International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor (2018): 7.426

Detection of ADHD using Machine Learning Algorithms

Rohit Kale

IT Department, Vishwakarma Institute of Technology, Bibwewadi, Pune, India

Abstract: ***Attention deficit hyperactivity disorder (ADHD) is a standout amongst the most widely recognized disorder in school-age kids. To date, the determination of ADHD is for the most part subjective and investigations of target symptomatic strategy are of awesome significance. Although numerous endeavors have been made as of late to explore the use of structural and functional images of brain, for diagnosis, out of which a few are relevant for ADHD. Since the available dataset is unlabeled, we present a programmed grouping structure (k-means) to create clusters in view of MRI image of ADHD patients and normal subjects and present in detail highlight determination, and classifier preparing techniques. Classification algorithm K-Nearest Neighbor creates a classifier which classifies the input into the respective category.***

Keywords: ADHD, s-MRI, f-MRI, Machine Learning, k-means, anomaly, random forest

1. Introduction

ADHD is a standout amongst the most widely recognized disorder in school-age kids. Currently, Clinical analysis is used for detection of ADHD. Our method proposes more convenient and advanced way for the same. In clinical analysis, more human intervention is needed, which may create a possibility of human error. Clinical analysis takes more time as the period over which the patient is being observed may vary. Machine learning makes it simpler and accurate. Use of machine learning makes sure that error generated by the program is minimum. Working with a larger data set ensure more accuracy. Direct analysis can be made from MRI which makes it easier than to physically do clinical analysis. Clustering technique ensure that data being clustered is the best fit for the data set. Anomaly Detection helps find irrelevant data so. Hypothesis obtained makes sure that patient is being classified with least error. Output obtained from hypothesis is based on numerical values which are more reliable.

2. Literature Survey

Machine learning has been used many times as a solution in a medical fields. Various research papers have been written which use machine learning as a key concept for a solution. number of diseases can be detected using machine learning.
One approach would be to combine the clinical analysis with machine learning. Clinical analysis include factors such as cognitive test ,solutions provide by students to cognitive tasks and images which specify the emotional state of the patient. This gives a brief information about behavioral conduct cognitive performance and emotional state. Face detection and feature extraction can be done by using machine perception toolbox. Emotions can be classified by using machine learning algorithms like Knn.This approach can classify data more accurately. Emotion classification can help boost the results and a better view can be obtained at the data using this. Solutions are not a part of analysis as they require biological aspects. Focusing only on ADHD is a prime factor as patient may also have other disorders.[1]

ELM can also be used for classification of ADHD using MRI. Usual dataset consists of 200 to 300 subjects. Dataset is made of subjects having ADHD and normal people. High Resolution multidimensional images are used . Multiple brain features such as cortical thickness and cortical volume ae taken into consideration as input features. Out of 340 cortical features, from 68 brain segments, 5 basic cortical features were considered. For feature extraction of cortical thickness, software called freesurfer was used. F-score and SFS methods were used. Support Vector Machine algorithm. Higher accuracy are being obtained by using machine learning than by using traditional method.Method proposed only focuses on ADHD detection and does not provide a solution on medical basis. Successful ratio obtained between detection of ADHD and normal subjects.

A better dataset means a data set which has more number of relevant input features and more subjects. Feature extraction can be performed using various tools. Freesurfer is a readymade software available for extraction of cortical thickness from structural MRI of brain. C-PAC software is used in extraction of functional connectivity. Unlabelled dataset can be labelled by using clustering algorithms. Once data is labelled it can be used in classification algorithms. MKL can be used to integrate multi modal features.

The approach is by considering three user characteristics, which can be also used as variables in different contexts. These variables are: behavioral conduct, executive functions performance, and emotional state. For inferring the ADHD symptomatic profile of a student and his/her emotional alterations, these features are used as input in a set of classification rules. Based on the testing of the proposed model, training examples are obtained. These examples are used to prepare a classification machine learning algorithm for performing, and improving, the task of profiling a student.

Next approach is classifying adult’s (ADHD) based on spectra of EEG estimations. The examined test incorporates 117 grown-ups (67 ADHD, 50 controls). The estimations are taken for four unique conditions: two resting conditions (eyes open and eyes shut) and two neuropsychological errands (visual nonstop execution test and passionate consistent execution test). We separate the example into four informational collections, one for each condition. Every data set is utilized for preparing of four distinctive support vector machine classifiers, while the output of classifiers is combined using logical expressions got from the Karnaugh graph. The outcomes demonstrate that this approach enhances the separation amongst ADHD and control groups, and also between ADHD subtypes.

According to one research paper about multiclass classification for the differential diagnosis on the ADHD subtypes using recursive feature elimination and hierarchical extreme learning machine: Structural MRI Study, 159 structural MRI images of children were taken. Depending upon this input we get the output of this study. The algorithm proposed in this research was Exeter Learning Environment (ELE). Advantage of this research is Using H-ELM; maximum classification accuracy can be obtained in both binary as well as multiclass classification. The SVM-based recursive feature elimination algorithm makes classification significantly more accurate. The major limitation of this study was the imbalanced number of subjects from each diagnosis group, shared from each site. Pre-processing of structural MRI data is required .Usage of SVM and ELM provide lower accuracy as compared to H-ELM .And the output of this algorithm is Mapping the images which gives disordering of that data .Concluding, by using H-ELM, the accuracy of the testing classification between multiple diagnosis groups can be increased.

From our literature survey we learnt that ADHD detection can be made by analyzing structural and functional MRI of the brain.

3. Proposed System

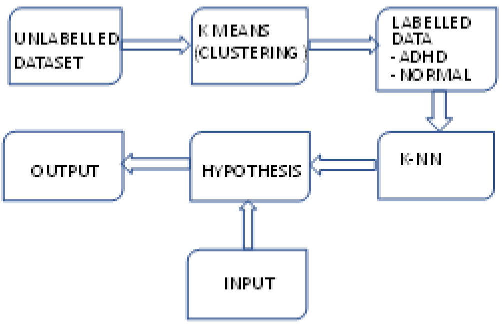

Figure 1: System Diagram

The proposed system consists of taking into parameters from the structural and functional MRI. Structural MRI of the brain gives anatomical features of the brain. Functional MRI measures brain activity by detecting changes associated with the blood flow. The input features necessary for detection are as follows

1) **Cortical Thickness:** It is the thickness of cerebral vertex in the brain. Cortical thickness of patients with ADHD is slightly less as compared to that of a normal human brain. It plays a key role in memory and consciousness. The human cerebral cortex is 2-4mm(0.08-0.16 inches)
2) **Gray Matter Probability:** Gray matter is a major component of the central nervous system, consisting of neuronal cell bodies. Lesser Gray matter in the CNS may also contribute to ADHD.
3) **Regional Homogeneity:** It is voxel based measure of the brain activity which evaluates the similarity of synchronization.
4) **Functional Connectivity:** It is measured as the correlation coefficient of time courses of any two voxels.

Above features are most relevant for detection of ADHD. Dataset available is in unlabelled form that is the data available does not specify whether the given subject has ADHD or whether it is a normal subject

A. Use of K- means clustering algorithm

1) Need for clustering

Initially we had a dataset of 850 people which includes MRIs of ADHD and normal people. As this dataset is unlabeled, we need to cluster this dataset in 2 parts (ADHD subjects and normal subjects). To do this clustering different algorithms are available such as K-means, fuzzy K-means hierarchal clustering, Gaussian clustering. Most accurate and faster to implement was K means and K means is most suitable for our dataset we used K means algorithm for clustering so as to label our data with ADHD subjects and normal subjects.

B. Use of Anamoly Detection

1) Need for anamoly detection:

After running k means algorithm on our dataset to label it, we need to find the inappropriate new input parameters provided by the user.If the values of new input parameters are not matching with our pre defined input parameter's ranges then our anamloy detection algorithm should mark them as anamoly and should not accept that input. In statistics, this is called an "independence assumption" on the values of the features inside training example x.More compactly, the above expression can be written as follows:

C. Use of K-NN (nearest neighbour) algorithm

Need of k- NN: For binary classification K-NN is one of the most preferred algorithm. For the new input given by the user to detect whether that person belongs to ADHD or normal category we need to give the new input to the hypothesis of K-NN .This hypothesis works on the input and gives the desired output. .

K-Nearest Neighbors is a standout amongst the most fundamental yet basic grouping calculations in Machine Learning. It has a place with the administered learning space and finds extreme application in design acknowledgment, information mining and interruption recognition. It is generally expendable, all things considered, situations since it is non-parametric, which means, it doesn't make any basic suspicions about the appropriation of information (rather than different calculations, for example, GMM, which accept a Gaussian conveyance of the given information). We are given some earlier information, which orders arranges into bunches recognized by a quality.

In design acknowledgment, the k-closest neighbors calculation (k-NN) is a non-parametric technique utilized for order and regression. In the two cases, the info comprises of the k nearest preparing cases in the element space. The yield relies upon whether k-NN is utilized for clustering or classification

In k-NN grouping, the yield is a class member. A protest is ordered by a greater part vote of its neighbors, with the question being appointed to the class most basic among its k closest neighbors (k is a positive number, normally little). In the event that k = 1, at that point the question is just relegated to the class of that solitary closest neighbor.

In these, we first find the 10 nearest neighbors of the new given input.

Introduction of algorithm:
In our research, we had unlabelled dataset of around 1,000 subjects. After applying k means algorithm, we got the labeling that is either zero or one(in our case).

Need:
As we have labeled dataset of 1000 disordered and normal individuals, we can apply supervised machine learning algorithm like knn to calssify a new input to existing groups

Algorithm:
K-Nearest Neighbors is a standout amongst the most fundamental yet basic grouping calculations in Machine Learning. It has a place with the administered learning space and finds extreme application in design acknowledgment, information mining and interruption recognition.

It is generally expendable, all things considered, situations since it is non-parametric, which means, it doesn't make any basic suspicions about the appropriation of information (rather than different calculations, for example, GMM, which accept a Gaussian conveyance of the given information). We are given some earlier information, which orders arranges into bunches recognized by a quality.

In design acknowledgment, the k-closest neighbors calculation (k-NN) is a non-parametric technique utilized for order and regression. In the two cases, the info comprises of the k nearest preparing cases in the element space. The yield relies upon whether k-NN is utilized for clustering or classification

In k-NN grouping, the yield is a class member. A protest is ordered by a greater part vote of its neighbors, with the question being appointed to the class most basic among its k closest neighbors (k is a positive number, normally little). In the event that k = 1, at that point the question is just relegated to the class of that solitary closest neighbor.

When we have a labeled dataset, we use it for finding nearest neighbors. After the new input is plot, we find the nearest available point in vicinity. So that the newly added point can be classified to the group of the nearest point. But in some case it may occur the case shown in fig 1. Here, 5 objects from class A are closer and one from class B. Using previous approach, it is classified as class B, which may be not much accurate. Hence we use 'k' nearest neighbors. After k nearest neighbors are found out, we do voting of the groups. The most occurring or the most voted group wins and new input is added to that group. This is how we can classify a new input into existing classes.

D. Normal Equation

Our system has a hypothesis function:

$$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \cdots + \theta_n x_n$$

Where theta is a matrix containing parameters for the hypothesis and x1,x2,x3,x4 are the input features used by the system which are cortical thickness, grey matter probability, regional homogeneity and functional connectivity respectively.

We'd like to minimize the least-squares cost:

$$J(\theta_{0\ldots n}) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

Where $x^{(i)}$ is the i-th subject (from a set of 135 samples) d $y^{(i)}$ is the i-th expected result.To proceed, we'll represent the problem in matrix notation; this is natural, since we essentially have a system of linear equations here. The regression coefficients θ we're looking for are the vector:

$$\begin{pmatrix}\theta_0\\ \theta_1\\ \ldots\\ \theta_n\end{pmatrix} \in \mathbb{R}^{n+1}$$

Each of the 135 input samples is similarly a column vector with 4 rows, x_0 being 1 for convenience. So we can now rewrite the hypothesis function as:

$$h_\theta(x) = \theta^T x$$

When this is summed over all samples, we can dip further into matrix notation. To find the values for the theta we need to use the equation given below.

$$\theta = (X^T X)^{-1} X^T y$$

which is the normal equation and X is the matrix which contains dataset of the system.

This equation provides us with the parameters which are used in our hypothesis h(x).This hypothesis is then used to calculate our output (ADHD subject or normal subject). Normal equation in our case where we have binary classification has a much less accuracy. We obtained a accuracy.

Abbreviations

ADHD: Attention Deficit Hyperactivity Disorder
KNN: K – nearest Neighbor
ELM: extreme machine learning
F-MRI: functional magnetic resonance imaging
S-MRI: structural magnetic resonance imaging
MKL :Multi Kernel Learning
BC : behavioral conduct
EEP: executive functions performance
ES : emotional state

4. Conclusion

The proposed system presented classification model of ADHD and normal peoples .In the given system 4 types of input parameters which are extracted from brain images .Feature selection is based on the advice of the radiologists. Then on the dataset of approx 1000 peoples which included MRIs of ADHD patients and normal peoples, we applied the k –means classification algorithm which classifies data in two clusters , one of ADHD patients and other of normal peoples. Then for a new input to detect whether an individual has ADHD or not, K- NN algorithm is used to detect the new input belongs to which category.

References

[1] Classification of ADHD children through multimodal magnetic resonance imaging Dai Dai, Jieqiong Wang, Jing Hua, and Huiguang
[2] Classification of ADHD patients on the basis of independent ERP components using a machine learning: Gian Candrian†,Juri D Kropotov†
[3] Machine learning approach for classification of ADHD in adults: Aleksander Tenev, Jordan Pop
[4] Biederman J., Mick E., Faraone S. V . (2000). Age-dependent decline of symptoms of attention deficit hyperactivity disorder: impact of remission definition and symptom type. Am. J. Psychiatry 157, 816–818. [PubMed]
[5] Bush G., Frazier J. A., Rauch S. L., Seidman L. J., Whalen P. J., Jenike M. A., Rosen B. R., Biederman J. (1999). Anterior cingulate cortex dysfunction in attention-deficit/hyperactivity disorder revealed by fMRI and the Counting Stroop. Biol. Psychiatry 45, 1542–1552. [PubMed]
[6] Cao Q., Zang Y ., Sun L., Sui M., Long X., Zou Q., Wang Y . (2006). Abnormal neural activity in children with attention deficit hyperactivity disorder: a resting-state functional magnetic resonance imaging study. Neuroreport 17, 1033–1036. 10.1097/01.wnr.0000224769.92454.5d [PubMed] [Cross Ref]
[7] Chawla N. V ., Japkowicz N., Kotcz A. (2004). Editorial: special issue on learning from imbalanced data set. ACM SIGKDD Explorations Newsletter 6, 1–6.
[8] Craddock R. C., James G. A., Holtzheimer P. E. III., Hu X., Mayberg H. S. (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. Hum. Brain Mapp. 33, 1914–1928.
[9] 1002/hbm.21333 [PMC free article] [PubMed] [Cross Ref] 10. Dale A. M., Fischl B., Sereno M. I. (1999). Cortical surface-based analysis I: segmentation and surface reconstruction. Neuroimage 9, 179–194. 10.1006/nimg.1998.0395 [PubMed] [Cross Ref]
[10] Durston S., Tottenham N. T., Thomas K. M., Davidson M. C., Eigsti I. M., Yang Y . (2003). Differential patterns of striatal activation in young children with and without ADHD. Biol. Psychiatry 53, 871–878. 10.1016/S0006-3223(02)01904-2 [PubMed] [Cross Ref]
[11] Fan Y ., Shen D., Gur R. C., Gur R. E., Davatzikos C. (2007). COMPARE: classification of morphological patterns using adaptive regional elements. IEEE Trans. Med. Imaging 26, 93–105. 10.1109/TMI.2006.886812 [PubMed] [Cross Ref]
[12] Fox M. D., Snyder A. Z., Vincent J. L., Corbetta M., Van Essen D. C., Raichle M. E. (2005). The human brain is intrinsically organized into dynamic anticorrelated functional networks. Proc. Natl. Acad. Sci. U.S.A. 102, 9673–9678. 10.1073/pnas.0504136102 [PMC free article] [PubMed] [Cross Ref]
[13] Guyon I., Weston J., Barnhill S., Vapnik V . (2002). Gene selection for cancer classification using support vector machines. Mach. Learn. 46, 389–422.
[14] Hinrichs C., Singh V, Mukherjee L., Xu G., Chung M. K., Johnson S. C. (2009). Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. Neuroimage 48, 138–149. 10.1016/j.neuroimage.2009.05.056 [PMC free article] [PubMed] [Cross Ref]
[15] Hinrichs C., Singh V ., Xu G., Johnson S. C. (2011). Predictive markers for AD in a multimodality framework: an analysis of MCI progression in the ADNI population. Neuroimage 55, 574–589. 10.1016/j.neuroimage.2010.10.081 [PMC free article] [PubMed] [Cross Ref]
[16] Kates W. R., Frederikse M., Mostofsky S. H., Folley B. S., Cooper K., Mazur-Hopkins P., Kofman O., Singer H. S., Denckla M. B. (2002). MRI parcellation of the frontal lobe in boys with attention deficit hyperactivity disorder or Tourette syndrome. Psychiatry Res. 116, 63–81. [PubMed]

Author Profile

Rohit Kale receiving the Bachelor of technology degree in Information Technology Engineering from Vishwakarma Institute of Technology.