

Classification of EEG based Diseases using Data Mining

Deepika Kundra
School of computer application
Lovely professional university
Phagwara, Punjab

Babita Pandey
School of computer application
Lovely professional university
Phagwara, Punjab

ABSTRACT

Brain is the most important organ of our body and it serve as centre of nervous system. It is a bioelectric generator. With neurological and psychological disordering there occur some inflammation and disruptions in the brain that lead some problems like vision loss, weakness, brain functioning, anxiety and abnormal behaviour etc. that leads to the death. In this work, J48 algorithm is deployed for the classification of EEG based diseases such as: dementia, Schizophrenia, ADHD, epilepsy and mood disorder. The results are shown in terms of, decision tree true positive rate, false positive rate and ROC curve.

Keywords

Classification, data mining, J48 algorithm, confusion matrix, ROC curve, EEG based diseases.

1. INTRODUCTION

Brain is the most important organ of our body and it serve as centre of nervous system. It is a bioelectric generator. With neurological and psychological disordering there occur some inflammation and disruptions in the brain that lead some problems like vision loss, weakness, brain functioning. EEG (electroencephalography) is a medical imaging techniques test that measures the electric activity of the human brain [1]. EEG signals related to the level of consciousness of the person, and contain very useful information relating to different physiological state of brain.

A lot of work has been done by many authors for the classification of bioelectric signal classification [2, 3, 4, and 5]. Few of the authors have deployed the RBR, CBR and ANN for the diagnosis of EEG, ECG and EMG based diseases [6]. Gangwar et al., [6] deployed rule based reasoning (RBR), case based reasoning (CBR) and artificial neural network (ANN) for the diagnosis of EEG based diseases. They use the muscular, cognitive, psychological and EEG signal parameter for the diagnosis. This model faces the problem of rule acquisition. Glover et al., [7] developed a system for the automated interpretation of bioelectrical signals. They use two Knowledge-based approach one for the signal analysis and other for the more subjective interpretation knowledge. Wu et al., [8] proposes an effective classification method on probabilistic neural network with supervised learning. Genetic algorithm is also used training the neuron's smoothing parameter and hidden central vector for determining hidden neurons. Ganesan et al., [9] proposes a system that integrate wavelet transform, Feature extraction and Artificial Neural Network for the detection and classification of Epileptic and non-Epileptic spike in EEG. Li et al., [10] developed an EEG based BCI system for classification, they apply two development Adaboost classifiers on the basis of an advanced boosting learning algorithm: AdaboostNN and Gentle Adaboost.

A very few literatures are available for the diagnosis of EEG based neuropsychiatric diseases. In this work, J48 algorithm is deployed for the classification of the different EEG based diseases such as: dementia (D), schizophrenia (SI), Attention deficit hyperactivity disorder (ADHD), epilepsy (E), mood disorder (MD) using physical, psychological, cognitive and EEG based parameters.

The organization of the paper is as follows. Section 2 presents the description of the various EEG based diseases with their psychological, cognitive and physiological parameters. The EEG parameters such as effect on different brain lobes and different wave (alpha, beta, theta, and delta) are also mentioned in this context. Section 3 describes the data mining technique with J48 classification method.

Section 4 shows the computation and results for diagnosis of EEG based disease using Weka tool (J48 algorithm). Section 5 deals with conclusion.

2. EEG BASED DISEASES

EEG based diseases such as: dementia, Schizophrenia, ADHD, epilepsy and mood disorder. The classifications of different diseases are based on two parameters [6]: psycho-physical (PP) and EEG signal characteristics. Psycho-physical parameter are further divided into three parts 1) psychological (PSY) parameter are anxiety (AN), abnormal behaviour (AB), fear (FR), hyper activity (HA), stress (ST), hallucination (HL), agitation (AG). 2) The cognitive parameter (COG) are confusion in decision making (CD), learning disability (LD), visual disability (VD), judgement (JD), speech disability (SD), memory loss (ML). 3) The physical (PHY) parameters are problem in walking (WL), climbing (CL), hearing (HR), vision (VS), speech (SH), oversleeping (OS).

EEG signal characteristic is divided into two parts :1) EEG signal abnormality observed in different brain lobe (BL) such as frontal lobe (FL), parietal lobe (PL), occipital lobe (OL), temporal lobe (TL). 2) Changing Behaviour of the different EEG waves (W) that are alpha (AL), beta (AB), theta (TH) and delta (DE) [6].

Dementia (D) Dementia causes the reduction in ability to learn reason or recall past experiences. All kind of dementia is caused by brain cell death, head injury and stroke. The sign and symptoms of D are: ML, AB, HL, AG, CD, ML, VS, HR, FL, PL and TL.

Schizophrenia (SI) Schizophrenia is a long term brain disorder; it strikes in between the ages of 15 and 25. The sign and symptoms of SI are AB, HL, SD, ML, OS, PL, DL and TL.

Attention deficit hyperactivity disorder (ADHD) ADHD is one of the most common childhood disorders. It is a psychiatric disorder of the neurodevelopment type. The sign

and symptoms of ADHD are AB, HA, LD, VD, CL, HR and VS.

Epilepsy (E) Epilepsy caused by some unexpected electrical disturbances in the brain. The sign and symptoms of E are AN, FR, ST, CD, LD, VD, SD, ML, WL, CL, HR, SH, FL and TL.

Mood disorder (MD) Basic kinds of mood disorder are depression, cyclothymia, seasonal affective disorder and mania. The sign and symptoms of MD are AN, HA, ST, CD, LD, VS, OS and DL.

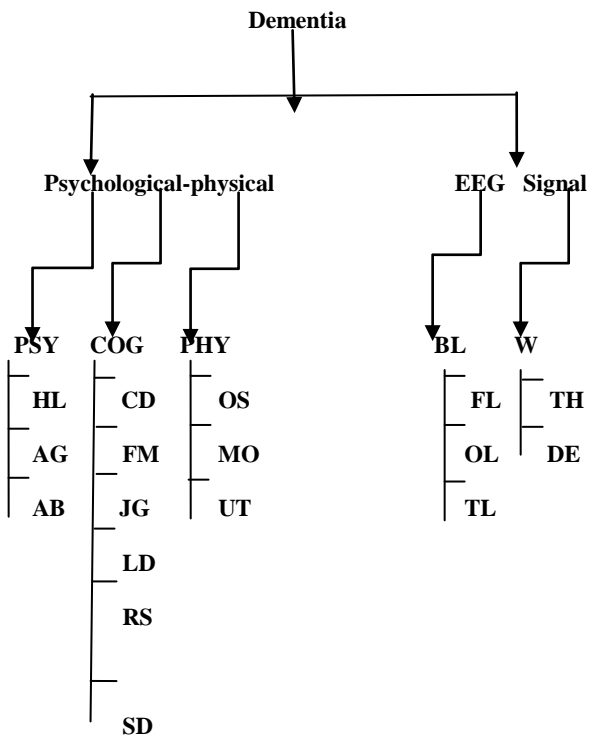


Fig 1: Hierarchical correlation of sign and symptoms of the dementia disease.

3. DATA MINING (DM)

Data mining is a non trivial extraction of implicit, previously unknown, and imaginable useful information from data. DM is used to find important information hidden in large volumes of data. Various DM techniques are Association rule, classification, regression, summarization and clustering. Classification methods such as: j48, it classify the data according to the features of the data with respect to the predefined set of classes [11].

J48 build univariate decision tree from a dataset using the concept of information gain by entropy. Each attribute of dataset is used to make a decision by splitting the data into smaller part. J48 finds the difference in entropy and the attribute heaving the highest normalized information gain is used to make the decision [12].

J48 can handle nominal and numeric attributes (continuous and discrete). It can also generate pruned tree, confusion matrix and decision tree. Classification is done by sensitivity (SN), specificity (SF) and accuracy. SN and SF values are obtained from true positive rate and false positive rate values in confusion matrix [11]. J48 can also handle the data set with missing attribute value.

4. RESULT

The diseases are classified using J48 algorithm of Weka tool. Data which is used for the experiment is **Fig 2: pruned tree generated by Weka** collected from hospital. The data set consist of 100 records and 5 attributes. In which 75 records are used for training and 25 records are used for testing. Detailed description of data set is given in Table 1.

Table 1: Detail of dataset

Attributes	Data type
Psychological	Categorical((high ,medium ,low)
Cognitive	Categorical(high, medium, low)
Physical	Categorical(high, medium, low)
Brain lobes	Categorical(high ,medium, low)
Brain waves	Categorical(high ,medium, low)

J48 generate rules for pruned tree and decision tree as shown in Figure 2 and 3 respectively.

J48 pruned tree

```

-----
brainlobe = high
| cognitive = high: dementia (7.0/1.0)
| cognitive = medium: SI (10.0/2.0)
| cognitive = low: SI (1.0)
brainlobe = low
| physical = medium: no disease (3.0/1.0)
| physical = high: ADHD (10.0)
| physical = low: no disease (3.0)
brainlobe = medium
| psychology = high: mood disorder (5.0)
| psychology = medium: epilepsy (3.0/1.0)
| psychology = low: epilepsy (4.0)
    
```

Number of Leaves : 9

Size of the tree : 13

Fig 3: pruned tree generated by Weka

Rules generated by J48 pruned tree

Rule 1: If brain lobe (BL) symptoms and cognitive (COG) symptoms are high, Then disease is “dementia”.

Rule 2: If brain lobe (BL) symptoms are high and cognitive (COG) symptoms are medium/low, Then disease is “SI”.

Rule 3: If brain lobe (BL) symptoms are low and physical (PHY) symptoms are medium/low, Then there is no disease.

Rule 4: If brain lobe (BL) symptoms are low and physical (PHY) symptoms are high, Then disease is “ADHD”.

Rule 5: If brain lobe (BL) symptoms are medium and psychological (PSY) symptoms are high, Then disease is “mood disorder”.

Rule 6: If brain lobe (BL) symptoms are medium and psychological (PSY) symptoms are medium/low, Then disease is “epilepsy”.

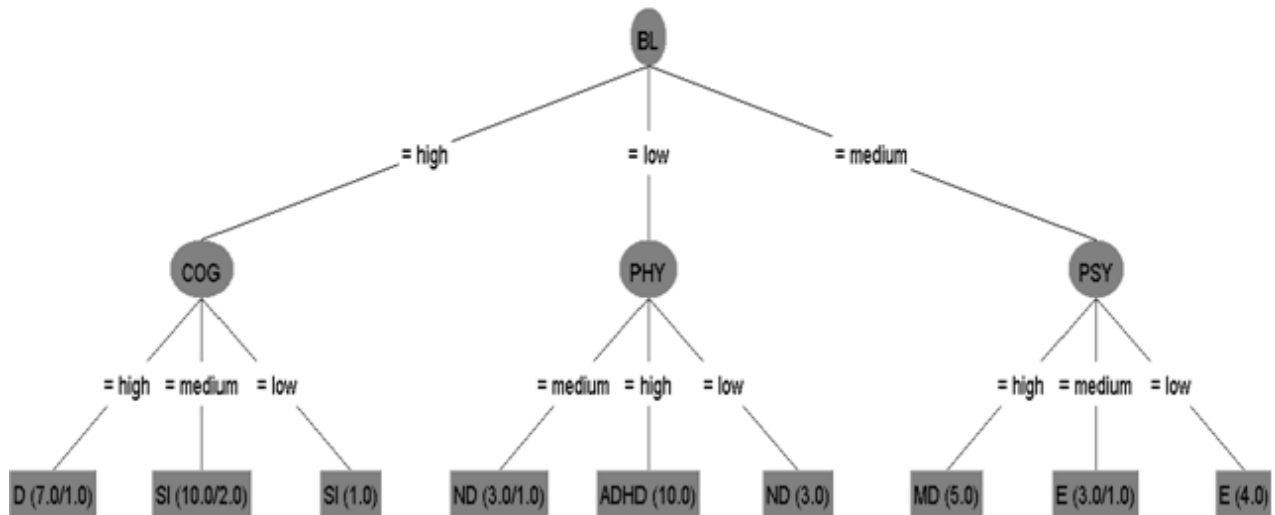


Fig 3: decision tree generated by Weka

Different attribute have been chosen randomly from the dataset. J48 is applied on the dataset and a confusion matrix is generated as shown in Fig 4.

=== Confusion Matrix ===

a	b	c	d	e	f	<-- classified as
6	0	0	0	0	0	a = dementia
1	9	0	0	0	0	b = SI
0	0	10	0	0	0	c = ADHD
0	0	0	5	1	0	d = no disease
0	0	0	1	6	0	e = epilepsy
0	2	0	0	0	5	f = mood disorder

Fig 4: confusion matrix

From above confusion matrix, true positive for class a =”dementia” is 6 while false positive values are 0, 0, 0, 0,

0. Whereas for class b =”SI”, true positive is 9 while false positive values are 1, 0, 0, 0, 0. For class c =”ADHD”, true positive is 10 while false positive values are 0, 0,0,0,0. For class d =”no disease”, true positive is 5 while false positive values are 0, 0, 0, 1, 0. For class e =”epilepsy”, true positive is 6 while false positive values are 0, 0, 0, 1, 0. For class f =”mood disorder”, true positive is 5 while false positive values are 0, 2, 0, 0, 0. Now simply diagonal elements of the confusion matrix represents the true positive (TP) values and the rest of the elements represent the false positive (FP) values. [13]. Different operative characteristics are defined as follows:

True positive (TP) = when test outcome is positive and condition is positive.

False positive (FP) = when test outcome is positive and condition is negative.

True negative (TN) = when test outcome is negative and condition is negative.

False negative (FN) = when test outcome is negative and condition is positive.

Table 2: TPR and FPR values

For Class	Sensitivity (TPR)	FPR	Specificity (1-FPR)
A	1	0.025	0.975
B	0.9	0.056	0.944
C	1	0	1
D	0.833	0.025	0.975
E	0.857	0.026	0.974
F	0.714	0	1

The ROC curve obtained from Weka (J48 algorithm) tool is shown in Fig 5,6,7,8 and 9 for: dementia, Schizophrenia, ADHD, epilepsy and mood disorder. It compares two operating characteristics that are true positive rate value and false positive rate value. Receiver operating characteristics curve (ROC) and area under curve (AUC) are calculated by Threshold Curve class in Weka. This tool is used in many fields for diagnostic test evaluation.

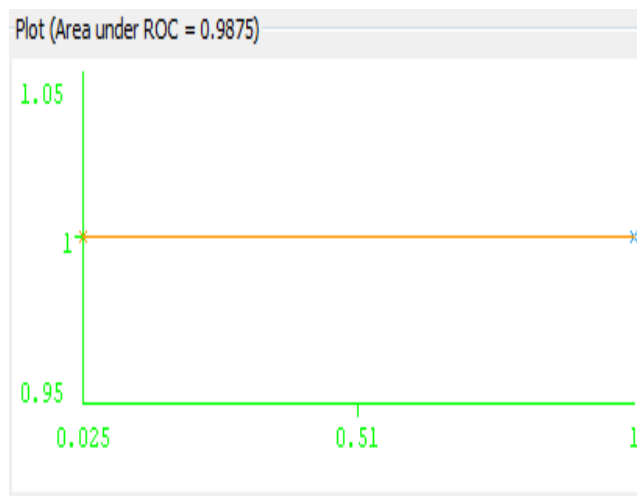


Fig 5: Class value D

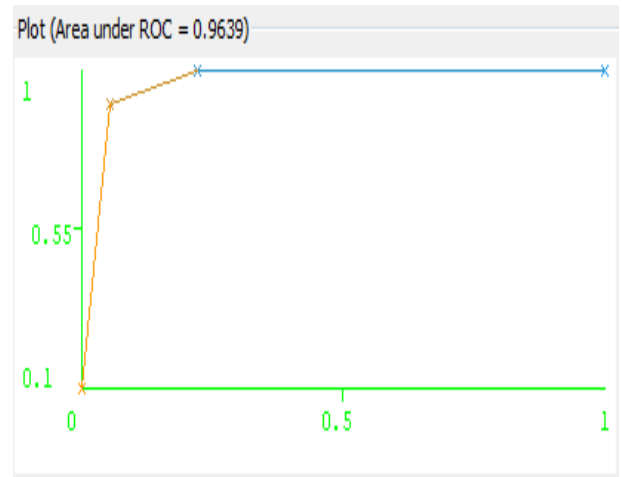


Fig 6: Class value SI

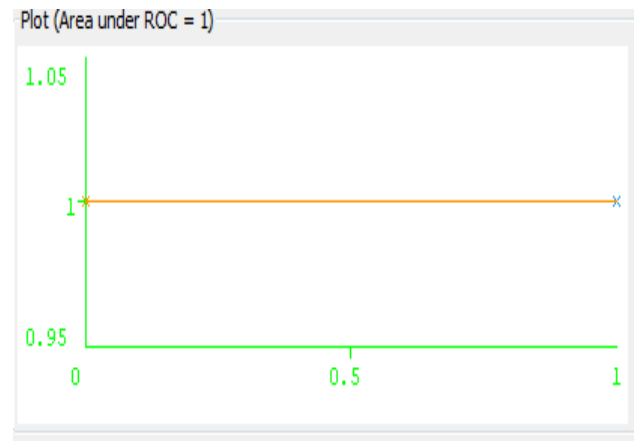


Fig 7: Class value ADHD

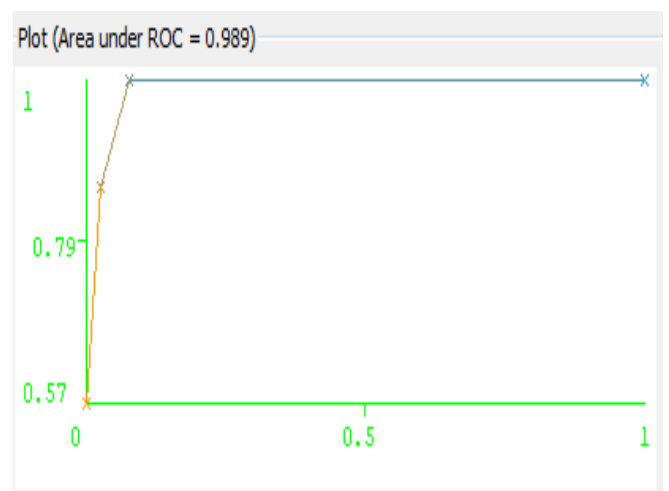


Fig 8: Class value E

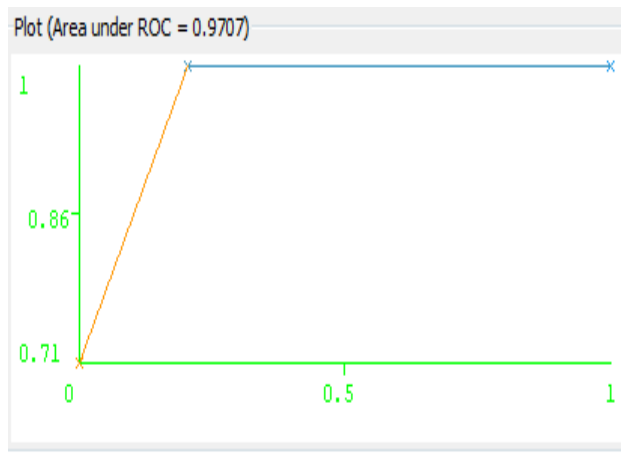


Fig 9: Class value MD

5. CONCLUSION

In this work, J48 algorithm is deployed for the classification of EEG based diseases such as: dementia, Schizophrenia, ADHD, epilepsy and mood disorder. The data is collected from hospitals and experts.

J48 algorithm generates rules and decision tree for the classification of EEG based diseases. The sensitivity for all diseases lies in the range of 70-100 % and specificity lies in the range of 94-100%. J48 algorithm in Weka generate decision tree that help us to diagnose the disease on the basis of taken attributes/symptoms. And the ROC curve compares the two operative characteristic that are true positive rate and false positive rate.

6. REFERENCES

- [1] M.Teplan,"fundamentals of eeg measurement", measurement science review, volume 2, section 2, 2002.
- [2] B. Pandey and R.B. Mishra, "Case-based reasoning and data mining integrated method for the diagnosis of some neuromuscular disease", IJMEI 3(1): 1-15 (2011).
- [3] B. Pandey and R.B. Mishra," An integrated intelligent computing method for the detection and interpretation of ECG based cardiac diseases", IJKESDP 2(2): 182-203 (2010).
- [4] B. Pandey and R.B. Mishra , "Data mining and CBR integrated methods in medicine: a review", IJMEI 2(2): 205-218 (2010).
- [5] B. Pandey and R.B. Mishra , "Knowledge and intelligent computing system in medicine", Comp. in Bio. And Med. 39(3): 215-230 (2009).
- [6] Gangwar et al ,," Intelligent Computing Methods for The Interpretation of Neuropsychiatric Diseases Based on Rbr-Cbr-Ann Integration", International Journal of Computer and Technology Vol 11, No.5 , ISSN 2277-3061, pp. 2490-2511.
- [7] Glover et al.,,"Knowledge based interpretation of bioelectrical signals: Examples in eeg analysis", Proceedings of the 25th Annual International Conference of the IEEE EMBS Cancun, Mexico September 17-21, 2003. 1988 MAPLE PRESS.
- [8] Wu et al.,," EEG classification based on artificial neural network in brain computer interface", LSMS/ICSEE 2010, part 1, CCIS 97, pp.154-162.
- [9] M.Ganesan, P.S.Sathidevi and K.P.Indradevi " A novel approach for the analysis of epileptic spikes in EEG", International Conference on Computational Intelligence and Multimedia Applications 2007 ,pp.297-301.
- [10] Li et al.,," An Effective Classification Approach for EEG-based BCI System", 2011 Sixth International Conference on Image and Graphics.
- [11] Patil and Sherekar," Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification", International Journal of Computer Science and Applications Vol. 6, No.2, Apr 2013 .
- [12] N. Bharhava, G. Sharma, R. Bhargava and M. Mathuria, "Decision Tree Analysis on J48 Algorithm for Data Mining "International Journal of Advanced Research in Computer Science and Software Engineering, volume 3, issue 6 june 2013.
- [13]<http://www.opentox.org/dev/documentation/components/j48>.