# Analysis of Brain Data Attributes To Detect the Prevalence of ADHD Using R Programming

[1]E. Radhamani, [2]K. Krishnaveni
[1]Research Scholar, [2]Associate Professor
Sri S.R.N.M.College, Sattur, Tamil Nadu, India
[1]radhamaniphd2015@gmail.com, [2]kkveni69@gmail.com

**Abstract—** The prevalence of Attention Deficit Hyperactivity Disorder due to brain damage or abnormality is emphasized in this paper. The data samples collected from various mind care hospitals are categorized and ADHD affected samples alone are taken and classified based on the level of Brain Damage. The Brain Damage Questionnaire with 15 attributes is prepared, the selected data samples are interviewed and each attribute in the Questionnaire is rated on a four point scale. The level of Brain Damage(BD) is identified based on the total Brain Damage Rating Value (BDRV). The attributes are reduced and the samples are classified into three classes like mildBD, moderateBD and severeBD using C5.0 Decision Tree classifier. The data sample is said to has mild brain damage if BDRV < 10; moderate for 10<= BDRV<=15 and if BDRV > 15then the sample is severely affected with Brain damage. The results of the classifier are analyzed using the performance metrics and the value of Fscore shows that 50 % of the ADHD samples are having moderate or severe brain damage. So, this research work predicts that the child will definitely have ADHD symptoms if it has brain damage or injury or abnormality.

**Keywords:** Attention Deficit Hyperactivity Disorder, Brain Damage, Brain Damage Rating Value, Decision Tree classifier.

## 1. Introduction

Computational and statistical assessment of data has become the most valuable resource in every field like social media, education, agriculture, medical, science etc. These fields are generating large volume of data day to day known as big data, since there are drastic changes in human lifestyle, environment, food etc. Among them many researches are to be performed in the medical field with large growth of data since there is a need to find best diagnosis and treatment plans.

Attention Deficit Hyperactivity Disorder (ADHD) is the neurological disorder caused by various factors. It has been diagnosed that the symptoms of ADHD may vary with Age and Gender of the children [1]. Initially, the data samples are classified into NOADHD and ADHD with different levels by Data Mining Classifiers like J48, Naïve Bayes, MLP (Multilayer Perception Network) and SVM classifiers. Once the ADHD samples are identified, the questionnaire is prepared with various symptoms related to the attributes of Brain damage or abnormality. Based on this questionnaire, the data samples are collected and the research tool R_programming is attempted to analyze the data. The main aim of this research work is to predict that ADHD may occur due to the cause of brain damage or abnormality or not.

This paper is organized into IV sections. The proposed Data Analysis methodology is discussed in Section II. The experimental analysis and evaluation of ADHD data samples using R-programming tool is provided in Section III. The conclusion and future direction of work is presented in Section IV.

## 2. Proposed Data Analysis Methodology

Data analysis is the process for acquiring raw data, cleaning, inspecting and modeling data with the goal of useful information and transforming into information for decision-making by users. The type of user's may vary from business analysts, engineers, end-user, scientist, etc., and the data analytic process is represented in Figure 1.

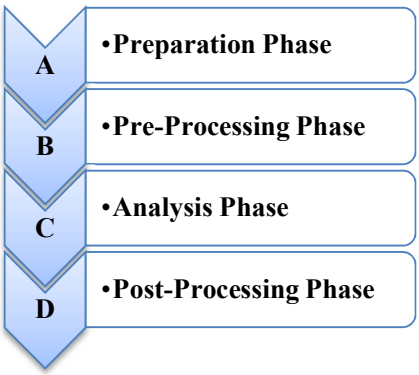


**Figure 1: Data Analytic Process**

### 2.1. Preparation Phase

The preparation phase comprises of Questionnaire preparation, Data collection and selection. The data preparation initiates with the preparation of a questionnaire focusing the Brain damage, Brain abnormality and Medical characteristics of the children.

#### 2.1.1. Preparation of Brain Damage Questionnaire

The Brain damage Questionnaire consists of 15 attributes which are the symptoms of Brain Damage(BD) and each attribute is rated on a four point scale from 0 to 3(0-Never, 1-Rare, 2-Often, 3-Very Often) as shown in Table I. The severity of the brain damage is measured from these rating measures.

**Table I: BRAIN DAMAGE QUESTIONNAIRE**

| S.No | Symptoms | Never | Rare | Often | Very Often |
|---|---|---|---|---|---|
| 1. | Headache | | | | |
| 2. | Vomiting | | | | |
| 3. | Convulsions | | | | |
| 4. | Abnormal dilation of the eyes | | | | |
| 5. | Inability to awaken from sleep | | | | |
| 6. | Weakness in extremities | | | | |
| 7. | Trouble with memory | | | | |
| 8. | Trouble with concentration | | | | |
| 9. | Trouble with attention or thinking | | | | |
| 10. | Loss of coordination | | | | |
| 11. | Confusion | | | | |
| 12. | Aggressive | | | | |
| 13. | Abnormal Behavior | | | | |
| 14. | Slurred Speech | | | | |
| 15. | Coma or Other disorders of consciousness | | | | |

#### *2.1.2.* Collection of Data Samples

Initially 105 data samples were collected from mind care hospitals, schools and from the parents and classified as ADHD and NOADHD samples by machine learning classifiers [2]. Out of 105 samples, 44 ADHD data samples were identified and the children were directly interviewed with the brain damage Questionnaire. A four-point rating scale ranging from 0 to 3 is awarded for each question and for each data sample the total score BDRV(Brain Damage Rating Value) to identify the ADHD levels like Mild, Moderate and Severe are calculated. The collected data samples along with

their attributes and rating score are stored in the Excel Spreadsheet which can be further analyzed and classified by R Programming Research tool [6].

### 2.2. Pre-Processing Phase

The pre-processing phase deals mainly with the syntax analysis and data correction. It includes data cleaning, filtering, correction, standardization, transformation and completion of data. In this paper, after collecting the input data, the BDRV score is calculated and the corresponding BD level is extracted. The sample data set is preprocessed to remove duplicate records; missing data and inconsistent data. During preprocessing, the total number of data sample attributes in the ADHD data set is reduced from 15 to only five attributes namely ID, Gender, Age, BDRV score, BD(Brain Damage) Level (Mild, Moderate and Severe) as listed in the Table II.

**Table II: Reduced ADHD Brain Damage Data Set Attributes**

| Number | Attribute Name | Data Type |
|---|---|---|
| 1 | ID | Numeric |
| 2 | Gender | Numeric |
| 3 | Age | Numeric |
| 4 | BDRV | Numeric |
| 5 | BD Level | Numeric |

### 2.3. Analysis Phase

The analysis phase restrains visualization, correlation, regression, forecasting, classification and clustering of data with similar patterns. The main aim of this research work is to identify the risk factor of ADHD due to brain damage. I.e., it has to be identified that the main cause of ADHD is brain damage or not. In this work, C5.0 decision tree classifier is proposed to classify the ADHD data samples into three classes based on the level of brain damage.

#### 2.3.1. C5.0 Decision Tree Classification

Decision Tree based C5.0 classifier is proposed to classify the data samples collected by using Brain Damage Questionnaire. Based on the BDRV score, the severity level of BD (Mild, Moderate and Severe) is identified. The following classification rules are intended for classification.

*IF BDRV <10 THEN*
*BD Level: Mild*
*IF BDRV >=10 AND BDRV <=15 THEN*
*BD Level: Moderate*
*IF BDRV > 15 THEN*
*BD Level: Severe*

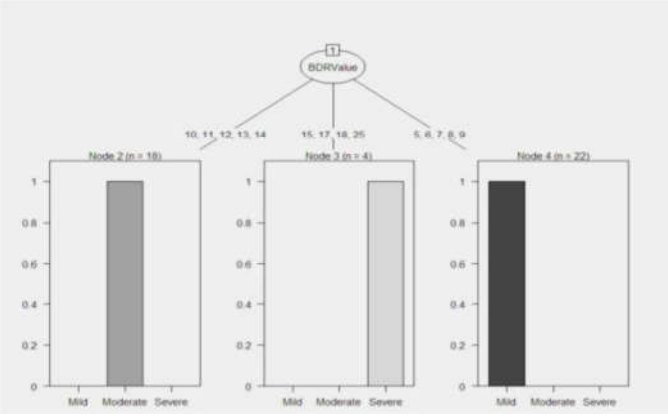


**Figure 2. Decision tree for Brain Damage Dataset**

The decision tree used by C5.0 classifier for Brain Damage Data set is shown in Figure 2. If the BDRV score is greater than 15 then the data sample is said to have ADHD due to Severe BD. If the BDRV score is greater than or equal to 10 and less than or equal to 15 then the data sample is said to have Moderate BD. Otherwise, the BD level is Mild.

### 2.4. Post-Processing Phase

The post processing phase is the final and important stage of data analytics. The post processing phase consists of data interpretation, prediction of data and evaluation of data.

## 3. Experimental Analysis & Results

In this research work, R programming tool is used to predict the prevalence of ADHD due to brain abnormality or damage. It contains the following steps:

- Data Pre-processing and Attribute Selection
- Classification
- Prediction

### 3.1. Data Pre-processing and Attribute Selection

The ADHD brain damage dataset with 44 instances is created and given as input to R Studio tool. The data samples are preprocessed and the description of the reduced attributes is given in Table III. The Brain Damage data set contains five attributes: ID, Gender, Age and BDRV (Brain Damage Rating Value) which are used to classify the samples into three classes of BD like mild, moderate and severe.

**Table III: Brain Damage Data Set Description**

| Attribute Name | Data Type | Possible Values | Values in data Set |
|---|---|---|---|
| ID | Numeric | 1-100 | (1 - 44) |
| Gender | Numeric | Male (0)<br>Female(1) | (0 , 1) |
| Age | Numeric | 2 - 15 | (2 - 15) |
| BDRV (Brain damage Questionnaire 15 items –Table II) | Numeric | 0 - Never<br>1 - Rare<br>2 - Often<br>3 - Very Often | (0 - 45) |
| BD Level | Numeric | Mild<br>Moderate<br>Severe | BDRV <10 Level=Mild<br>BDRV >=10 & BDRV <15 Level=Moderate<br>BDRV >=15 Level=Severe |

### 3.2. Classification

The main goal of the classification is to predict the target class. Primarily, the data samples are classified into ADHD and NOADHD and ADHD data samples are further classified based on different levels of BD as Mild Decision Tree, Moderate Decision Tree and Severe Decision Tree).

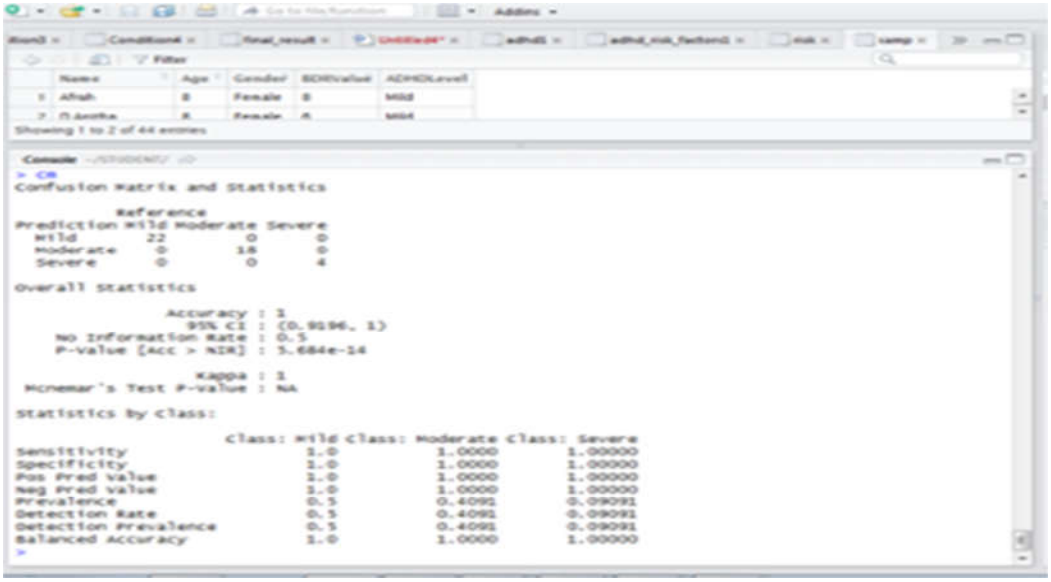


**Figure 4. Output of Decision Tree Based C5.0 Classifier**

From Figure 4, it has been identified that 4 samples are affected with severe brain damage, 18 by moderate and 22 have mild damage. The analysis shows that out of 44 ADHD data samples, 50% of the samples are having severe brain damage or abnormality. It emphasis that the occurrence of ADHD depends on brain damage or abnormality.

### 3.3. Prediction

After the data samples are classified into various levels of BD, any sample can be predicted to have ADHD due to brain damage or brain abnormality based on some performance metrics listed in Table IV. These metrics are used to evaluate the performance of the classifier. Sensitivity and specificity are the characteristics of the test, whereas prevalence values depend on the disease predictive in the population. The Line Chart of the performance metrics is shown in Figure 5.

**Table IV: Performance Metrics for Various Brain Damage Class**

| Metrics | Class | | |
|---|---|---|---|
| | Mild | Moderate | Severe |
| Sensitivity | 1.0000 | 1.0000 | 1.0000 |
| Specificity | 1.0000 | 1.0000 | 1.0000 |
| Prevalence | 0.5000 | 0.4091 | 0.0909 |
| Accuracy | 1.0000 | 1.0000 | 1.0000 |

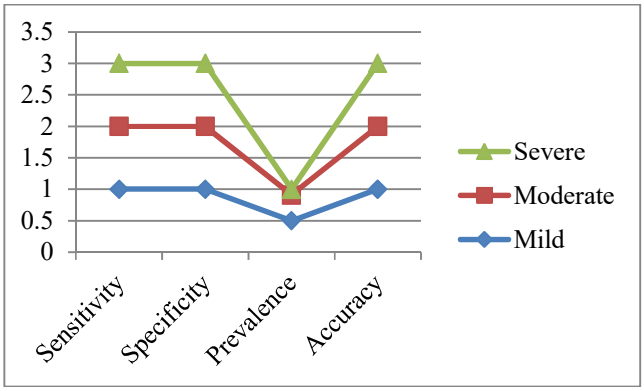


**Figure 5. Performance Metrics**

From the confusion matrix listed in Table V, it has been identified that out of 44 instances, the C5.0 algorithm classifies all the samples accurately (22- Mild, 18-Moderate, 4-Severe) for BD and the pictorial representation is given in Figure 6.

**Table V: Confusion Matrix for Brain Damage Data Set**

| Prediction | Mild | Moderate | Severe |
|---|---|---|---|
| Mild | 22 | 0 | 0 |
| Moderate | 0 | 18 | 0 |
| Severe | 0 | 0 | 4 |

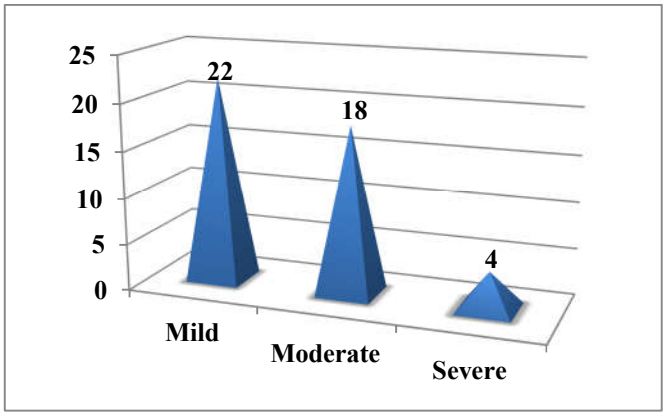


**Figure 6. Level of ADHD Children**

From these experiments, it has been proved that if a child has any brain damage or injury or brain abnormality, it will definitely be affected with ADHD.

## 4. Conclusion

In this research work, the ADHD data samples are taken, interviewed, brain damage attributes for each sample is collected, classified using C5.0 Decision Tree Classifier and the results are analyzed. The results show that brain damage is one of the main causes for the prevalence of ADHD ie, a child with brain abnormality will definitely have the symptoms of ADHD. So the research work is planned to continue in future to identify the abnormal brain parts by affording the volumetric analysis on MRI and FMRI brain images of ADHD children.

## Acknowledgments

**E.Radhamani** has received her MCA & M.Phil degree in Computer Science from Sri S.Ramasamy Naidu Memorial College, Sattur affiliated to Madurai Kamaraj University, Madurai, Tamil Nadu. She is pursuing Ph.D in Computer Science. She is also working as a Project Fellow for UGC Major Research Project. Her area of interests are Medical Image processing and Data mining.

**Dr.K.Krishnaveni**, is an Associate Professor and Head of the Department of Computer Science, Sri S.Ramasamy Naidu Memorial College, Sattur, India. She has 25 years of teaching experience. She received her B.E. degree in Computer Science and Engineering from Bharathiar University, Coimbatore, India in 1990 and M.Tech. degree in Computer and Information Technology from Center for Information Technology and Engineering of Manonmaniam Sundaranar University, Tirunelveli, India in 2004 and received Ph.D degree from the same University. Currently her research interests are Medical Image

Processing, Data mining, and Mathematical Morphology. She is a member of IEEE and Editor of various reputed journals.

## References


[1] Radhamani E, Krishnaveni K, "Diagnosis of ADHD using Statistical measures", International Journal of Engineering Research in Computer Science and Engineering (IJERCSE), ISSN : 2394-2320,Vol. 5, Issue 3, March 2018.

[2] Radhamani E, Krishnaveni K, "Prognosis of ADHD using R Programming and MATLAB Tools", International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, ISO 9001:2008, Vol 7, 2017.

[3] Radhamani E, Krishnaveni K, "Diagnosis and Evaluation of ADHD using MLP and SVM Classifiers", Indian Journals of Science and Technology, Vol 9(19)/93853, ISSN 0974-5645, 2016.

[4] Radhamani E, Krishnaveni K, "Diagnosis and Evaluation of ADHD using Naïve Bayes and J48 Classifiers", International Conference on Computing for Sustainable Global Development (INDIACom), 978-9-3805-4421-2/16/$31.00 © 2016 IEEE.

[5] Shwetha G, Visali Lakshmi P R, "Analysis of Medical Image and Health Informatics Using Big data", 2017 Third International Conference On Science Technology Engineering and Management (ICONSTEM), 978-1-5090-4855-7/17/$31.00 ©2016 IEEE.

[6] Tran Duc Chung, Rosdiazli Ibrahim, "Fast Approach for Automatic Data Retrieval using R Programming Language", Robotics and Manufacturing Automation (ROMA), ISBN: 978-1-5090-0928-2, 2016.

[7] Ms. Ayushi Malviya, Mr. Amit Udhani, "R-Tool: Data Analytic Framework for Big Data", Colossal Data Analysis and Networking (CDAN), ISBN: 978-1-5090-0669-4, 2016.

[8] Sunghae Jun, "Patent Big Data Analysis by R Data Language for Technology Management", International Journal of Software Engineering and Its Applications, ISSN: 1738-9984, Vol. 10, No. 1, 2016.

[9] Sanchita Patil, "WhatsApp Group Data Analysis with R", International Journal of Computer Applications (0975 – 8887), Volume 154 – No.4, November 2016.

[10] Ruizhu Huang, Weijia Xu, "Performance Evaluation of Enabling Logistic Regression for Big Data with R", IEEE International Conference on Big Data, 978-1-4799-9926-2/15/$31.00 ©2015 IEEE.

[11] Ilir Keka, Mentor Hamiti, "Load Profile Analyses Using R Language", Information Technology Interfaces (ITI), ISBN: 978-953-7138-32-5, 2013.