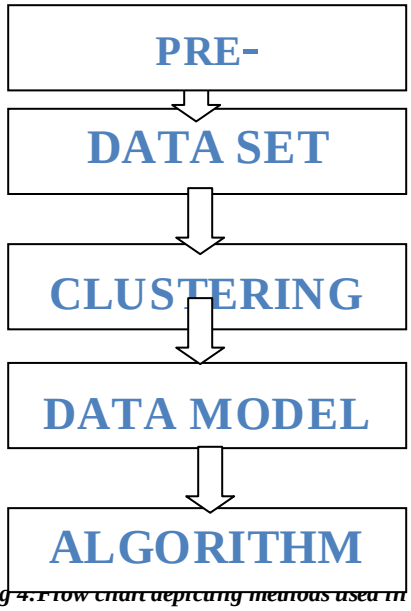

Fig 4.Flow chart depicting methods used in data mining

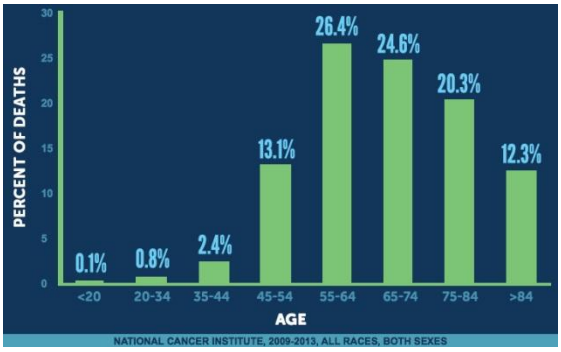

Fig 5: Graph depicting deaths due to oral cancer

A. DATA SETS

The datasethave hold of a number of variables including all thefields whichis controlled by the standard medical record type. National Minimum Data Set is one of the slightest quantity of the data that all expertor other specialist, doctors, surgeons are managing the patients which are expected to gather the data or particulars ofeach and every patients suffering with cancer.

B. PRE-PROCESSING

It comprises of the information ofboth cancer patients and patients without cancer, and the accumulated data is pre-processed for duplicate data, missing details or for any inconsistent data and offer a classification approach that is utilized. The pre-processing of the data is very crucial for good results in any data analyzing work.It also tells about the character and features of the data, obstacles existing in it and even can change the structure of the data.Data pre-processing involves various sub-techniques among which one is mentioned:-

1) DATA CLEANING-

It is the first step in data pre-processing which is used to take away or detach the noise and have errorless inconsistencies from the data.Datacleaning methodis used for cleaning of the data and to resolve all the inconsistencies which generallyincreases the confusion in data mining procedure that further results into inappropriate output.

C. CLUSTERING

It is used for analysing the noisy data and some symptoms are detected to group liver disarray. It generally clusters the data into several sub-parts.

1) 

Fig 6: Graph showing k-means clustering

Clustering analyses the series of procedural decision making steps. It is itself a very difficult task. There is a relationship between one cluster and the other. Clustering algorithms include:-

Connectivity-based clustering which is also called as hierarchical clustering

Centroid-based clustering

Distribution-based clustering

Density-based clustering

2)

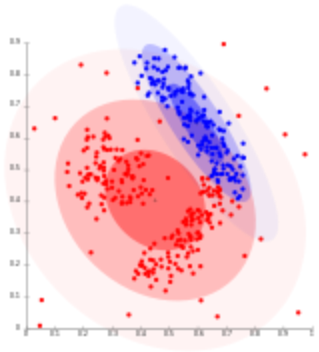

Fig 7: Graph showing Density-Based clustering

D. DATA MODEL

The assessment methods are mutual together using the group of datasetthat contains NMDS (National Minimum Data Set), which arranges it in some class patternin the list in order to pick out the whole database.

E. ALGORITHMS

The Data Mining algorithm discovers the calculations, which create a data mining type from data. In this project we use the efficient algorithm to utilize the result.

IV. TECHNIQUES

A. DECISION TREE

A decision tree acts as a supportive tool in Data Mining technique. It is a kind of a flow chart giving all the necessary details just like a tree structure having one root node, internal and external nodes. The internal nodes being the children of the root node and the external node or so called leaf node which is only represent at the bottom having no further children. Hence a simple tree structure would consist of a Root node on the top of the tree, in the middle it would have the internal nodes and at the bottom it holds the leaf nodes. This structure in Data Mining technique constitutes internal node which furtherindicates a test on an attribute,each branch would represents an outcome of the test and each leaf node holds a class label. The value is tested against this tree that is called as a decision tree. Theroute is then investigated from root node to the leaf node. Decision tree are the ones that can be easily and effortlessly transformedinto some classification rules. This tree in Data Mining process is used to generate some frequent patterns in the dataset. The data and the respective items in the sets that occur frequently in the database are known as frequent patterns. Significant frequent patterns are found with the help of frequent patterns that are greatly related to some cancer types or the others.

B. APRIORI ALGORITHM

In our project we have used two algorithms. One of them being is Apriori which is a classic algorithm that is used for frequent item set mining and association rule learning over the transactional databases. It works on the principle of identifying the frequent individual items of the database and using them in bigger item setstill the time those item sets are seen frequently in the database. The frequent item sets found usingApriori can be used to set the rules for the rest of the system or network. In our projectsome basic rules for mining like Apriori algorithm uses a bottom-up approach, along with breadth-first search and a hash tree structure are employed to count the candidate item sets efficiently.

V. CONCLUSION

Thus, oral cancer or more conveniently, a throat cancer, being a very chronic disease still persisting in our societies, can be cured when detected on time and analyzed thoroughly. Despite being the most dangerous disease among all, still a needful job can help the suffering patient to at least reduce his ailment.

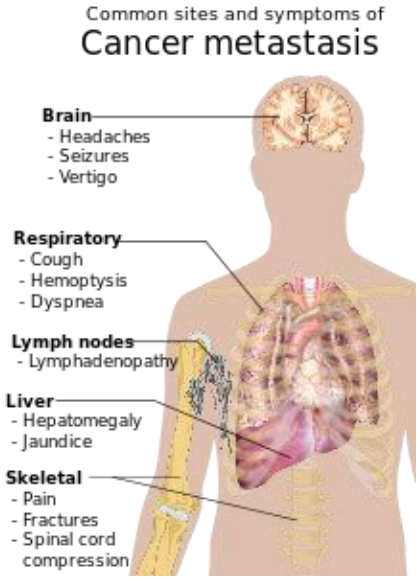

Fig 8: Picture showing Cancer metastasis

A person can prevent himself / herself from getting attackedby cancer by not smoke or using any tobacco products, not drinking alcohol, eating a well-balanced diet, limiting the body's exposure to the sun as repeated exposure can increase the risk of cancer on the lips, especially the lower lip. When in the sun, using UV-A/B-blocking sun protective lotions can reduce the direct sun exposure on yourskin, as well as your lips it can help a lot.

When detected, regular diagnosis and check-ups should be made a habit for better outcomes. Trying not to miss any check-up would be the best and the wise decision. If known that it can come as an inherited disease, early precautions are must. Avoiding obesity, making a habit of eating fruits and vegetables daily in a large quantity and limiting processed

meats will help the normal humans not to catch cancer so early.

A data predicts that in 2015, about 90.5 million people were there having cancer. Every year, it is found that approximately 15 million cancer cases are seen. Hence causing 16% of human deaths due to cancer. Children are mostly suffering from the brain tumors. Males are affected with mostly prostate cancer, lung cancerand stomach cancer. In females, breast cancer, cervical cancer and colorectal cancer are seen on a wide range. The living styles in this developing world is changing on a great extent hence the rate of cancer patients is increasing day by day and therefore deaths caused by cancer to the humans is also increasing tremendously.

RFERENCES

[1] G. Ries L, Hankey B, Miller B, Harras A, EBK, SEER cancer statistics review, 1973–1994. Bethesda, MD: US Department of Health and Human Services, Public Health Service, National Institutes of Health, 1997.

[2] Multidisciplinary Mgt. Guidelines (4th ed.) British Association of Head and Neck Oncologists.

[3] S S Jr. Oral cancer. (4th ed.) Hamilton, Ontario of American Cancer Society-1998.

[4] Landis S, Murray T, Bolden S, Wingo P. Cancer statistics, 1998.

[5] Improving outcomes in head and neck cancers National Institute for Health and Care Excellence (NICE), 2004.

[6] CDC and the National Institutes of Health. Cancers of the oral cavity and pharynx: a statistics review monograph, 1973–1987. Atlanta: US Department of Health and Human Services, Public Health Service, CDC, 1991.

[7] Mashberg A, Samit A. Early diagnosis of asymptomatic oral and oropharyngeal squamous cancers.

[8] Principles and practice of oncology (9th ed.) VT D Vita, S Hellman and S Rosenberg Lippincott, Williams and Wilkins for Cancer- 2011

