RESEARCH ARTICLE OPEN ACCESS

Predicting Colon Cancer Using Data Mining Techniques

K. Siva Sankari [1], Mrs M. Logambal [2]
M.Phil scholar [1], Assistant Professor [2]
Department of Computer Science
Navarasam Arts and Science College for Women Arachalur
Tamil Nadu - India

ABSTRACT
The most central element for death about the world by cancer. In 2015, there are 9.5 million cancer demise worldwide and future anticipated that would have 13 million deaths by growth in 2030. Early prediction of cancer the stage a very significant role in dropping deaths originate by cancer. In this research is to predict colon cancer. This research using data mining technology for instance clustering to identify potential colon cancer patients. The main aspire of this model is to provide the earlier warning to the colon cancer. In future, a prediction system is residential to examine risk levels which help in prognosis.
Keywords:- Colon Cancer, Clustering, Data Mining

I. INTRODUCTION

Data mining is a term from computer science. Sometimes it is also called **knowledge discovery in databases** (KDD). Data mining is about finding new in a lot of. The information obtained from data mining is hopefully both new and useful. In many cases, data is stored so it can be used later. The data is saved with a goal. For example, a store wants to save what has been bought. They want to do this to know how much they should buy themselves, to have enough to sell later. Saving this information, makes a lot of data. For data, there are lot of different kinds of data mining for getting new information. There are represented the predicted results.

Cancer is a disease of the body's own cells. Our bodies are made up of billions of cells and each one has a specific role to play. We are complex beings and there are many different types of cell – liver cells, brain cells, and blood cells and so on. Normally these cells are kept in check so that they only grow and divide when they are told to – such as when old cells need replacing or an organ needs repairing. In cancer these molecular checks are broken so cells are no longer kept under strict control. This can cause them to divide uncontrollably ultimately leading to a mass of cells known as a tumour – the physical manifestation of the disease we call cancer.

Colon cancer is also known as bowel cancer and colorectal cancer. A cancer is the abnormal growth of cells that have ability to invade or spread to another part of the body. Colon cancer is cancer of the large intestine (colon), which is the final part of your digestive tract. Most cases of colon cancer begin as small, noncancerous (benign) clumps of cells called adenomatous polyps. Over time some of these polyps can become colon cancers.

II. RELATED WORK

Data mining is used in various medical applications like tumor classification, protein structure prediction, gene classification, cancer classification based on microarray data, clustering of gene expression data, statistical model of protein-protein interaction etc. Adverse drug events in prediction of medical test effectiveness can be done based on genomics and proteomics through data mining approaches. Cancer detection is one of the hot research topics in the bioinformatics. Data mining techniques, such as pattern recognition, classification and clustering is applied over gene expression data for detection of cancer occurrence and survivability. Classification of colon cancer dataset using weka 3.6, in which Logistics, Ibk, Kstar, NNge, ADTree, Random Forest Algorithms show 100 % correctly classified instances, followed by Navie Bayes and PART with 97.22 %, Simple Cart and ZeroR has shown the least with 50 % of correctly classified instances. Kappa Statistic for Logistics, Ibk, Kstar, NNge, ADTree, Random Forest has shown Maximum. Mean absolute error and Root mean squared error are shown low for Logistics, Kstar and NNge. Using various Classification algorithms the cancer dataset can be easily analyzed.

III. CAUSES

In most cases, it's not clear what causes colon cancer. Doctors know that colon cancer occurs when healthy cells in the colon develop errors in their genetic blueprint, the DNA.

Healthy cells grow and divide in an orderly way to keep your body functioning normally. But when a cell's DNA is damaged and becomes cancerous, cells continue to divide — even when new cells aren't needed. As the cells accumulate, they form a tumor.

With time, the cancer cells can grow to invade and destroy normal tissue nearby. And cancerous cells can travel to other parts of the body to form deposits there (metastasis).

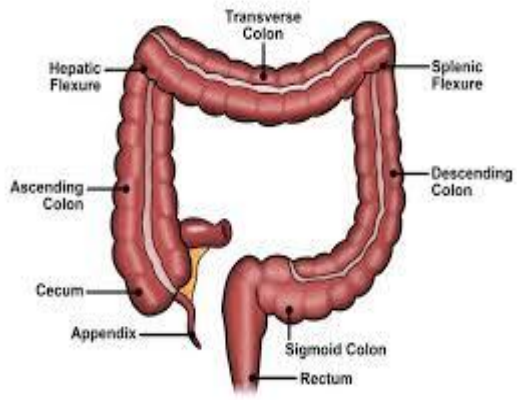

Figure 1: parts of colon

IV. RISK FACTORS

Factors that may increase your risk of colon cancer include:

1. **Older age.** The great majority of people diagnosed with colon cancer are older than 50. Colon cancer can occur in younger people, but it occurs much less frequently.
2. **African-American race.** African-Americans have a greater risk of colon cancer than do people of other races.
3. **A personal history of colorectal cancer or polyps.** If you've already had colon cancer or adenomatous polyps, you have a greater risk of colon cancer in the future.
4. **Inflammatory intestinal conditions.** Chronic inflammatory diseases of the colon, such as ulcerative colitis and Crohn's disease can increase your risk of colon cancer.
5. **Inherited syndromes that increase colon cancer risk.** Genetic syndromes passed through generations of your family can increase your risk of colon cancer. These syndromes include familial adenomatous polyposis and hereditary nonpolyposis colorectal cancer, which is also known as Lynch syndrome.
6. **Family history of colon cancer.** You're more likely to develop colon cancer if you have a parent, sibling or child with the disease. If more than one family member has colon cancer or rectal cancer, your risk is even greater.
7. **Low-fiber, high-fat diet.** Colon cancer and rectal cancer may be associated with a diet low in fiber and high in fat and calories. Research in this area has had mixed results. Some studies have found an increased risk of colon cancer in people who eat diets high in red meat and processed meat.
8. **A sedentary lifestyle.** If you're inactive, you're more likely to develop colon cancer. Getting regular physical activity may reduce your risk of colon cancer.
9. **Diabetes.** People with diabetes and insulin resistance have an increased risk of colon cancer.
10. **Obesity.** People who are obese have an increased risk of colon cancer and an increased risk of dying of colon cancer when compared with people considered normal weight.
11. **Smoking.** People who smoke may have an increased risk of colon cancer.
12. **Alcohol.** Heavy use of alcohol increases your risk of colon cancer.
13. **Radiation therapy for cancer.** Radiation therapy directed at the abdomen to treat previous cancers increases the risk of colon and rectal cancer

V. SYMPTOMS OF COLON CANCER

Signs and symptoms of colon cancer include:

1. A change in your bowel habits, including diarrheal or constipation or a change in the consistency of your stool, that lasts longer than four weeks
2. Rectal bleeding or blood in your stool
3. Persistent abdominal discomfort, such as cramps, gas or pain
4. A feeling that your bowel doesn't empty completely
5. Weakness or fatigue

6. Unexplained weight loss.

VI. METHODOLOGY

In this research study, using clustering technique to forecast colon cancer. Clustering be capable of measured the most significant unsupervised learning difficulty; so, as each another difficulty of this variety, it deals with judgment a structure in a locate of unlabeled data. A cluster be subsequently a set of substance which are "similar" along with them and are "dissimilar" to the substance belong to other clusters.

VII. CONCLUSION AND FUTURE WORK

Cancer is potentially incurable sickness. Detecting cancer is immobile difficult for the doctors in the field of medicine. Yet now the concrete reason and complete treatment of cancer is not invented. Detection of cancer in before period is curable. Prediction and clustering are the principal of data mining skills; they are largely used in healthcare sectors for medical diagnosis and predicting diseases.

In this research, work provides a valuable knowledge on colon cancer symptoms and its factors. The most important intend of this model is to afford the earlier warning to the users. In future, would like to implement clustering algorithm to predict colon cancer. Then generate a data set for colon cancer.

REFERENCES

[1]. Mohammed J. Zaki, Shinichi Morishita, Isidore Rigoutsos, "Report on BIOKDD04: Workshop on Data Mining in Bioinformatics", in SIGKDD Explorations, vol. 6, no. 2, pp. 153-154, 2004.

[2]. J. Li, L. Wong, Q. Yang, "Data Mining in Bioinformatics", IEEE Intelligent System, IEEE Computer Society. Indian Journal of Computer Science and Engineering, vol 1 no 2, pp. 114-118, 2005.

[3]. R. P. Kumar, M. Rao, D. Kaladhar, "Data Categorization and Noise Analysis in Mobile Communication Using Machine Learning Algorithms", Wireless Sensor Network, vol. 4, no.4, pp. 113-116, 2012.

[4]. Mark H. E. Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten, "The WEKA data mining software: an update", SIGKDD Explorations, vol. 11, no.1, pp.10-18, 2009.

[5]. D. J. Hand, "Statistics and data mining: intersecting disciplines", SIGKDD Explorations, vol. 1, no. 1, pp. 16-19, 1999. technology", Proceedings of PADD99: The Practical Application of Knowledge Discovery and Data Mining, pp.39-47, 1999.

[6]. C Apte, E Grossman, E Pednault, B Rosen, F Tipu, B White, "Insurance risk modeling using data mining

[7]. Liu, Bing, Chee Wee Chin, Hwee Tou Ng. "Mining topic-specific concepts and definitions on the web." Proceedings of the 12th international conference on World Wide Web. ACM, pp.251-260, 2003.

[8]. M.K. Jakubowski, Q. Guo, M. Kelly, "Tradeoffs between lidar pulse density and forest measurement accuracy", Remote Sensing of Environment, vol. 130, pp. 245-253, 2013.

[9]. E. Frank, M. Hall, L. Trigg, G. Holmes, I. H. Witten, "Data mining in bioinformatics using Weka", Bioinformatics, vol. 20, no. 15, pp. 2479-2481, 2004.

[10]. R. W. Burt, J. S. Barthel, K. B. Dunn, D. S. David, E. Drelichman, J. M. Ford, et al, "Colorectal cancer screening", Journal of the National Comprehensive Cancer Network, vol. 8, no. 1, pp. 8-61, 2010.

[11]. Data mining information's are available. Online. Available https://simple.wikipedia.org/wiki/Data_mining

[12]. Colon cancer details are available. Online. Available http://www.cancer.org.