RESEARCH ARTICLE   OPEN ACCESS

Predicting Blood Tumor and Hematology Using Data Mining Techniques

J.Abinaya [1], Mrs.A.Sumathi [2]

M.Phil scholar [1], Assistant Professor [2]

Department of Computer Science

Navarasam Arts and Science College for Women, Arachalur

Tamil Nadu - India.

ABSTRACT

Human healthcare systems generate a huge data collected from medical tests. Data mining is the compute process of learn pattern in huge statistics sets such as wellbeing ensure examinations. Blood diseases are not an exception; there are many test data can be collected from their patients. Loads of explore has been finished to spawn information from medical statistics by using data mining techniques. Though, there still needs to take out buried in sequence in the medical data, which can help in detecting diseases in the early stage or even before phenomenon. In this journal, we applied data mining techniques to settle on the connection between blood test uniqueness (Hematology) and blood swelling, in order to foretell the disease in an early stage, which can be used to develop the therapeutic capacity. The aim of our experiments is to produce model that can inform apart patients with regular blood sickness from patients who have blood tumor. The elucidation of policy that describes both tumor in blood and ordinary hematology.

Keywords:- Hematology diseases, Blood tumor, Clustering, Knowledge

I. INTRODUCTION

Data generated from healthcare domain is vast and complex. These data contain many hidden patterns which can help to discover and predict diseases in the medical field. The prophecy course of these diseases can diminish the numbers of mortalities and improve the superiority of life for the patients tainted with these diseases. Data mining was usually used in the healthcare area, for example, data mining can facilitate to perceive fraud and neglect of health insurance, craft customer bond supervision decisions by supervision, recognize efficient treatments and best practices by physicians .

Hematologic diseases study the blood diseases such as leukemia, thalassemia, lymphoma …etc. The medical feature of Hematology is troubled with the treatment of blood disorders. Hematologic diseases, like any additional healthcare fields, produce an vast amount of data. Traditional statistics is not adequate to study these data, using data mining techniques is superior option. Researchers academic information withdrawal techniques to Hematologic diseases normally utilize data generated from tests similar to Complete Blood Count (CBC) test. CBC test check the figure of blood cells circulating in the bloodstream. The test is a normal laboratory blood check that can be used to notice blood tumor and monitor tumor treatment. The aim of our study is to use data mining techniques to order CBC sample of a blood disease patient as regular hematology disease or blood tumor.

II. RELATED WORKS

Because of the enormous numbers of data in medical fields, which are available today, many researchers depending on data mining techniques to get new knowledge. Some of these research done in hematology diseases, such as: Abdullah and Al-Asmari in [1] used data mining to specify the anemia type for the anemic patients through a predictive model. They used real data constructed from the Complete Blood Count (CBC) test results of the patients. These data filtered and eliminated undesirable variables, then implemented on data mining techniques in the field of Shouval et al. in [5] used Allogeneic Hematopoietic Stem Cell Transplantation (SCT), that predicts transplantation outcome and donor selection. They proposed to use decision trees, Artificial Neural Networks (ANNs) and Support Vector Machines (SVM). No actual experiments were done. Al-shami and Al-halees in [2] used the data mining techniques on CBC tests to detect Thalassemia disease. They conducted four

type of experiments on the data with all attributes in their data set samples and then repeats the experiments after reducing some features from the dataset. The accuracy results of their experiments exceeded 90%, and it showed that the critical point which can be the first indicator of the thalassemia existence is MCV ≤ 77.65.

They used association rules and clusters on the collected data. Saichanma et al. in [3] used data mining technique to predict abnormality in peripheral blood smear from 1,362 students by using 13 data set of hematological parameters gathered from automated blood cell counter. They found that the decision tree, which is created by the algorithm, can be used as a practical guideline for RBC morphology prediction by using four hematological parameters (MCV, MCH, Hct, and RBC). They investigated which algorithm is most suitable for user working on hematological data. Their model can predict hematological data comment and developed a mobile application that can make diagnosis hematological data comments. The best algorithm based on the hematological data was J48 classifier with an accuracy of 97.16%. Finally, Vijayarani and Sudha in [4] developed weight based k-means algorithm for identifying leukemia, inflammatory, bacterial or viral infection, HIV infection and pernicious anemia diseases from the hemogram blood test samples data set. They found that the clustering accuracy of weight based k-means algorithm is better when compared to k means and fuzzy c means.

III. METHODOLOGY

In this paper, by means of clustering method. Clustering can be measured the most important unsupervised learning crisis. So, as each extra trouble of this sort, it deals with decision a arrangement in a anthology of unlabeled data. A *cluster* is consequently a assortment of item which are "similar" among them and are "disimilar" to the objects belonging to other clusters. Clusters are grouped into data into piece of objects. Clusters are very practical for data publishing.
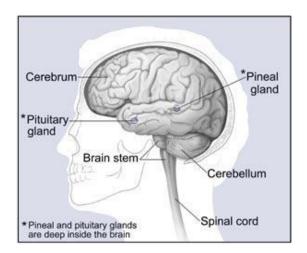
IV. COLLECTED DATASET

The dataset we used in this paper was composed from Coimbatore, KMCH Hospital. The dataset belongs to the Department of Oncology and Blood psychotherapy partition. The dataset

contains 5350 CBS samples subsequent to crackdown with special blood diseases. We separated the dataset into two groups , cluster one has 1764 CBC samples of blood tumor patients we labeled them as 'Tumor' and cluster two which has 3586 CBS samples of patients have other blood diseases, we labeled them as 'Hematology'.

Attributes of CBC sample

NO	SYMBOL	MEANING
1	WBC	White Blood Cell
2	RBC	Red Blood Cell
3	HGB	Hemoglobin
4	HCT	Hematocrit
5	MID	Mid-range Absolute Count
6	MCV	Mean Cellular Volume
7	MCH	Mean Cellular Hemoglobin
8	MCHC	Mean Cellular Hemoglobin Concentration
9	RDW	RBC Distribution Width
10	PLT	Platelets Count
11	MVW	Platelets Volume
12	GRAD	Percentage of white blood cells with granules in their cytoplasm
13	Gender	Male, Female
14	Class	Tumor, Hematology

V. PARTS OF THE BRAIN

The three main parts of the brain manage different measures:

- **Cerebrum**: The cerebrum use in spot aperture our true mind to notify us what is leaving on roughly us and tells our corpse how to oppose. It gearshift therapy, confidence, training, dialect, and emotions.

The cerebrum is separated into the left with right brainy hemispheres. The detailed hemisphere reins the muscles on the omitted side of the body. The misplaced hemisphere reins the strength on the specific part of the body.

- **Cerebellum**: The cerebellum pedals poise meant for on foot and footing, and extra compound measures.
- **Brain stem**: The brain stem connect the brain among the spinal cord. It reins inhalation, body hotness, with other vital corpse functions.

VI. SYMPTOMS OF BLOOD TUMOR

Symptoms (signs) of kind brain tumors time and again are not precise. The subsequent is a inventory of symptoms that, unaided or collective, can be caused by kindly brain tumors; unhappily, these symptoms can arise in lots of added diseases:

- dream evils
- trial harms
- poise troubles
- changes in mind skill (for example, attention, recall, talking)
- seizures, power jerking
- change in brains of stink
- vomiting/sickness
- facial paralysis
- headaches
- lack of feeling in extremities

VII. CONCLUSION

Attractive the superiority of living is the main idea of all healthcare explore. In this paper, we tried to include a few data to this position. We intrinsic our knowledge from CBC blood test personality. The experiments gave diverse

correctness rate according to the type of blood disease. In our prospect work, we resolve select a great dataset to experiment our copy on it, extra clustering also can be used.

REFERENCE

[1] M. Abdullah and S. Al-Asmari "Anemia types prediction based on data mining classificationalgorithms," Communication, Management and Information Technology – Sampaio de Alencar (Ed.)2017.

[2] H. Alshami and A. M. Alhalees, "Automated Diagnosis of Thalassemia Based on DataMiningClassifiers," Int. Conf. Informatics Appl., pp. 440–445, 2012.

[3] D. Minnie and S. Srinivasan, "Clustering the PreprocessedS. Saichanma, S. Chulsomlee, N. Thangrua, P. Pongsuchart, and D. Sanmun, "The observation reportof red blood cell morphology in Thailand teenager by using data mining technique," Adv. Hematol.,vol. 2014, pp. 4–8, 2014.

[4] S. Vijayarani and S. Sudha . "An Efficient Clustering Algorithm for Predicting Diseases fromHemogram Blood Test Samples". Indian Journal of Science and Technology, Vol 8(17), August 2015.

[5] R. Shouval, O. Bondi, H. Mishan, a Shimoni, R. Unger, and A. Nagler, "Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in SCT.," Bone Marrow Transplant., vol. 49, no. 3, pp. 332–7, 2014.