Prediction and Diagnosis of Dravet Syndrome Data Set for Myoclone Disease

C.Radha, K. Ramesh Kumar

Abstract*: This article presents the detailed introduction on the dravet syndrome data set which has been collected from various medical organizations. The data set has various informations about the patients and it has been framed with different attribute. The myoclone disease has been identified in various patients of different geographic locations in different age groups. It occurs with the patients with various neurology conditions according to different other factors. We display the list of parameters like Prediction Accuracy, False Prediction Ratio and Rule Generation Efficiency and features considered.

Keyword:- Dravet syndrome, data set, myoclone disease

I. INTRODUCTION

This paper presents the detailed introduction on the dravet syndrome data set which has been collected from various medical organizations. The data set has various information about the patients and it has been framed with different attribute. This paper explains different information about each feature considered.

The myoclone disease has been identified in various patients of different geographic locations in different age groups. It occurs with the patients with various neurology conditions according to different other factors. To explain the features considered, it is necessary to pop a single instance here. We display the list of parameters and features considered here in this section.

Table 1: Details of features considered

S. No	Feature	Details	Possible Values and Explanation
1	PID	The patient ID is the value assigned as the unique factor for each patient by the medical organization	-
2	PNAME	Denotes the patient name	-
3	GENDER	Denotes the gender of the patient	M-Male F-Female
4	Age in Years	Denotes the age of patient in years	-
5	Age in Months	Denotes the age of patient in months	-
6	Mother Feed in Months	Denotes the number of months the child has given mother feed	-
7	Ist Occurrence	The month at which the first occurrence of the seizure has appeared.	-
8	Number of visits	The number of times the patient has visited to hospital	-
9	History of Parents	Denotes whether the parents has the seizure	Y- Yes N-No
10	Frequency in Seasons	Denotes the frequency of seizure in different seasons	Winter- Number Summer-Number
11	General Reason for Seizure Occurrence	Denotes the reason for the seizure occurrence	GR1-Sleeplessness-Number GR2-Tiredness-Number GR3-Weight Increase-Number GR4-Fever-Number
12	General Symptoms	Denotes the symptoms for the seizure occurrence.	GR1 – Number of mild jerks in finger - number GR2- Number of mild jerks in legs - number GR3- Number of upward starring in eyes- number GR4 – Number of whole body jerks - number
13	Tests	Denotes the result of different diagnosis	Blood Test- AN, NR Metablolic Disorder – AN,R EEG- AN, NR MRI- AN, NR Mscle Biopsy- AN, NR Bone Marrow- AN, NR

Revised Manuscript Received on September 15, 2019

C.Radha, Research scholar, Mass College of Arts & Science, Kumbakonam-612 501 .

Dr. K. Ramesh Kumar, Assistant Professor, Dept. of Computer Science, Mass College of Arts & Science, Kumbakonam-612 501

			CSF - AN, NR
14	Seizure Type	Denotes the type of seizure	Generalilzed – Y-yes N-No Pettimal – Y, N Absent – Y,N Myoclonic – Y,N Mitochondrial – Y, N Head Drops – Y, N
15	Vitamin Drugs	Name of drug given	Benadon in Mg **Omacartil in mg**

The above table shows the details of different factors considered to construct the data set. It has number of features and each feature vector has the above mentioned features.

The author collected different dravet syndrome data from various hospitals and stored in form of data set. The available data set has been converted into above mentioned form and has been used to perform diagnosis and myoclonic prediction.

II. PREDICTION OF MYOCLONIC DISEASE

This section discusses the various features considered for the myoclonic prediction and explains their importance in myoclonic disease prediction.

Mother Feed:

The mother feed plays the vital role in securing the babies from different diseases. It provides higher immunity to the child. The immunity of the child depends on the number of months it has given the mother feed. In order to perform myoclonic disease prediction, the mother feed in months must be considered.

First Occurrence:

The child would get the jerk at any age but getting the first appearance is the most important one. This features is used to compute the frequency of jerk being appeared.

History:

The heredity is the most important factor in identifying the presence of myoclonic. The parents would have suffered with the same problem and if they have then it must be considered to measure the possibility of getting appeared.

Frequency in Seasons:

The climate factors plays the most important role in identifying the presence of myoclonic. The disease has more influence when the climate is highly chill so that the temperature must be considered.

General Reasons:

The sleeping is the most important one for the human to all the parts of human body to function properly. If any body skipps the sleep then it will affect the functioning of the entire one. Similarly if the patient is tiered then also the user would have more chance of getting that.

Diagnosis Test:

The test results are more important in measuring or predicting the presence of myoclonic occurrence. The methods have been evaluated using different data sets and the dravet syndrome data set has been framed by the author which has been collected from different organization.

Table 2 Details of data set

Parameter	Value
Number of Instances	600
Number of attributes	15
Number of Patient records	200
Tool Used	Advanced Java

Table 2 shows the details of data set being used and present the information about the data sets. The data set contains hardly 600 records of 200 patients which has been collected from different medical organizations. Each record has 15 number of attributes which has been considered for the prediction of myoclonic disease.

For the evaluation number of methods has been considered and their performance has been measured on various measures.

Prediction Accuracy:

The prediction accuracy is the measure which represent the efficiency of the method in predicting the upcoming or future occurrence of the jerk. It has been measured using the below formula.

$$\text{Prediction Accuracy} = \frac{Number\ of\ successful\ Prediction}{Total\ number\ of\ tests} \times 100$$

False Prediction Ratio:

The false prediction is the measure which represent wrong prediction or wrong classification produced on a given sample. It can be measured as follows:

$$\text{False Prediction Ratio} = \frac{Number\ of\ false\ prediction\ or\ false\ classification}{Total\ number\ of\ samples}$$

Rule Generation Efficiency:

The rule generation efficiency is the measure which is computed based on the reduction of number of rules and shows the precise factor. It has been computed as follows:

$$\text{Rule Generation Efficiency} = \frac{Number\ of\ rules\ generated}{Number\ of\ rules\ selected} \times 100$$

III. CLASS ASSOCIATION RULES-BASED FEATURE SELECTION FOR DIAGNOSIS OF DRAVET SYNDROME

The class association rules based feature selection approach for diagnosis of dravet syndrome has been implemented and evaluated for its efficiency in dravet syndrome prediction. The method has produced efficient results in disease prediction with higher prediction accuracy.

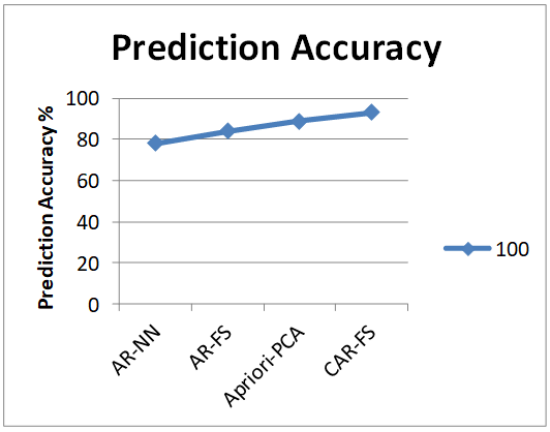

Figure 1: Comparison on prediction accuracy

The above figure shows the comparative result on myoclone disease prediction accuracy and shows that the proposed method has produced higher prediction accuracy than other methods. The CAR-FS approach produced the prediction accuracy upto 98%.

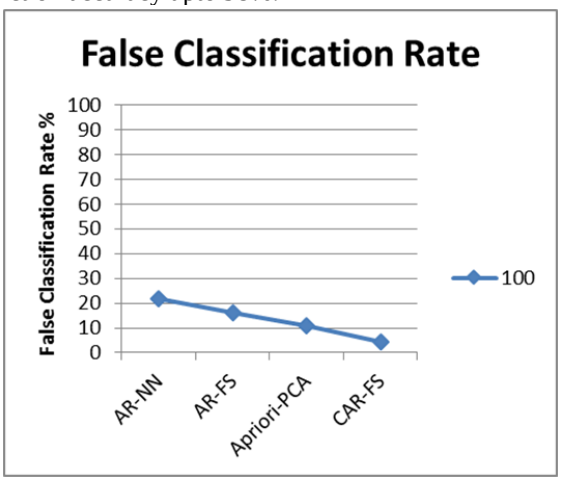

Figure 2: comparison on false classification rate

The false classification on myoclonic disease prediction has been measured and compared with different methods. The comparison results shows that the proposed CAR-FS approach has produced less false classification rate than other methods. The false classification ratio has been reduced upto 1.3%.

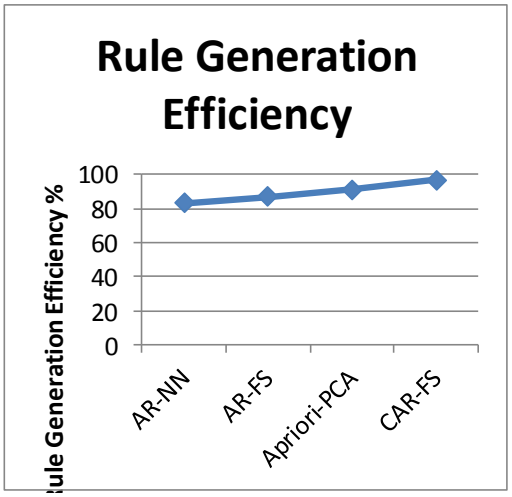

Figure 3: Comparison on rule generation efficiency

The efficiency in rule generation has been measured and compared with different methods. The evaluation results show clearly that the proposed method has produced higher rule generation efficiency than other methods. The rule generation efficiency has been increased upto 96.5%.

IV. CONCLUSION

The article presented the detailed information regarding the data set considered. It also shows the different factors and their values in measuring myoclonic. The data set obtained has more than 200 entities of various conditions and has been used to evaluate the performance of the proposed algorithms. The class association rules based feature selection approach for diagnosis of dravet syndrome has been implemented and evaluated for its efficiency in dravet syndrome prediction. This article has produced efficient results in disease prediction with higher prediction accuracy.

REFERENCES

1. Ceulemans B, Boel M, Claes L, Dom L, Willekens H, et al (2004), "Severe myoclonic epilepsy in infancy", Toward an optimal treatment. J Child Neurol, Vol. 19, pp.516–521.
2. Ingrid E. Scheffer (2012), "Diagnosis and long-term course of Dravet syndrome", European journal of pediatric neurology, pp.1-4.
3. Nienke E. Verbeek, Nicoline A. T. van der Maas, Floor E. Jansen, Marjan J. A. van Kempen, Dick Lindhout and Eva H. Brilstra (2013), "Prevalence ofSCN1A-Related Dravet Syndrome among Children Reported with Seizures following Vaccination: A Population-Based Ten-Year Cohort Study", PLOS ONE Journal, Vol. 8, No.6.
4. Peter C. Austin, Jack V. Tu, Jennifer E. Ho, Daniel Levy and Douglas S. Lee (2013), "Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes", Journal of Clinical Epidemiology, Vol. 66, pp.398-407.
5. Swati Gupta (2015), "A Regression Modeling Technique on Data Mining." International Journal of Computer Applications, Volume 116 – No. 9, pp.27-29, 2015.