

Headache Disease Type Classification and Predicting System using Data Mining Techniques

Kalaivani J

Lecturer, Image college of arts animation and technology

ABSTRACT: In this Paper, Migraine Headache and the types of Migraine Headache were analyzed using Data mining Techniques. From the headache diary the data is being collected. **Naïve Bayes** is used on the collected data to determine the data corresponds to which category of Headache. It is also called as probabilistic algorithm which is used to predict and classify the category of Headache into the classes like **Migraine, Cluster, Sinus and Tension Headache**. Apply **Decision tree, C4.5** algorithm on the Migraine Class and classify the data set into three class (**Pediatric migraine, Episodic migraine, Chronic migraine**) based on the on certain conditions.

Keywords: C4.5, Chronic Migraine, Data Mining, Decision tree, Episodic Migraine, Migraine, Naïve Bayes, Pediatric Migraine

I. INTRODUCTION

Data mining is a non trivial extraction of raw data into some useful information. It is process of discovering hidden valuable knowledge by analyzing large amounts of data from different perspective using various data mining techniques. The analyzed knowledge can be used for various applications. Disease prediction plays a vital role in data mining.

Headache disease occurs due to work overload, stress and many reasons behind it. Based on doctor's experience and knowledge the diagnosis of the disease is often made. Sometimes this leads to unwanted and undesirable results and lots of medical loss of treatments provided to the patients.

This paper analyzes the migraine headache disease prediction and classification using different classification algorithm. Medical data mining techniques like Association Rule Mining, Clustering, Classification Algorithms such as Naïve Bayes, Decision tree, C4.5 Algorithm are implemented to analyze the different kinds of headaches and different types of migraine headaches.

II. BACKGROUND

There are two types of headaches such as primary headache and secondary headache. Primary headaches such as Migraine, Sinus, Tension, and Cluster are not associated with other diseases. The secondary headaches are associated with major or minor diseases.

Based on the literature survey tension headache is the most common type of primary headache and migraine is the second most common type of primary headaches. Sinus is the third most common type of primary headaches. Cluster headaches are a rare but important type of primary headache,

affecting mainly men.

Headaches are common in adults and children. Migraine headache is a disability. The first attack of migraine occurs at or before the age of 10. Migraine was often undiagnosed in children. There was about 10% of school-age children suffer from migraine. Many parents misthought that their child's headache is the sign of a brain tumor or other serious medical condition. Before puberty both girl child and boy child having same percent of chances to suffer in migraine. After puberty women are suffering in migraine more than men. Among 17% of women and 7% of men are suffering in migraine. Migraine pain exists from 4 to 72 hours. Based on the pain duration the severity was calculated. The children has less pain duration. The treatment of the headache based on the type and severity of the headache.

To undertake the classification, I obtain values taken from disease database and classify the dataset using different types of data mining techniques. Data mining is a powerful tool to extract knowledge from datasets. Disease prediction plays a vital role in data mining. The medical data mining provides lots of technique like association rule mining, classification technique, cluster technique. Classification is one of the supervised learning methods to extract models describing important classes of data. **Nikunj C. Oza** [16] proposed ensemble data mining techniques. The ensemble data mining technique is one of the emerging technique. The ensemble data mining technique is used to increase the classification rate and improved the majority voting of classification technique for individual classification algorithms like Bayes, decision trees.

III. WORK

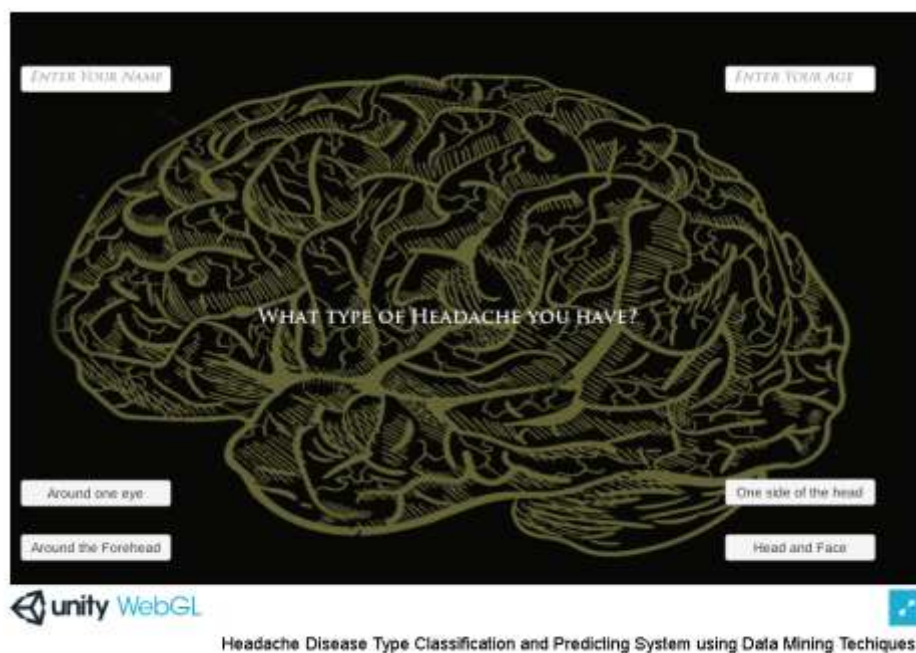
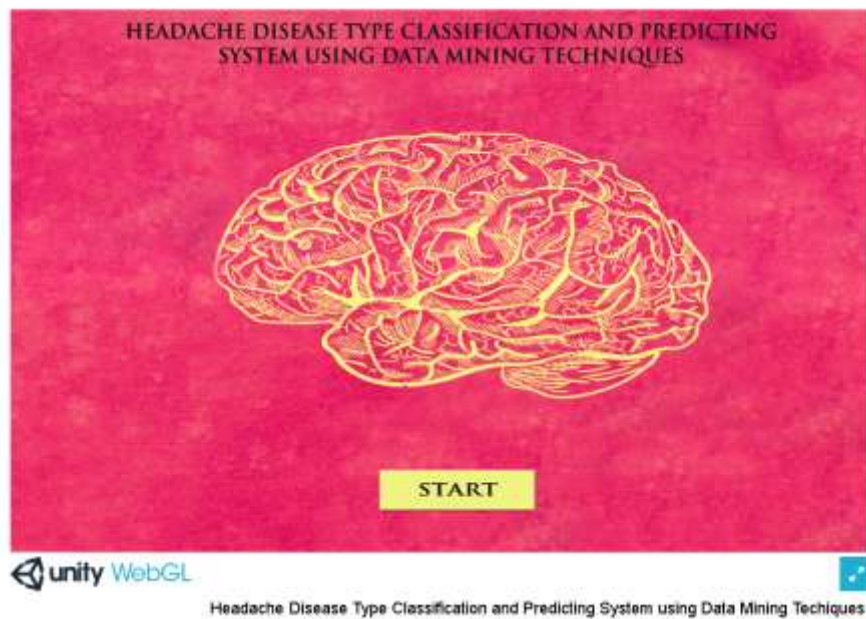
Now the data is being collected from headache diary for the research concern. The database having information about

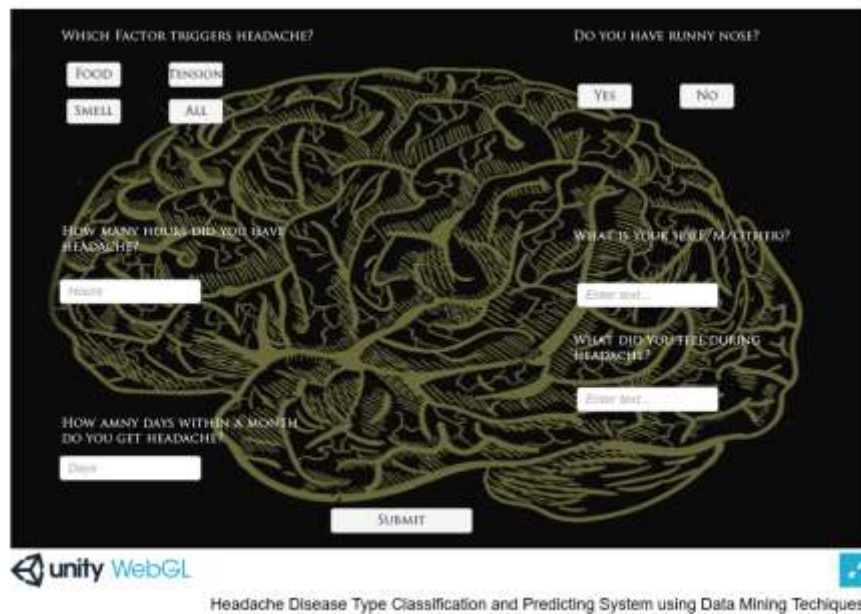
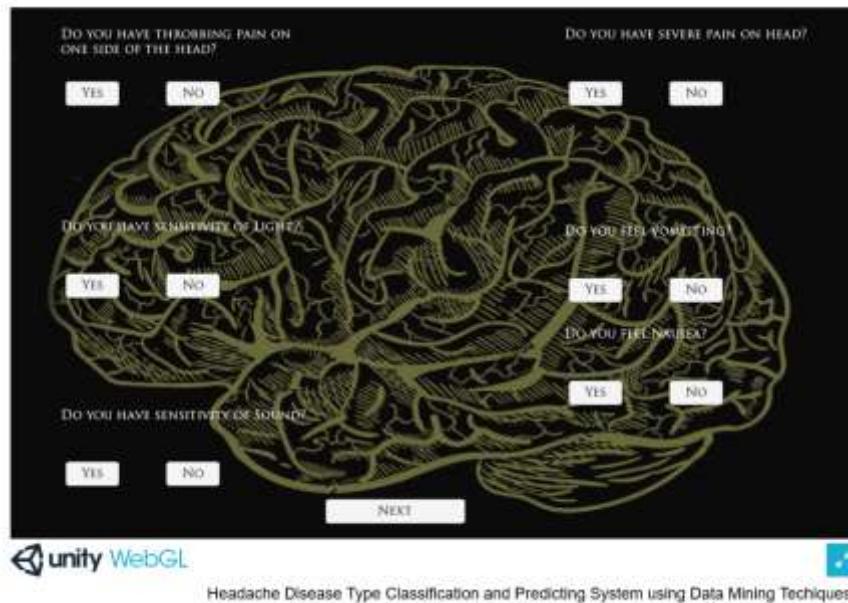
the headache datasets. The Naïve Bayes algorithm is a supervised function, which is used for the classification and prediction of Migraine headache and other primary headaches. Based on the result(Migraine Class) use Decision tree C4.5 algorithm to classify the types of Migraine headache.

In this study to select features or dataset to classify & predict Migraine and then classify the sub class based on dataset collected via questionnaires. The questionnaires are especially prepared for the headache-related disability patients are mentioned below. Patients should answer all the

questions. The answers should be saved on the database.

First convert the data saved on the into frequency table. Create Likelihood table by finding the probabilities based on that predict and classify the classes. Now, use Naïve Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction(Migraine). Then, Apply Decision Tree C4.5 algorithm to the Migraine class and classify it into three classes are Pediatric migraine, Episodic migraine, Chronic migraine



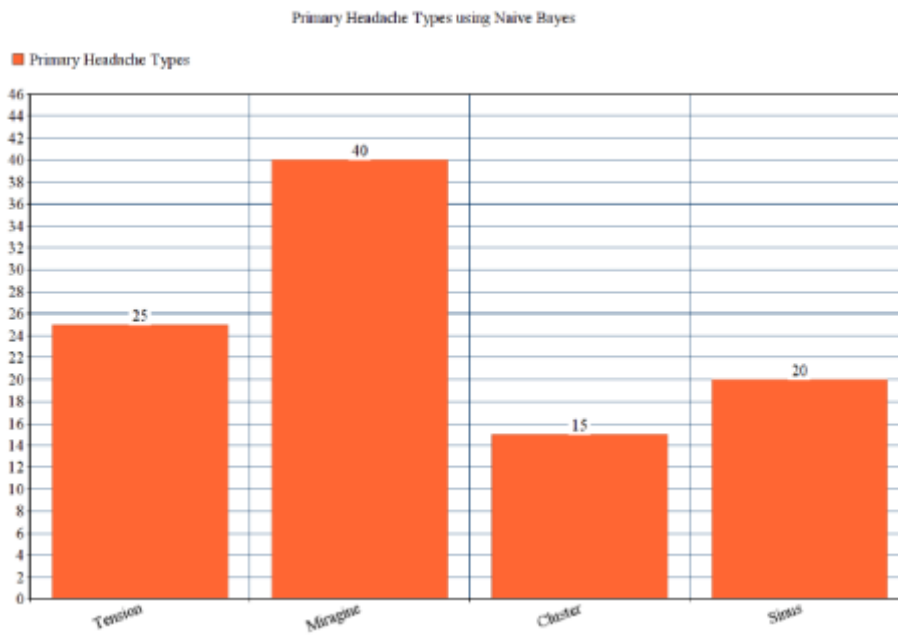


Naïve Bayes is a probabilistic model and it is used to generate multiple classes based on probability. This algorithm is also well known for multi-class prediction feature of target variable. Bayesian classification is based on Bayes theorem.

Bayes classifier has various features, among which each training data in the database affects the hypothesis probability. It associates a confidence percentage with the predictions by combining these predictions based on their confidence score. It classifies every new instance as a prediction function of multiple hypotheses that are weighted by their probabilities. By using the Bayes algorithm, construct pseudo code to find the probability of migraine and other classes.

Pseudo Code:

X ← set of variables or data in the database(x1, x2... xn)
 C ← set of classes or labels (c1, c2... cn)
 Condition ← set of conditions(condition1, condition2... conditionn)
 If X equals to condition1
 Label it as c1
 Else if X equals to condition2
 Label it as c2
 Else if X equals to condition3
 Label it as c3
 Else if X equals to condition4
 Label it as c4



Data mining tool is used to draw bar chart by applying Naïve Bayes algorithm

Apply the above pseudo code to the dataset. If the set of variables(data) present in the database equals to the condition then based on the posterior probabilities classify it and label it into the Primary headache type class. The Migraine symptoms and some of the Patient details is

needed. So, save the values in two separate tables called Patient Details and Migraine. The Migraine table(*Table-I*) contains the symptoms of migraine. The Patient Details table(*Table-II*) contains Patient’s Age, Pain Duration and Primary headache type.

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G
1	18	ThrobbingPain	SensitivityOfLight	SensitivityOfSound	SeverePain	Vomit	
2	1	Yes	Yes	Yes	Yes	Yes	
3	2	Yes	Yes	Yes	Yes	Yes	
4	3	Yes	Yes	Yes	Yes	No	
5	4	Yes	Yes	Yes	Yes	No	
6	5	Yes	Yes	Yes	Yes	Yes	
7	6	Yes	Yes	Yes	Yes	Yes	
8	7	Yes	Yes	Yes	Yes	Yes	
9	8	Yes	Yes	Yes	Yes	Yes	
10	9	Yes	Yes	Yes	Yes	No	
11	10	Yes	Yes	Yes	Yes	Yes	
12	11	Yes	Yes	Yes	Yes	Yes	
13	12	Yes	Yes	Yes	Yes	Yes	
14	13	Yes	Yes	Yes	Yes	Yes	
15	14	Yes	Yes	Yes	Yes	Yes	
16	15	Yes	Yes	Yes	Yes	Yes	
17	16	Yes	Yes	Yes	Yes	Yes	
18	17	Yes	Yes	Yes	Yes	Yes	

Table-I

	A	B	C	D	E
1	Age	PainDuration	HeadacheType		
2	5	5	Migraine		
3	42	6	Sinus		
4	24	8	Tension		
5	27	5	Cluster		
6	8	8	Migraine		
7	24	4	Sinus		
8	7	48	Sinus		
9	6	8	Migraine		
10	10	10	Migraine		
11	12	12	Migraine		
12	11	5	Cluster		
13	30	12	Migraine		
14	13	15	Migraine		
15	10	24	Migraine		
16	27	24	Migraine		
17	20	13	Sinus		
18	18	20	Sinus		
19	15	8	Tension		
20	27	24	Migraine		

Table-II

Decision trees are the popular tool for classification. The structure of the decision tree contains root node, branches and leaf nodes. Each internal node (non leaf node) denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (or terminal node) holds a class label. The most predictive variable is placed at the root node of the tree. C4.5 algorithm is especially designed for the decision tree. C4.5 is one of widely-used learning algorithms. Decision tree is a flow chart like tree structure simply constructed using if-then rules. C4.5 is also known as a statistical classifier.

Pseudo Code:

C ← target Classes

- Check for root node.
- For each element K, discover the normalized
- Let K_{best} be the element with the highest normalized information gain.
- Create a decision node that breaks on a best.
- Repeats on the sub nodes obtained by dividing it into C.

If age is less than & equals to 10 and pain duration is less then classify it then create leaf node with **Pediatric migraine**. If age is greater than 10, pain duration is more and pain occurs more than two days then create leaf node with **Episodic migraine**. If age is greater than 10, pain duration is more and pain occurs less than or equals two days then create leaf node with **Chronic migraine**.

	A	B	C	D	E
1	Age	Duration of Pain (hrs)	Number of Days (30)	Type of Migraine	
2	1	10	24	1	Pediatric migraine
3	2	16	48	2	Chronic migraine
4	3	27	24	1	Chronic migraine
5	4	12	12	5	Episodic migraine
6	5	10	10	6	Episodic migraine
7	6	25	24	1	Chronic migraine
8	7	23	72	10	Episodic migraine
9	8	8	9	4	Pediatric migraine
10	9	5	4	1	Chronic migraine
11	10	9	50	2	Chronic migraine
12	11	27	27	10	Episodic migraine
13	12	35	24	1	Chronic migraine
14	13	35	56	1	Chronic migraine
15	14	22	56	9	Episodic migraine

Table-III

Based on the above pseudo code construct conditions using if then rules. Then save the values in separate table called Type of Migraine(*Table III*).

The Unity, Sqlite are tools used for conduct the experiments. There are 25 records used for this experiments. I used some classifiers such as Naïve Bayes, Decision Tree C4.5 to find out the results appropriately.

IV. CONCLUSION

Data mining holds great potential in medical industry. It provides various techniques for the prediction of many diseases. In this paper i have presented an efficient approach for classification and prediction of Primary headache types,especially Migraine and also for the classification of Migraine types. This paper helps patients to understand about migraine. The migraine can be curable if the proper medication is taken based on the type of migraine. The episodic migraine should treated with medication and make it into chronic migraine. The chronic migraine and pediatric migraine should curable with appropriate treatments. In my future work i have planned to conduct experiments on big real time health data sets to predict the Migraine disease and classify it based on the trigger factors.

REFERENCES

- [1] Yolanda Garcia-Chimeno and Begonya Garcia-Zapirain, Automatic migraine classification via feature selection committee and machine learning techniques over imaging and questionnaire data,BMC Medical Informatics and Decision, 2017
- [2] M.A.Nishara Banu and B Gomathy,Disease Predicting System Using Data Mining Techniques,International Journal of Technical Research and Applications,2013
- [3] S.Vijayarani and S.Sudha,Disease Prediction in Data Mining Technique – A Survey,International Journal of Computer Applications & Information Technology and Applications,2013
- [4] Girish B Kulkarni and Girish N Rao,Headache disorders and public ill-health in India: prevalence estimates in Karnataka State,The Journal of Headache and Pain,2015
- [5] Rahul Deo Sah and Dr. Jitendra Sheeltani,Migraine (Headaches) Disease Data Classification Using Data Mining Classifiers,Quest Journals Journal of Research in Environmental and Earth Science,2017
- [6] Andrew H. Ahn and Peter J. Goadsby,Migraine and Sleep: New Connections,Cerebrum,2013
- [7] Josué Pagán and M. Irene De Orbe,Robust and Accurate Modeling Approaches for Migraine Per-Patient Prediction from Ambulatory Data,Sensors,2015
- [8] Marina de Tommaso and Marianna Delussi,Sleep features and central sensitization symptoms in primary headache patients,The Journal of Headache and Pain,2014
- [9] Smita and Priti Sharma,Use of Data Mining in Various Field: A Survey Paper,IOSR Journal of Computer Engineering,2014
- [10] Kenneth J. Mukamal and Gregory A. Wellenius, Weather and air pollution as triggers of severe headaches,Neurology 72,2009
- [11] Yong Peng and Yanming Ye, Jianwei Yin,Decision Tree Construction Algorithm Based on Association Rules,The 2nd International Conference on Computer Application and System Modeling,2012
- [12] Zaza Katsarava & Dawn C. Buse & Aubrey N. Manack & Richard B. Lipton, Jianwei Yin,Defining the Differences Between Episodic Migraine and Chronic Migraine,Springer,2012
- [13] Dorota Talarska, MaBgorzata Zgorzalewicz-Stachowiak, MichaB Michalak, Agrypina Czajkowska, and Karolina HudaV,Functioning of Women with Migraine Headaches,Hindawi Publishing Corporation the Scientific World Journal,2014
- [14] Gaurav L. Agrawal and Prof. Hitesh Gupta,Optimization of C4.5 Decision Tree Algorithm for Data Mining ,International Journal of Emerging Technology and Advanced Engineering,2013
- [15] Raj Kumar and Dr. Rajesh Verma,Classification Algorithms for Data Mining: A Survey,International Journal of Innovations in Engineering and Technology (IJET),2012
- [16] Nikunj C. Oza,Ensemble Data Mining Methods