

Review of Data Mining Techniques for Glaucoma Detection

Prof. (Dr.) Ritu Sindhu *

^{1*}Department of Computing Sciences and Engineering, Galgotias University, Greater Noida, UP, India
e-mail: ritu.sindhu2628@gmail.com

Abstract- This review paper describes the data mining technique to detect Glaucoma disease. It is an eye disease and a typical disorder of the optic nerve, mainly this disease is a neurodegenerative disease which causes partial loss of vision. Large number of people suffers from eye diseases in rural and semi urban areas all over the world. As patients realize about the disease when their eyes gets already damaged upto 40%. So to detection and prediction for this disease as early as possible is needed, so that it can be cured at early stages and this will help not only patients but also the doctors to make it necessary practice to follow for every new patients. This process of diagnosis and detection can be achieved through various data mining technique and algorithm such as classification, clustering, association and fuzzy decision tree etc. In this paper we study attribute such as Age, B.P, Diabetes, Myopia, IOP, Family History of various patient. This dataset is taken as to the intension of predicting the accuracy result for glaucoma. Through this approach of data mining a patients can easily reach to a conclusion, whether there are chances of having him/her infected with glaucoma or not

Keywords- *Fuzzy, Decision Tree, Clustering, Classification, Myopia, Glaucoma, IOP, Data Mining, Health Care.*

I. INTRODUCTION

Lots of people in rural and semi-urban areas suffer from eye diseases such as Diabetic Retinopathy, Glaucoma, Age based Macular Degradation etc. Glaucoma is a pathological condition of optic nerve damage and is second leading cause of vision loss. It is a neurodegenerative issue of an eye that causes blindness if we does not take the proper initiative. It is known as silent killer of sight. It leads to the destruction of optic nerve head (ONH) that is caused by an increase in intraocular pressure within eye. These optic nerve carries image information signal to the brain. Due to the damage caused to various number of nerve fibers, a blind spot is created and that leads to the loss of vision for a person. One of the indicators of glaucomatous (person suffering from glaucoma disease) eye is the change in the appearance of optic disk. Approximately, 5 million people live with a glaucoma risk while around 800.000 people suffer from

glaucomatous damages in Germany. In Glaucoma the damage to the optic nerve is somewhat related to change in the fluid pressure of eyes that means IOP which flows around the eyeball. Glaucoma occurs mainly in those cases whose eye fluid pressure is high but it can also happened when IOP is normal. Fluid known as aqueous humor displace all the way through the way of the pupil, passes into the region connecting the iris and the cornea, & comes out of the eye through a tissue known as the angle. When moving of fluid, via the angle is moreover reduced or abruptly discontinues or stops, this increase the amount of fluid inside the eye, which enhance the chances of occurrence of glaucoma. This fluid increases the pressure in the eyes and it damages the nerve fiber and eventually hurts the eye optic nerve also, which is responsible for blind spot and this pressure is called as

IOP (Intra Ocular Pressure). Less common causes of glaucoma include a blunt or chemical injury to the eye, severe eye infection, blockage of blood vessels in the eye, inflammatory condition of the eye, and occasionally eye surgery to correct another condition. Glaucoma usually occurs in both eyes, but it may involve each eye to a different extent. If you are over age 40 and have a family history of glaucoma, you should have a complete eye exam with an eye specialist every one to two years. If you have both problems such as diabetes or a family

history of glaucoma or are at risk for other eye disease, you may need to visit your eye doctor more frequently.

II. DATA MINING TECHNIQUES

As Healthcare business or a health organization generates a large amount of composite data about patients, hospital resources, disease analysis, electronic record of patients, medical device etc we construct a dataset for using all these parameter to check the accuracy result for glaucoma. So based on these large amount, data is processed and analyzed for knowledge extraction for accuracy which enables cost cutting and decision making for a person based on relevant information. It brings a set of tool and technique that can be applied to processes data to discover hidden patterns.

Data Mining involves these following aspects:-

1. Data acquisition and preprocessing
2. Data representation
3. Cross validation
4. Association rule

Thus it correlates a specific set of symptoms that are related to a particular disease extracted from medical history of patient, diagnose and predict the possibility of a typical disease, so as to detect it early as possible so that it can be cured in early stages without being much damage. Data can be extracted from medical domains or an organization recorded in a database, basically we called it as dataset. After gathering data, possibility of having any of typical disease can be predicted or diagnosed by the use of classification and prediction data mining technique.

Issues in Data Mining Technique:-

1. Data cleaning:-It mainly refers to the process which is used to detect and identify inaccurate, incomplete, unnecessary data from database or record sets and to modify.

2. Relevance analysis:-Measure of attribute relevance analysis can be used to help identify irrelevant or weakly relevant attribute that can be excluded from the analysis process. The preprocessing of data and comparison is termed as analytical characterization.

3. Data Transformation:-Data transformation converts data values set from a sourced data system to a destination data system. It includes data mapping and code generation.

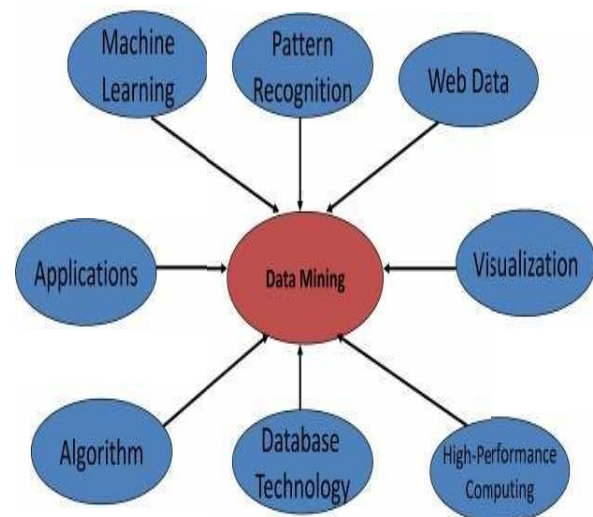


Figure 1- Data Mining Procedure

III. LITERATURE REVIEW

Zhao et.al. performed a work “**Web Based Decision Support and Telemonitoring System for the management of Diabetes**” which explains that advanced methods of time series analysis are used to find pattern and interpret any unwanted changes in the home data. It also provide daily management support and allows access to electronic diabetic record [1].

Kouznetsova et.al. performed a work “**Locating blood vessels in Retinal Images by Picewise Threshold Probing of a Matched Filter Response**” describes automation method to outline and locate blood vessels in image of the ocular fundus. Such a tool is useful to eye care specialist for purpose of treatment evaluation [2].

Balasubramanian et.al performed a work “**Glaucoma Progression Detection Using Variational Expectation Maximization Algorithm**” which present a new

framework for glaucoma progression detection using HRT images. And in this Maximization algorithm the task of inferring glaucomatous changes used for the first time[3].

Moniz et.al. performed a work “**Mining Electronic Medical Records for Patient Care Patterns**” describes the full electronic medical record of victim. Special care has been taken on performing clustering of highly dimensional dataset and cluster validation is also introduced into this[4].

Smith et.al performed a work “**Extending CRISP-DM to Incorporate Temporal Data Mining of Multi Dimensional Medical Data Streams: A Neonatal Intensive Care Unit Case Study**” explains the process of current cross industry standard for Data Mining approach. This is basically used to model intelligent data analysis based system which perform the technique of Temporal Data Mining[5].

Berndt et.al. performed a work “**Introduction to the Minitrack Database, Data Warehousing and Data Mining in health care**” explained that the nature of healthcare services is essentially based on the information and can be improved with effective information support, including data modeling, retrieval and analysis.[6].

Ahmed et.al. performed a work “**An Empirical Study of the Application of Web Mining Technique in health care**” explained that data mining technique are analytical tools that can be used with the use of dataset with different attribute to extract precious information specially when we talk about in health sector. This information is thus used for decision making process to predict the accuracy for given attribute by taking real life example [7].

Lee et. Al. performed a work “**A Framework for Mining Signatures from Event Sequence and Its Application in Healthcare Data**” described the process of novel temporal knowledge representation and learning framework that perform large scale temporal signature mining of longitudinal heterogeneous data. It enables extraction, representation and mining of high order event with relationship within single and multiple sequence[8].

Eikelboom et.al. performed a work “**Texture Analysis of Retinal Images to Determine Nerve Fibre Loss**” explained the concept of retinal nerve fiber where

several loss of these fibers due to glaucoma can be identified. Though this is more difficult with early loss. To determine texture analysis this tool is very useful as it measure nerve fiber loss[9].

Canlas et.al. performed a work “**Data Mining in Healthcare: Current Application and Issues**” explained a survey of current techniques of KDD using data mining tool and discuss critical issue and challenges that are related with data mining. And the application of Data Mining are in e-business, retail and marketing [10].

IV. ANTICIPATION IN DISEASE USING DATA MINING TECHNIQUE

Data Mining technique as we were discussing is a very effective technique for anticipation of any health disease with the attribute of a given dataset. This technique is employed with its different technique. Once data is collected and represented we can construct the data entry table which actually runs on different tool that gives the accuracy result of a person to be infected from a disease or not. Attributes has their meaning in themselves and should be selected very carefully since wrong input in data entry will lead to an interruption running the dataset. Here in this paper we are discussing the aspect of healthcare. Now what kind of attribute these can be is based on the abundance of data provided by health organizations. Thus to predict a disease at early stage by simply adding the information of different attribute this technique is quite useful. After entry of these attribute we have used a weka tool to calculate the accuracy precision of a person suffering from that disease or not.

Prediction of Glaucoma disease

As we know Glaucoma is a disease that relates to eye and it is a neurodegenerative disease which causes eye blindness if does not predicted early. We have used different attribute in the this kind of disease and these attribute are Age, BP, Diabetes, Glaucoma for predicting Glaucoma. Any value that is described once cannot repeat again. These attributes are created in the notepad and then saved file with either .arff or .xls extension. The data in notepad is then imported and preprocessed in WEKA. Here in our prediction for Glaucoma we have used association rule where the notepad file shows nominal data.

V. SIMULATION

```

@relation GLAUCOMA_ANALYSIS

@attribute AGE {25,22,42,10,67,30,38,56,54,28,44,18,52,64,71,51,35,45,62,32}
@attribute BP {HIGH,VERYHIGH,NORMAL,LOW}
@attribute DIABETES {100,124,135,80,142,102,134,87,129,65,93,70,137,122,95,149,119,126,132,107}
@attribute GLAUCOMA {YES, NO}

@data
25,NORMAL,100,NO
22,HIGH,124,YES
42,VERYHIGH,135,YES
10,NORMAL,80,NO
67,NORMAL,142,YES
30,NORMAL,102,NO
38,LOW,134,NO
56,LOW,87,YES
54,NORMAL,129,YES
28,NORMAL,65,NO
44,HIGH,93,NO
18,NORMAL,70,NO
52,VERYHIGH,137,YES
64,HIGH,122,YES
71,NORMAL,95,NO
51,VERYHIGH,149,YES
35,NORMAL,119,YES
45,HIGH,126,YES
62,VERYHIGH,132,YES
32,LOW,107,NO
    
```

Figure 2- Dataset for Glaucoma

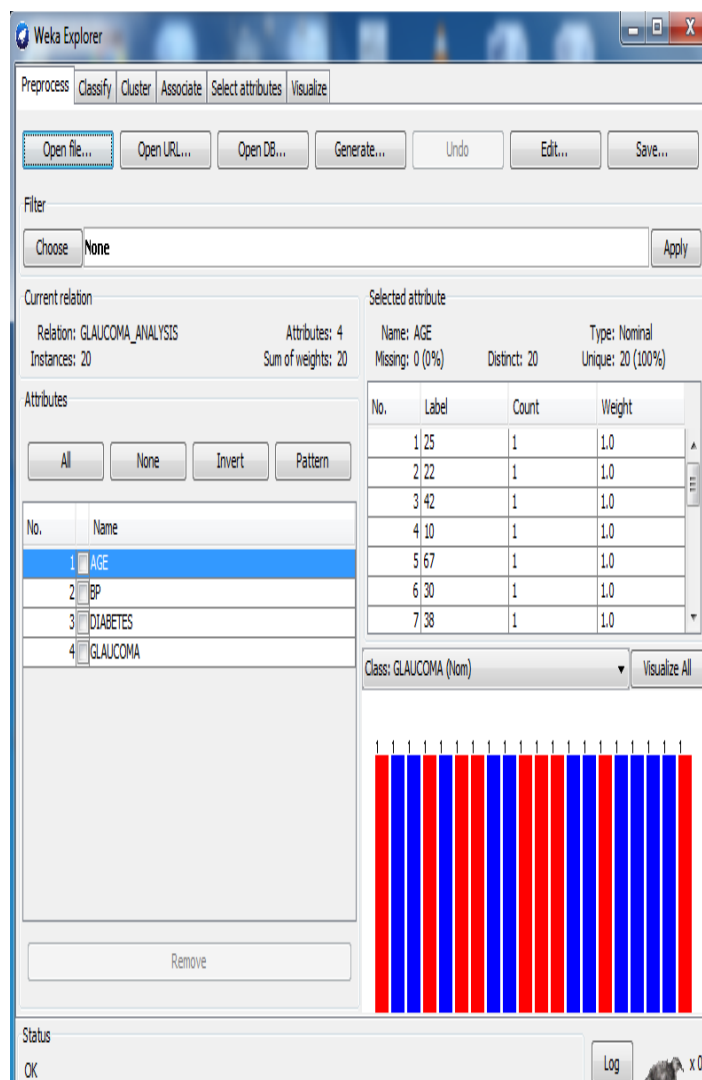


Figure 3- WEKA Output for Glaucoma

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	13	65	%
Incorrectly Classified Instances	7	35	%
Kappa statistic	0.3		
Mean absolute error	0.3958		
Root mean squared error	0.5015		
Relative absolute error	79.1667	%	
Root relative squared error	99.8601	%	
Coverage of cases (0.95 level)	90	%	
Mean rel. region size (0.95 level)	82.5	%	
Total Number of Instances	20		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.636	0.333	0.7	0.636	0.667	0.611	YES
	0.667	0.364	0.6	0.667	0.632	0.611	NO
Weighted Avg.	0.65	0.347	0.655	0.65	0.651	0.611	

=== Confusion Matrix ===

a b <-- classified as

7 4 | a = YES

3 6 | b = NO

Figure 4- Accuracy result using cross validation

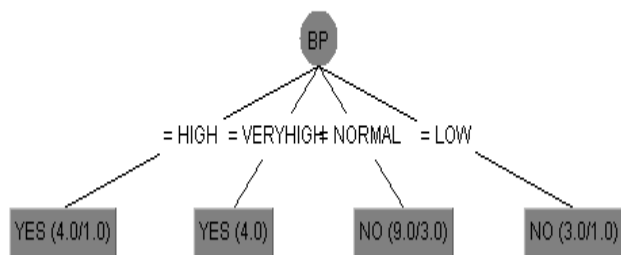


Figure 5- Tree view for glaucoma disease

VI. ANALYSIS

The figure 2 shows the dataset for the detection of glaucoma disease with its attribute given in notepad file with an extension of .arff file. The figure 3 shows the WEKA output for the given set of values. WEKA output shows the range of the attribute values with blue and the red line. The figure 4 shows the accuracy result using cross validation algorithm where every detailed summary is given for the given set of values. And at last the figure 5 shows the tree view for the result obtained using cross validation algorithm. Thus all these analysis shows that for given set of attribute the chance of suffering from Glaucoma disease can be counted using WEKA tool .

VII. CONCLUSION

In Healthcare department the development of web based application which is the part of Data Mining Application has been increasing with a rapid rate and their significance with the accuracy of predicting disease gives this technique a lots of importance. Implementation of system is to detect disease as early as possible so that if a person who might be suffering from this disease does not causes blindness. With the attributes we have given we comes to a conclusion that person suffering from this disease might cause partial or full blindness. So to detect it early and proper diagnosis will prevent person to be suffered from such a silent eye killer disease. As we discussed earlier these attributes have been collected from health organization as it is stored there in abundance. With attributes using various tools such as Weka tool, Rapid Miner Tool we concludes the accuracy precision for a person of having chances of disease in percentage. More the percentage there will be greater chance of having disease or a

person to be infected. So at last we comes to a conclusion that data mining techniques with the use of effective result is a big plus to our life and to predict a disease in its early stage is quite helpful as it reduces the chance of any severe problem and can be cured more precision.

REFERENCES

[1] A. Roudsari, S. Zhao, E. Carson, “Web-Based Decision and Telemonitoring System for the Management of Diabetes” , Proceedings of the 22nd Annual EMBS International Conference, July 23-28, 2000, Chicago IL.

[2] Adam Hoover, Valentina Kouznetsova, Michael Goldbaum, “Locating Blood Vessels in Retinal Images by Piecewise Threshold Probing of a Matched Filter Response”, IEEE Transactions on Medical Imaging , Vol.19, No. 3, March 2000

[3]Akram Belghith, Madhusudhanan Balasubramanian, Christopher Bowd, Robert N Weinreband Linda M. Zangwill, “Glaucoma Progression Detection Using Variational Expectation Maximization Algorithm” ,2013IEEE 10th International Symposium on Biomedical Imaging: From Nano to Macro San Francisco, CA, USA, April 7-11, 2013

[4] Anna L. Buczak, Linda J. Moniz, Brian H. Feighner, Joseph S. Lombardo, “ Mining Electronic Medical Records for Patient Care Patterns”, 978-1-4244-2765-9/09/\$25.00 ©2009 IEEE

[5]Christina Catley, Kathy Smith, Carlolyn McGregor, Mark Tracy, “ Extending CRISP – DM to Incorporate Temporal Data Mining of Multi-Dimensional Medical Data Streams: A Neonatal Intensive Care Unit Case Study”, 978-1-4244-4878-4/09/\$25.00 ©2009 IEEE.

[6] Donald J. Berndt, Paul J. Hu, Chih -Ping Wei, “Introduction to the Minitrac : Database, Data Warehousing, And Data Mining in Health Care” , Proceedings of the 33rd Hawaii International Conference On SystemScience-2000.

[7] Dr. Varun Kumar, MD. Ezaz Ahmed, “An Emperical Study of the Application of Web Mining Techniques in Health Care”, (IJACSA) International Journal of Advanced Computer Science and Application, Vol. 2, No. 10, 2011.

[8] Fei Wang, Noah Lee, Jianying Hu, Jimeng Sun, Shahram Ebadollahi, Andrew L aine, Fellow ,“A

Frameworkfor Mining Signatures from Event Sequence and Its Applications in Healthcare Data”, IEEE Transaction OnPattern Analysis And Machine Intelligence, Vol. 35, No.2, Feb 2013.

[9] K. Yogesan, R.H Eikelboom, C.J. Barry, “Texture Analysis of Retinal Images to Determine Nerve Fiber Loss

[10] Ruben D. Canlas Jr., “Data Mining in Healthcare : Current Application and Issues”, 2010 2nd International Conference on Computing, Communication and Networking Technologies, 9781-4244-6589-7/10/\$26.00©2010IEEE.