

# **Implementation of Decision Tree Technique in the Diagnosis of Psychiatric Disorder**

**Preeti Singh**  
M.Tech. Research Scholar  
Department of CSE  
AIET, Lucknow

**Atma Prakash Singh**  
Asst. Prof.  
Department of CSE  
AIET, Lucknow

**Shafeeq Ahmad**  
Professor  
Department of CSE  
AIET, Lucknow

## **ABSTRACT**

In this work, the Electro-encephalogram and Magnetic Resonance Imaging (MRI) features along with physical, cognitive and psychological features combined together to diagnose the psychiatric disorder. The disorders are taken for diagnoses are Hyperactivity Disorder (HD), Memory Disorder (MD), Anxiety Disorder (AD), Obsession Disorder (OD) and Alzheimer (AZ). The diagnosis procedure of the disorder depends upon the different types of features. In this research paper we are considering 20 factors divided into five categories and with the help of decision tree based C5.0 method to know the important factors in the diagnosis of disorder. The design of decision tree for C5.0 method in consideration to Clementine tool is also matched with the manual calculation. The decision tree model supports doctor's to get easier way to interpret and diagnoses disorder on the basis of important factors.

## **Keywords**

Artificial Intelligence, Decision Tree, Medical Computing, Psychiatric Disorder, Computational Intelligence.

## **1. INTRODUCTION**

Psychiatry is the knowing and application of the field that joints brain disorder and behavior disorder. In this paper, said about the underlying belief of the regulation to be the brain and thought, of mind and human body, and of mental body and physical body. Psychiatry tries to link the artificial limitations between neurology and psychiatry in order to treat the multitude of clinical manifestations of the singular brain. Psychiatry is primarily focused on the assessment and treatment of the cognitive, behavioral, and mood symptoms of patients with neurological disorders [1, 13, 16]. However, an equally important focus for psychiatrists is the understanding of the role of brain dysfunction in the pathogenesis of primary psychiatric disorders. Therefore, not only does psychiatry bring the psychiatric assessment and treatment of psychotic to the neurology arena, it also does neuropsychiatry bring the psychiatric assessment and treatment of psychotic or psychiatric symptoms to the neurology arena, it also returns clinicians to the objective rigors of physical diagnosis and testing, which are often not practiced in psychiatry today [2, 14, 17].

Psychiatric diagnosis is based on elicitation of clinical symptoms, identification of psychiatric syndromes, construction of a diagnosis, use of different types of laboratory tests and imaging to support or exclude specific disorder diagnoses, and identification of the primary etiology of the behavioral disturbance. In some cases, longitudinal assessment and careful monitoring of treatment responses may be necessary to clarify obscure diagnosis. Treatment depends on accurate disorder diagnosis [3, 15].

The psychiatric examination is a specialized clinical evaluation that supports clinical diagnosis and management in behavioral diagnosis of disorder. The elements of the psychiatric examination include the psychiatric evaluation described in the Psychiatric Association Practice guidelines for the psychiatric evaluation; the neurologic examination; and the behavioral mental status examination. This examination includes imaging and neurophysiologic testing [4, 14].

Data mining is a process of extracting implicit, previously unknown, and potentially useful information from complex and large amount of databases. The mining method is used to extract and analyze meaningful patterns and correlations in a large relational database. Various data mining techniques such as: instance-based learning, decision trees, neural networks, rule induction and sensitivity analysis have been successfully applied in medical problem domains. This research work describes the application of data mining technique, i.e., decision tree (C5.0 technique) in the diagnosis of five psychiatric disorders. The data mining problems may involve hundreds or even thousands of fields that can potentially be used as predictors. As a result, a great deal of time and effort may be spent in examining which fields or variables to be included in the model. The combined model helps in choosing the factor that has greater influence on predictions [5].

The different data mining method used to know the importance and contribution of different factors involved in the diagnosis process are further using for reducing the parameters and applying reduced set of parameters to generate decision tree model and rule set in the diagnosis of psychiatric disorder. Data mining methods gives an idea to know the importance and contribution of different symptoms in the diagnosis process but unable to produce the confidence factor, cumulative confidence factor and reasoning [6]. So, in this research work decision tree model is design and implement to form decision tree based on the training data of 20 factors with their corresponding disease's output class. So that physician can easily take decision on the basis of importance factors.

The data mining method using C5.0 technique has been implemented for the diagnosis of psychiatric disorder in this research work. The paper is organized as: Section 1 described about the introduction. Section 2 describes related work in the research area. Sections 3 define problem definition. Section 4 shows results and discussion and finally Section 5 presents conclusion.

## **2. RELATED WORK**

In this research work [7], Rule based technique applied using 38 factors i.e., symptoms to complete diagnosis to further provide support to psychiatrists to make diagnosis. These

Rule based technique applied using JAVA language. Rule based calculating confidence factor.

In this research paper [8], data mining based technique in the automatic Electro-encephalogram detection of transient events. Several classifications are generated based on spikes, physical activity, eye movement activity and alpha movement. The implementation of the models contains clustering methods of several events, feature recording from each segment and cluster, feature extraction and rule mining and different classification. The overall efficiency in this approach is 84.35%. The major benefit of this method is that it provides explanation for the diagnosis decisions.

In this paper [9], research work show that application of C5.0 technique used for the creation of decision structure to find the possibility of depression. The data take in this study from Survey. The input parameters taken for this work are regional, medical history related and economic features. The selection criterion uses Statistical techniques comprised of different regression techniques. The result shows 81.6% accuracy for the data. The usefulness of this technique may be assist for doctors to monitor and vigilant with possibility of depression.

This research work [10] contains of data mining method to test the imaging data for prevention of disease. Methods that classify make analysis correlation analysis and clustering brain formation for different types of data sets. Two types of brain imaging data: structural and functional used for this study. Statistical methods used to discovery of interesting associations and patterns between brain images and other clinical data.

### 3. PROBLEM DEFINITION

To describe the problem firstly, the information (symptoms) of different psychiatric disorder from different resources were collected. On the basis of above information a total of 750 records of five psychiatric disorders were collected and created in consultation with doctors from the neurology and psychiatry departments. The above 750 records includes: 130 of Hyperactivity disorder, 161 of Memory Disorder, 165 of

Anxiety Disorder, 145 of Obsession disorder and 149 of Alzheimer. There are total twenty factors that are divided into two major groups: The former is divided into: muscular, motor activity, cognitive and psychological parameters and the latter is divided in EEG signal and MRI characteristics. The sample data based on above criteria is shown in Table 1.

All the twenty parameters are coded in the symbol High (H) or Medium (M) or Low (L). The criterion for coding is based on the signs and symptoms rating. The symptom rating is based on the following criteria i.e. onset, frequency and intensity [11, 13]. Based on these criteria the severity of the sign and symptom is defined i.e. mild, moderate and sever. We have defined the three severity level based on the previously defined rating scales i.e. low (mild), medium (moderate) and high (Sever). These three severity levels are used to assess the patient by medical evaluator and to calculate the rating of signs and symptoms. The sign and symptom rating is defined as follows:

- Onset [Low <= 1 year < Medium <= 2 year < High <= Remaining]
- Frequency [Low <= 2 times/day < Medium <= 4 times/day < High <= Remaining]
- Intensity (0.0 < Low <= 0.3 < Medium <= 0.7 < High <= 1.0]
- Rating sale
  - Mild (LOW) - (0.0 < L <= 0.3)
  - Moderate (MIDEUM) - (0.3 < M <= 0.7)
  - Sever (HIGH) - (0.7 < H <= 1.0)
- Rating = (Onset value + Frequency value + Intensity value) / 3

Table 1. Disorder data

Diseases	Physio Parameters																				
	Muscular physiology Factors							Cognitive Factors					Psychological Factors					EEG		MRI	
	Muscular			Motor				LD	RS	H A	A G	SW	DE	FL	PL	OL	TL	FL	AC	CG	
	OS	MW	MO	LO	UT	CS	W														
HD	H	H	M	M	H	H	L	H	M	L	L	L	L	H	L	L	L	M	L	L	
MD	H	H	M	M	H	H	L	H	M	L	L	L	L	H	L	L	L	M	L	L	
AD	H	M	M	M	H	H	L	H	M	L	L	L	L	H	L	L	L	M	L	L	
OD	H	H	M	M	H	H	L	H	L	M	L	L	M	H	L	H	L	M	L	L	
AZ	H	H	L	L	H	H	L	H	M	L	H	L	L	H	L	L	L	M	L	L	

## 4. RESULTS

The experimentations are performed with the data mining method, i.e., decision tree (C5.0 algorithm) on the data set of 750 records. The experiment with their corresponding result is giving below. In the decision tree model given below 750 records have been used and among which 500 records are used for training and 250 records are used for testing the accuracy of model. In this section firstly we are generating decision tree and rule set with the help of Clementine software as shown in Figure 1 and 2 & secondly we are calculating different values and showing detailed steps for verification of decision tree generated by Clementine software. In the next section we are generating decision tree and rule set with the help of Clementine software.

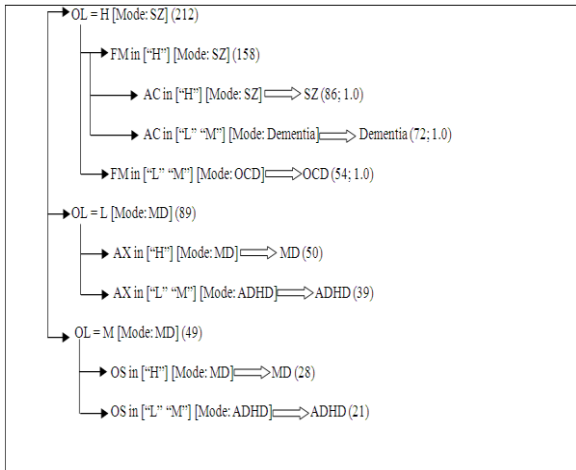


Figure 1: Decision tree for disorder

In the above tree structure in Figure 1 the right side of each rule, two values are given which are known as instances and confidence. Instance gives information about the number of records to which the rule applies and the confidence gives proportion of those records for which the rule is true.

The rules in the Figure 2 have been generated by the program based upon the tree structure as shown in the Figure 1. There are 7 sets of rules each containing 5 to 6 lines for description in Figure 2. Each set of rule reads like in order of lines as such: the first line says about the number of rules in a particular disease.

Rules	
Rules for ADHD - contains 2 rule(s)	and AX in [\"H\"] then MD
Rule 1 for ADHD (39; 1.0)	Rule 2 for MD (28; 1.0)
if OL in [\"L\"]	if OL in [\"M\"]
and AX in [\"L\" \"M\"]	and OS in [\"H\"]
then ADHD	then MD
Rule 2 for ADHD (21; 1.0)	Rules for OCD - contains 1 rule(s)
if OL in [\"M\"]	Rule 1 for OCD (54; 1.0)
and OS in [\"L\" \"M\"]	if OL in [\"H\"]
then ADHD	and FM in [\"L\" \"M\"]
Rules for Dementia - contains 1 rule(s)	then OCD
Rule 1 for Dementia (72; 1.0)	Rules for SZ - contains 1 rule(s)
if OL in [\"H\"]	Rule 1 for SZ (86; 1.0)
and FM in [\"H\"]	if OL in [\"H\"]
and AC in [\"L\" \"M\"]	and FM in [\"H\"]
then Dementia	and AC in [\"H\"]
Rules for MD - contains 2 rule(s)	then SZ
Rule 1 for MD (50; 1.0)	
if OL in [\"L\"]	

Figure 2: Rules for diagnosing psychiatric disorders

### 4.1 Steps for Verification

The calculated values measure on the basis of 750 data sets to create decision tree. Different types of values are display to set the splitting criteria for the attribute to create the decision using Equations one to five [12]. These values are shown in Table 2, the first column of table shows label of the tree during upper to lower movement in generated decision tree, second column of the table shows maximum value attribute name, third column in the table shows gain value for these attribute finally forth column in the table represent GainRatio values for these attributes. The value represented in terms of Info (D) shows the expected information required to classify a tuple in training records.

At first level all twenty factors values calculated for 750 training data sets. These calculated values are represented in second row of the Table. These values are for highest vales factors in the first level.

$$\text{Info (D)} = 2.296, \text{ for twenty parameters.}$$

$$\text{Info (D)} = - \sum_{i=1}^m p_i \log_2(p_i) \quad (1)$$

$$\text{InfoA (D)} = \sum_{j=1}^v \frac{|D_j|}{|D|} \times \text{Info} (D_j) \quad (2)$$

$$\text{Gain (A)} = \text{Info (D)} - \text{InfoA (D)} \quad (3)$$

$$\text{SplitInfoA (D)} = - \sum_{j=1}^v \frac{|D_j|}{|D|} \times \log_2 \left( \frac{|D_j|}{|D|} \right) \quad (4)$$

$$\text{GainRatio(A)} = \frac{\text{Gain (A)}}{\text{SplitInfoA(A)}} \quad (5)$$

Table 2. Calculated values

Label	Highest value attribute	Gain (Att)	Gain Ratio (Att)
1	OL	1.487	1.667
2	FM	1.053	3.078
2	OS	0.765	4.421
2	AX	1.086	4.981
3	AC	1.186	4.763

In this section we have calculated different values to decide the splitting attribute at different levels to generate the decision tree for 500 records.

The above decision tree using C5.0 method tested for 250 data set of psychiatric patient disorder. Out of 250 test data set, 240 data sets are correctively classified by method and efficiency showed 96% for psychiatric disorders as shown in Figure 3.

Method	Total test data	Correctly predicted	Wrongly predicted	Accuracy
Decision Tree	250	240	10	96%

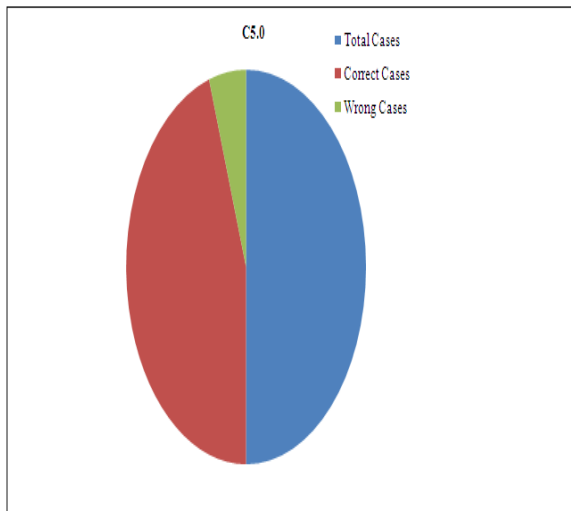


Figure 3: Performance of C5.0 on test data

## 5. CONCLUSION

Decision tree technique is useful in finding the important and considerable factors that involves in the diagnosis of disorder. On the basis of experimentation results, C5.0 method form decision tree structure and rules. Out of twenty factors only five factors finally used by C5.0 method for creation of tree and rule. These parameters are OL, FM, AC, AX and OS. The C5.0 techniques applied criteria of information gain to making the braches of tree. Based upon the division in tree OL, FM, AC, AX and OS produce maximum information gain and gain ration at different levels. Finally, the efficiency of C5.0 method tested for 250 data. The decision tree based C5.0 technique gives 96% accuracy in terms of diagnosis result. The benefit of using decision tree based C5.0 technique is that it makes an easy and quick ways for the psychiatrists for diagnosis of disorder. The value represented in this way (A; B) in the right of rules shows the number of rules applies (A: instance) and the amount of those data for which the rule is correct (B: confidence). Therefore, data mining based C5.0 technique for training records also creates and produces confidence factor value for diagnosis of disorder.

## 6. REFERENCES

[1] Brown, G.W. and Harris, T. eds., 2012. Social origins of depression: A study of psychiatric disorder in women. Routledge.

[2] Swanson, J.W., Holzer III, C.E., Ganju, V.K. and Jono, R.T., 1990. Violence and psychiatric disorder in the community: evidence from the Epidemiologic Catchment Area surveys. *Psychiatric Services*, 41(7), pp.761-770.

[3] Fergusson, D.M., Horwood, L.J. and Lynskey, M.T., 1996. Childhood sexual abuse and psychiatric disorder in young adulthood: II. Psychiatric outcomes of childhood sexual abuse. *Journal of the American Academy of Child & Adolescent Psychiatry*, 35(10), pp.1365-1374.

[4] Ford, T., Vostanis, P., Meltzer, H. and Goodman, R., 2007. Psychiatric disorder among British children looked after by local authorities: comparison with children living in private households. *The British Journal of Psychiatry*, 190(4), pp.319-325.

[5] Mining, D., 2001. Concepts and Techniques. Jiawei Han and Micheline Kamber.

[6] Strub, R.L. and Black, F.W., 1993. The mental status examination in neurology. FA Davis Company.

[7] Haerer, A.F., 1992. DeJong's the neurologic examination (No. 1992). Lippincott Williams & Wilkins.

[8] Strub, R.L. and Black, F.W., 1993. The mental status examination in neurology. FA Davis Company.

[9] Trzepacz, P.T. and Baker, R.W., 1993. The psychiatric mental status examination. Oxford University Press.

[10] Weintraub, S., 1985. Mental state assessment of young and elderly adults in behavioral neurology. *Principles of behavioral neurology*, pp.71-123.

[11] Exarchos, T.P., Tzallas, A.T., Fotiadis, D.I., Konitsiotis, S. and Giannopoulos, S., 2006. EEG transient event detection and classification using association rules. *Information Technology in Biomedicine, IEEE Transactions on*, 10(3), pp.451-457.

[12] Herrmann, C.S. and Demiralp, T., 2005. Human EEG gamma oscillations in neuropsychiatric disorders. *Clinical Neurophysiology*, 116(12), pp.2719-2733.

[13] Gangwar, M., Mishra, R.B. and Yadav, R.S., 2014. Classical and intelligent computing methods in psychiatry and neuropsychiatry: an overview. *International Journal of Advanced Research in IT and Engineering*, 3(12), pp.1-24.

[14] Gangwar, M., Yadav, R.S. and Mishra, R.B., 2012. March. Semantic Web Services for medical health planning. In *Recent Advances in Information Technology (RAIT), 2012 1st International Conference on* (pp. 614-618). IEEE.

[15] Gangwar, M., Mishra, R.B. and Yadav, R.S., 2013. Intelligent Computing Methods for The Interpretation of Neuropsychiatric Diseases Based on Rbr-Cbr-Ann Integration. *International Journal Of Computers & Technology*, 11(5), pp.2490-2511.

[16] Servan-Schreiber, D., 1986. Artificial intelligence and psychiatry. *The Journal of nervous and mental disease*, 174(4), pp.191-202.

[17] Montague, P.R., Dolan, R.J., Friston, K.J. and Dayan, P., 2012. Erratum: Computational psychiatry. [*Trends in Cognitive Sciences* 16 (2012), 72-80](DOI: 10.1016/j.tics.2011.11.018). *Trends in Cognitive Sciences*.