ISSN 2278-3091
Volume 9 No.2, March -April 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse231922020.pdf
https://doi.org/10.30534/ijatcse/2020/231922020

Autism Detection and Sub-grouping

Neha Reddy Bonthu[1], Swarna Kuchibhotla [2] , Prasuna Kotturu [3] ,Niranjan S.R.Mellacheruvu[4], B. Manjula Josephine[5]

[1,2,3,5] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, nehareddybonthu725@gmail.com, drkswarna@kluniversity.in, prasukotturu@kluniversity.in , manjulajosephine@kluniversity.in, India

[4]Department of Electronics and Communication Engineering, Vikas college of Engineering and Technology Guntur ,niranjanmsr@gmail.com, India

ABSTRACT

This paper is regarding Autism spectrum disorder(ASD) and its prevalence based on age. Three datasets from different sources are collected and an additional dataset was obtained from NSCH which consists of several attributed have undergone preprocessing and principal component analysis was applied in-order to achieve highest accuracy i.e., an approximate of 99% for all the datasets used . In addition to this, the study also includes the sub-grouping i.e., the disorders under the spectrum of ASD and the correlation between age and the disorder.

Key words : Autism Spectrum Disorder (ASD), DSM-5 (The Diagnostic and Statistical Manual of Mental Disorders), NSCH (National survey of child health).

1. INTRODUCTION

Autism is defined as a developmental disorder with the difficulty of cognition, physical impairment, etc. Several types of research are conducted on autism in different areas. However, we are concerned regarding the types of disorders in the spectrum of ASD. As early studies reveal that even autism was treated as schizophrenia until 1943. By referring to the existing database, we can provide an alternative diagnosis with the aid of machine learning techniques.

This study a continuation of our earlier study of subgrouping the disorders using DSM-5. We considered several factors like gender, history of disorders, symptoms, etc. We would like to add the new attribute i.e, the age. The datasets we are dealing with are of three different age groups and we have so far conducted four machine learning classification techniques.

The proposed study aims at building a mobile screening tool that may be used by the citizens at the comfort of their own home and can assess regarding the type of disorder, Whether the disorder is in the spectrum of autism and treatments that are currently preferred by the medical experts for the disorders.

2. LITERATURE SURVEY

The three datasets of different age groups are collected from the survey conducted by F.Abdeliaber in [1]. Several screening processes were conducted in [2,4,5,12]. However, we used different classifiers for the existing datasets. In [1] and [4] same datasets were used generated from a mobile screening app where the former used ten-fold cross-validation using if-then rules and obtained an accuracy around 90% and the latter performed KNN (K-Nearest Neighbor) and LDA(Linear Discriminant analysis) and best of them turned out to be LDA with an accuracy of 90%.

The new screening process was introduced in [10] using MRI Scans and [13] has introduced similar technique using Neural networks classifier. The idea of considering age as an attribute as it was mentioned in [8] that early symptoms prediction in children is difficult.

The role of DSM in the decision making of clinicians and medical experts is cited in [3,7]. In [7], DSM-IV and DSM-V are compared and the stringent nature, advancements in DSM-V that meet the criteria of ASD diagnosis are incorporated. In [3], the ideas of sub grouping of ASD and the significance of it during diagnosis are acknowledged. The basis of sub grouping in [3] is from clinical samples of [7,9] and some existing datasets. The specifiers in DSM-5 are mentioned which aid the phenotypic characterization.

3. METHODOLOGY

In this study, data is collected from three different sources in raw formats. The datasets are of different age groups i.e., adult, adolescent, and toddler. Besides this we collected a dataset from NSCH. All these datasets are preprocessed for further applying the classification algorithms of machine learning . Later, machine learning algorithms Naive Bayes, Neural Networks, Support vector machine and random forest are applied. For achieving better accuracy, principal component analysis is applied to the existing datasets.

.

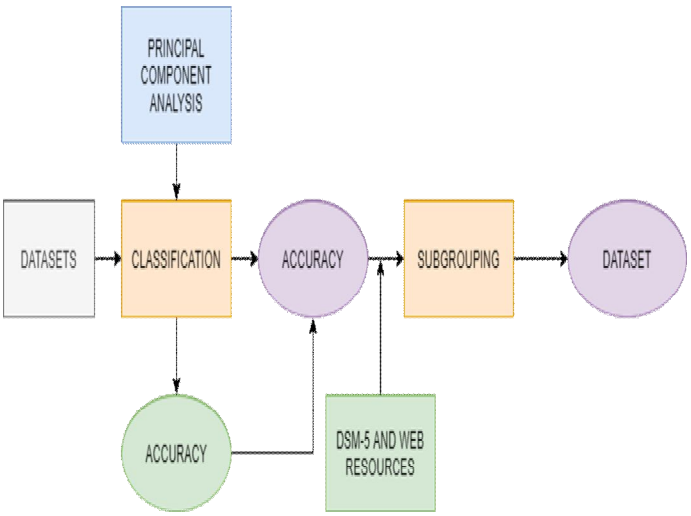

Figure 1: Illustration of classifying the dataset using effective algorithms and sub grouping

These results aided in the continuation of our study which is sub-grouping of the disorder. Age is the new specifier added to the study.

4. FEATURE EXTRACTION

Feature extraction in machine learning is used for the dimensionality reduction of the data which is suitable for building the architecture of the system. The transformed attributes are a linear combination of original attributes. The attributes in the datasets are questionnaires that are reduced into these features.

4.1. Social Interaction

Social interaction of children is to be known as it is the first indication of autism. It is to analyze how comfortable are children communicating with an individual or with a block of society.

4.2. Focus

Children may deviate their attention for a bit of time which is normal. The focus attribute is to determine the amount of information obtained from both vocal and non-vocal formats.

4.3. Gender

Gender attribute is mentioned in the dataset as researchers have consistently found that more boys/ men than women/girls are prone to have autism.

4.4. Multi-Tasking

Performing multiple tasks at a single -go requires certain cognitive ability and both physical and mental coordination. It is to determine whether comfort levels of children while multitasking.

4.5. Age

Age attribute is used just to classify the data in which age group is most likely to have autism as previous researchers mostly concluded based on this feature.

4.6. Medical History

Medical history is essential to analyze both physical and mental disorders. As, diseases such as Jaundice , seizures, etc are the most prevalent ones in autistic children.

5. EXPERIMENTAL RESULTS

Four different classifiers runs are conducted on three datasets collected from UCI and Kaggle. These datasets as mentioned earlier are of different age groups.

Further, we considered NSCH dataset with 22000 samples and conducted four classifiers runs on it

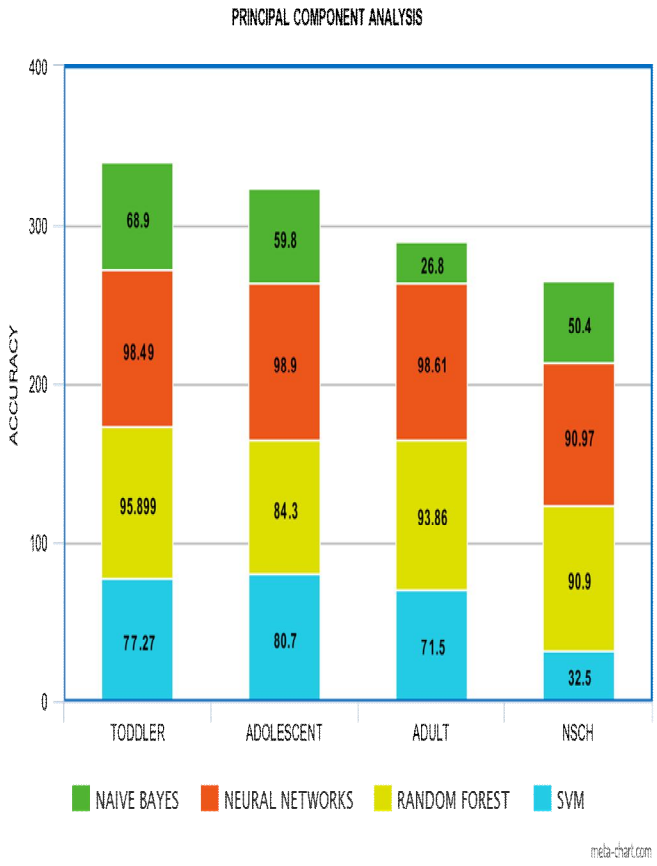

Figure 2: Illustration of classifying the dataset using effective algorithms and sub grouping.

For attaining best accuracy we conducted principal component analysis on the four existing datasets. A huge variation of inaccuracy has been observed after performing PCA.

Table 1: Results of classifying the dataset using effective algorithms

.

DATASET	NAÏVE BAYES	NEURAL NETWORKS	RANDOM FOREST	SVM
TODDLER	68.9	99.99	97.4	73.8
ADOLESCENT	59.8	99.99	25	76.9
ADULT	26.87	98.627	92.13	74.43
PCA TODDLER	68.9	98.49	95.899	77.27
PCA ADOLESCENT	59.8	98.9	84.3	80.7
PCA ADULT	26.8	98.61	93.86	71.5
NSCH	50.4	90.97	90.9	72.5

In Table 1, the accuracy values of datasets using four different classifiers are represented in which neural networks outperformed the remaining classifiers.

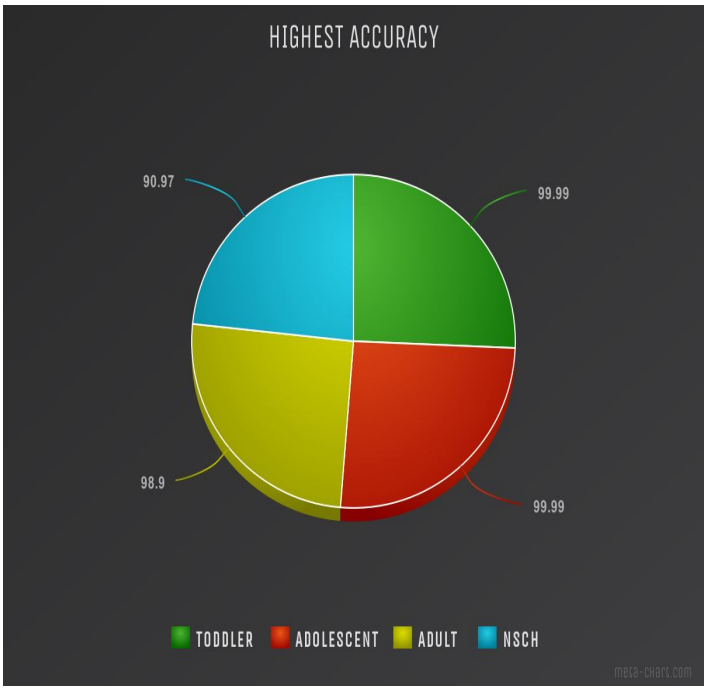

Figure 3: Highest accuracy obtained in three datasets .

From the subgrouping dataset generated, 60% of the disorders didn't make it into the spectrum. In most of the disorders, toddlers are more prone to disorders compared to adults and adolescents.

6. CONCLUSION AND FUTURE SCOPE

Cleaning the data set (which recorded the characteristics associated with ASD) was difficult in that we had mainly categorical variables and only two numerical variables, but eventually, we were able to build these models and found that all machine learning algorithms are performed by the algorithm which performs best in all aspects and neural networks.

Although the information association made the prediction very easy, we believe that this research will serve as useful aid for doctors to identify new autistic cases.

In our consideration, we need to have larger datasets to build a precise and robust model. Here, after cleaning the data, the number of instances was not sufficient to say this model is optimal. With this current data collection, nothing can be changed by looking at the quality of our learning systems, as models are already at their peak. After discussing this issue with a researcher working directly on child autism, we realized that collecting a lot of well-documented ASD-related data is extremely difficult. Recently, this ASD dataset has been made public (available since December 2017), and so little work has been done. With this in mind, our research has led to well-developed models that can be accurately detect ASD in people with unique behavioral and health knowledge attributes. Such models may serve as benchmarks for any machine learning researcher / practitioner interested in further exploring this dataset or other Autism screening disorder related data sets.

REFERENCES

1. F.Abdeljaber , **Detecting Autistic Traits using Computational Intelligence & Machine Learning Techniques** , 2018 .

2. Matthew J.Maenner, Kim Van, Kim Van *.et.al* , **Development of a Machine Learning algorithm for the Surveillance of Autism Spectrum Disorder** 2016.

3. Meng-Chuan Lai, Bhismadev Chakrabati *.et.al* , **Subgrouping the Autism "Spectrum" : Reflections on DSM-5** 2013

4. Osman Altay, Mustafa Ulas , **Prediction of the Autism Spectrum Disorder Diagnosis with Linear Discriminant analysis and K-Nearest Neighbor in Children** 2018.

5. Matthew P.Black, Catherine Lord *.et.al* , **Use of machine learning to improve autism screening and diagnostic instruments : effectiveness, efficiency and multi instrument fusion** 2016.

6. Rhiannon Luyster , Jae-Yoon Lung *.et.al* ,**Use of artificial intelligence to shorten the behavioral diagnosis of autism** 2012.

7. Vicki Gibbs, Felicity Chandler *.et.al* , **An exploratory study comparing diagnostic outcomes for autism spectrum disorders under DSM-IV-TR with the proposed DSM-5 revision** 2012.

8. Sushama Rani Dutta, Soumyajit Gi, *.et.al* , 2005**, "A Machine Learning-based Method for Autism Diagnosis Assistance ".**

9 .Swarna Kuchibhotla, HD Vankayalapati and KR Anne, **"An optimal two stage feature selection for speech emotion recognition using acoustic features,"** International Journal of Speech Technology, pages 1-11, 2016.

10. Cooper Mellema, Alex Treacher *.et.al* **Multiple Deep-Learning architectures achieve superior performance diagnosing autism spectrum disorder using features previously extracted from structural and functional MRI ,** 2019.

11. Joshi , Sujatha *.et.al***, "Applications of chatbots in marketing ,"** International Journal of Advanced Trends in Computer Science and Engineering, 2019.

12. Mohamed A , Saleh.N *.et.al***, "A Deep Learning Approach in Robot-Assisted Behavioral Therapy for Autistic Chlidren ,"** International Journal of Advanced Trends in Computer Science and Engineering, 2019.

13. Rajesh R , Dr S.N.Kale**, "Neural Network Classifiers for Cardiac Arrhythmia From Scanned ECG Documents",** International Journal of Advanced Trends in Computer Science and Engineering, 2019.