International Journal of Scientific Research & Engineering Trends
Volume 6, Issue2, Mar-Apr-2020, ISSN (Online): 2395-566X

Enhancement using Adaboost Algorithm in Machine Learning to Predict ASD in Children

Mary Stella. J, Dr. Shashi Kumar. D.R
Department of CSE
Cambridge Institute of Technology
Bangalore, India
marystella12890@gmail.com,shashikumar.cse@cambridge.edu.in

***Abstract* – Autism is a neurobehavioral disorder of development which is seen with social behavioral and interactive behavioral impairments in children. The perspective of detecting and predicting the ASD in the children is to diagnose and treat the children in early age which could be a life saving aspect in leading a normal life. Though different techniques have been studied it did not provide any beneficial conclusion in predicting traits on autism in children of small age groups. Different Data mining techniques and prediction model of ML techniques were explored to performance evaluation and accuracy validation. In our study, the required classification method of ASD diagnosis has been used with children of very early age. Our proposed paper compares many algorithms of machine learning along with AdaBoost algorithm. AdaBoost combines the weak classifiers to a strong classifier and achieves more accuracy finally.**

***Keywords*– MLTechniques, ASD diagnosis etc.**

I. INTRODUCTION

In Recent days, Autism Spectrum Disorder is outgrowing faster, even diagnosing of autism could be traced in any age and its symptoms appear generally in the first 2 years of age in life that can be developed through time which drives to worst situations. Autism patients generally face many challenges like concentration difficulty, learning disabilities, anxiety, depression like mental health issues, sensory problems, motor difficulties etc. [1]. The upsetting disorders in the childhood in the terms of prevalence, outcome, morbidity, the effects on family, and the cost to the society where the recent estimate report in United States that, 1 in 68 children has been diagnosed autism [3].

Children who exhibit ASD lacks in communication and cognition skills, but the symptoms may vary from child to child. Autism, from research as a study, it develops as a combination of genetic, non-genetic, environmental influences. As a result, it appears to develop autism as an increased risk in a child. Child should be diagnosed within three years, earlier the diagnoses, the intervention can be sooner. ASD symptoms promptly need to be diagnosed with children age up to 36 months. Diagnosis usually made when children react or doesn't react with parents, friends or in interactivity [2].

Intervention and early diagnosis faced difficulties with number of ASD patients increasing and with scarce clinical resources. To overcome this difficulty the researchers, focus on observing possibly the early signs from clinical observations of children before age of 2 years [10]. Early interventions of diagnosis in ASD children can benefit in high success rate [4].

II. LITERATURE REVIEW

The autoimmune influenza and the ASD cases has been increasing in numbers in the world which reveals an immediate and urgent necessity to develop the easily applicable and the effective screening methods. Here, the performance comparisons are made with many different classification methods like, Naïve Bayes, Radial Basis Function Network, IBk (k-nearest neighbors) and the Random Forest, with UCI 2017 Autism Screening Data of the Children dataset. Analysing the result of experiment, RF method (Random Forest) has been the more successful methods than the Naive Bayes, RBFN methods, and IBk.[11]

Machine learning and Deep learning algorithms are widely and rapidly utilized in the medical sciences where the biomedical data has been transformed into valuable knowledge. It is also widely involved in bioinformatics like building predictive models in diagnosis and detection of diseases, segmenting medical images, protein fold prediction, gene finding and others. This paper provides the comparison study on predicting performance of selected learning classifiers. The classifiers predict the models that provides good precision, high accuracy, recall percentages and Receiver Operating Characteristics (ROC) performance for the given problem. Thus

prediction concludes with that, the Deep Learning method of classifier has achieved highest performance, in the terms of precision, accuracy, recall percentages and the ROC, whereas, the Random Tree method of classifier achieved the least performance in all the metrics. Results had been achieved with the performance of classification with Naïve Bayes, k-nn, and the Decision Tree (DT) classifiers and used as the reference for the future work in ASD screening.[12]

Ensemble method integrates many types of classifiers in building a powerful model like the system of pattern recognition and it enhances robustness and recognition effect by the cooperation and complementation between the individual Support Vector Machine (SVM) classifiers. Along this, on incorporating the traditional AdaBoost, feature selection and sample selection, and added to the improved algorithm in AdaBoost operational structure. In AdaBoost application, the selection algorithm used in selecting training samples is the neighbor-mean selection algorithm. In this algorithm the samples are selected, those which are the nearest to mean, as the selected samples and by calculating the mean of the sample from the selected sample sets. Then the feature selection relies on the relative entropy that selects the features ensuring the optimal features and samples had been selected and made to train individual SVM classifier for each cycle. Then the trained individual classifiers, by optimizing the subset would have the higher accuracy of classification, and the complexity is greatly reduced with the individual classifier by the removal of redundant features and noise samples in the training set.[13]

Cluster analysis is applied in this application to the sample of maximum 2,116 children of Autism Spectrum Disorder (ASD) to identify the patterns of the challenging behaviors of children observed in houses and clinical settings. Furthermore, the study has been observed that, when training the cluster models by separating the male and the female samples. Here, first machine learning has been presented based in the analysis of 2,116 patient samples. Challenging behavioral uses K- means clustering, to identify the behavior profiles that indicate the presence of dominant single behavioral challenge in most of the clusters.

Additionally, we also identified scarce potential differences in the behavioral challenge profiles across the male and the female populations. This study for future work, lays the foundation to model, how the treatment intensity, stimulus response, learning objective mastery were correlated in defining the cluster of challenging behavior. This might provide an important basis for the individualized treatment, also with the goal setting of the higher efficacy and at the lower cost.[18] Machine learning techniques have been applied and the performance on the Autism Spectrum Disorder related dataset has validated. In this paper the comparison of the performances is shown on different algorithms in ASD screening. In this work, the person is checked to find the traits of ASD(autism spectrum disorder) if any and so it's been implemented many machine learning algorithms and techniques like Naïve Bayes (NB), LDA (Analysis of Linear Discriminant), (KNN)the K-Nearest Neighbor, (CART) Classification And Regression Trees, (SVM) the Support Vector Machine and (LR) Linear Regression for classification of the ASD data. For the implementation of the algorithms, it's been selected 70 percentage of data as the trained data and the remaining 30 percentage of data as the test data. The data set of adults consists of 702 instances and 19 attributes. Data usually has selected very randomly. Firstly, the data set is categorized with their age, class/asd, gender and the result, in which it found, that the persons between 17-22 of age are high in number. Same way, dataset contains males higher in number than females and 189 instances has been suffered from ASD.[5]

Incorporating ensemble learning techniques involves two main key issues namely diversity of the base classifiers and the integration of multiple classifiers. In this application, a decision group is designed to the diversity of base classifier in order to increase the performance. The Genetic algorithm is incorporated to which the weights are assigned for each base classifier. AdaBoost algorithm and genetic algorithm are considered as the proposed ensemble classification algorithm for diagnosing cancer by classifying the gene data. Two main challenges in ensemble learning is how the base classifiers can be improved and how the integration can improve the performances. The base classifiers used are the KNN, decision trees and Naïve Bayes algorithms. Our proposed algorithm is AdaBoost- GA and it improves the performance of our selected base classifiers very effectively than the strong classifiers.[7]

III. BACKGROUND INFORMATION AND ALGORITHMS

1. Machine Learning:

Machine Learning is one of the applications of AI which facilitates the systems by providing ability to learn automatically and improve fromthe training datasets as experience, to predict the output, without using explicit instructions, instead relying on patterns and inferences. The aim is to make the computers learn something automatically without the human intervention and generate the output. ML is mainly categorized as the supervised learning and the unsupervised learning.

2. Supervised Learning:

Supervised learning algorithms are used by applying the learned patterns and parameters from the past to the input instance using labelled data. Starting with the analysis

from a training dataset, algorithm usually generates the inferred function for predicting the output values. The computer now able to provide the targets for any input instance after ample training. Additionally, the algorithms can compare the output with the intended output and predicts errors in order to adjust the model to produce high accuracy level.

3. Support Vector Machine (SVM):
SVM is an e-learning technology used for both classification as well as regression challenges. It has reached heights in biology for its well high accuracy. SVM doesn't require mathematical theory in depth to understand, its easier than the Neural networks [9]. In SVM algorithm, we point each data in n-dimensional space with its feature value of the particular coordinate. Later, we do classification by finding hyper plane which differentiate the two classes well. This SVM is the frontier that segregates 2 classes with the hyperplane very well. A technique named kernel trick transforming data and finds optimal boundary between possible outputs based on the transformations.

4. Random Forest:
Random Forest consists of huge number of decision trees, individual trees, which would operate as the ensemble. Every tree in Random Forest gives out the class prediction, those classes with more votes declared as model's prediction. Random Forest uses bagging, also feature randomness in the building of each tree to create the uncorrelated forest of the trees. This algorithm is widely used for better accuracy and low prediction error. Random Forest differs from decision tree, in which, the decision tree built on entire dataset with all features, but RF randomly selects the rows and features to build the tree and averages the outcomes.

5. AdaBoost
Ensemble Learning is modern ML Technique in solving very complex classifications by integration of many simple classifiers. Ensemble Algorithm designed basically on 2 approaches, Bagging and Boosting. AdaBoost is one of boosting algorithms which determines the weights of individual sample subset, iteratively and determines, relying on the accuracy of previous values of the classification results [7]. Adaptive Boosting (AdaBoost), proposed by Robert Schapire and Yoav Freund, can be used for classification and in conjunction with other learning algorithms to improve performance. The algorithm retrains iteratively, by choosing training dataset based on accuracy of the previous training. The weight of each trained classifier of any iteration just depends on the accuracy achieved. AdaBoost is very sensitive to the noisy data and the outliners.

IV. DATA ACQUISITIONS

The Californian University, Irvine (UCI) has a numerous dataset for machine learning (ML) algorithms. Mostly, the data gathered in this study was obtained from Tabtah, F. [6]. In this dataset, there are questionnaire related to the ASD and the score attributes consists of sum of questions in these datasets. To test with the algorithms, 30 percent of dataset has been selected for test data, 70 percent selected as the training data. Dataset consists of twenty-one attributes with the mix of binary and categorical data including age, ethnicity, gender etc. With needed information the autism traits are predicted with our proposed work. In our proposed work, different algorithms are implemented, calculating the accuracy of prediction and comparing it with the algorithms to elucidate the prediction of ASD with high accuracy. The purpose of implementing machine learning is to minimize the diagnose time and increase the accuracy rate of diagnosis to distinguish ASD and Non-ASD children [14].

V. PROPOSED WORK

Machine Learning comprises of many classification algorithms. Performance analysis and comparison of algorithms in data classification is a complicated process as it depends on several of evaluation dimensions and featuring and selecting the dimensions also to be considered. The proposed methods, can select features and samples very effectively in the different statics of the data sets and by getting rid of noise samples as well and the redundant features for selecting reliable samples for the training of data to effectively improve, the performance of the classifiers with trained data in each of the iteration process. In our work three algorithms are used to achieve the best performance.

Algorithms used are SVM, Random Forest, and AdaBoost algorithms. AdaBoost algorithm an effective tool in order to improve the ability of prediction learning system and also one of the representative methods with combined learning. Two observed issues with AdaBoost are the way of adjusting the training set to implement weak classifiers in training set and the way to combine weak classifiers to make strong classifiers. Selecting the training sets using random subsets and selecting the appropriate features, assigning the weights to each classifier measures are handled to boost the accuracy in AdaBoost methods. The accuracy level and the performance measures based on the attributes selected is analyzed to boost the accuracy levels of diagnosing autism in children.

VI. CONCLUSION

In our study and work, the performance and the accuracy has been studied and the comparative performance have been made with the algorithms. The investigation

promotes the accuracy among the classifiers, that worked on the ASD Screening. In this paper its been taken the children data of very early age and diagnosed ASD using the classifiers and predict which one produces high accuracy and performance evaluation in diagnosing autism. Algorithms and Classifiers are compared for accuracy and performance and the highest accuracy is selected as the best classifier for prediction. In Medical Research, high and accurate prediction is very essential in diagnosing patients who desperately in need of medical help, therefore the best intervention strategies can be provided in children those who been diagnosed with autism. In future, further modifications can be made to algorithm and on selecting attributes to attain even more better results.

REFERENCES

[1]. Kazi Omar, Nabila Khan, Prodipta Mondal, Md. Rezaul, Md Nazrul Islam, "A Machine Learning Approach to Predict Autism Spectrum Disorder," International Conference on ECCE, February 2019.

[2]. Osman Altay, Ulas, "Prediction of the Autism Spectrum Disorder Diagnosis with Linear Discriminant Analysis Classifier and K-Nearest Neighbor in Children," Firat University, Turkey 2018.

[3]. Khalifeh S, Yassin W, and Kourtian S, "Autism in Review," Lebanese Medical Journal, 2016.

[4]. DiCicco-Bloom et al. E, "The Developmental Neurobiology of Autism Spectrum Disorder," A Neuroscience journal, vol.26, Jun. 2006.

[5]. Bhawana Tyagi, Neha Bajpai, Rahul Mishra, "Machine Learning Techniques to Predict Autism Spectrum Disorder," C-DAC Noida, India.

[6]. Thabtah, F. "Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment," 2017.

[7]. Huijuan Lu, Ke Yan, "A Hybrid Ensemble Algorithm combining AdaBoost Algorithm with Gene Expression Data," 9th - International Conference in IT on Medicine & Education, 2018.

[8]. TU Chengsheng, LIU Huacheng, XU Bing," The Application of the AdaBoost Algorithm in the Text Classification," IEEE AIM, Electronics and ACC (IMCEC 2018)

[9]. Hardi Talabani, Engin AVCI," Performance Comparison of SVM Kernel Types on Child Autism Disease Database," Faculty of Technology, Firat University, Software Engineering Department, Elazig, Turkey, 2018 IEEE.

[10]. Alanoud Bin Dris, Abdulmalik Alsalman, Areej Al-Wabil, Mohammed Aldosari, Centre for Cyber Security Technology, King Abdulaziz City for Science and Technology (KACST) Riyadh, Saudi Arabia, 2019 IEEE.

[11]. Fatiha Nur Büyükoflaz, Ali Öztürk, "Early Autism Diagnosis of Children with Machine Learning Algorithms," Bilgisayar Mühendisli˘gi Bölümü, KTO Karatay Üniversitesi Konya, Türkiye, 2018 IEEE.

[12]. Alrence Santiago Halibas, Leslyn Bonachita Reazol, Erbeth Gerald Tongco Delvo, Jannette Cui Tibudan, "Performance Analysis of Machine Learning Classifiers for ASD Screening," 2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies.

[13]. Yiming Tian, Xitai Wang, "SVM ensemble method based on improved iteration process of Adaboost algorithm," Hebei University of Technology, Tianjin, 2017 IEEE.

[14]. Sushama Rani Dutta, Sujoy Datta, Monideepa Roy, "Using Cogency and Machine Learning for autism detection from a preliminary symptom," KIIT University, Odisha, India, 2019 IEEE.

[15]. Kajaree Das, Rabi Narayan Behera, "A Survey on Machine Learning: Concept, Algorithms and Applications," Institute of Engineering and Management, Kolkata, India.

[16]. Aaron J. Masino, Daniel Forsyth, Heather Nuske, John Herrington, Jeffrey Pennington, Yelena Kushleyeva, Christopher P. Bonafide, "m-Health and Autism: Recognizing Stress and Anxiety with Machine Learning and Wearables Data," 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS).

[17]. Halim Abbas, Eric Glover, Dennis P Wall, "Machine learning for early detection of autism (and other conditions) using a parental questionnaire and home video screening," 2017 IEEE International Conference on Big Data (BIGDATA).

[18]. Elizabeth Stevens, Abigail Atchison, Laura Stevens, Esther Hong, Doreen Granpeesheh, Dennis Dixon, Erik Linstead, "A Cluster Analysis of Challenging Behaviors in Autism Spectrum Disorder" 2017 16th IEEE International Conference on Machine Learning and Applications.