International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 07 | July 2021 www.irjet.net p-ISSN: 2395-0072

Comparison Analysis of Autism Spectrum Disorder Using Machine Learning Approaches

M C Ranjitha*1, Dr. Jalaja G*2

*1Student, Department of CSE, BNM Institute of technology, Bangalore, Karnataka, India
*2Associate Professor, Department of CSE, BNM Institute of technology, Bangalore, Karnataka, India

---

Abstract - *Autism Spectrum Disorder (ASD) is gaining transaction faster than ever today. Screening approach to measure autistic features are very complex and expensive. Through use of artificial intelligence and machine learning (ML), autism could be predicted at an appropriate initial stage. While multiple studies using various techniques have been carried out on these and did not provide any definite conclusion on the prediction of autism traits in terms of different age groups. The suggested validation was performed using the autism spectrum quotient-10 (AQ-10) dataset and the 1000 dataset from Kaggle, both with and without autistic features. For both types of datasets, the evaluation results will also reveal that the suggested prediction approach provided greater benefits in order to improve accuracy, recall, and thus the f1-score. As a result, the purpose of this paper is to establish an appropriate prediction model. The goal of this project is to present an autism machine learning-based scheme is developed techniques and to construct an implementation for portable devices that can accurately identify autistic features in people of a certain age. The purpose of this project is to create an autism screening app that can predict ASD symptoms in children aged 3 years and under. The proposed strategy focuses on a machine learning procedure for Autism spectrum disorder (ASD) classification and prediction, thus overcoming the existing problem. The model's performance and accuracy will be enhanced by incorporating the Random Forest (RF), Support Vector Machine (SVM), and AdaBooster algorithms.*

***Key Words*: Machine learning, AQ-10 dataset, random forest, Support Vector Machine (SVM), and AdaBooster, ASD.**

1.INTRODUCTION

Autism spectrum disorder (ASD) is a neurological condition that impairs a person's capacity to speak, engage, and learn. Autism could perhaps emerge during any era, nevertheless, the clinical manifestations usually begin during the first three years of life and gradually worsen. Autism is currently exploding all over the world, and it is escalating at a rapid stage. According to the WHO, ASD affects around one out of every 160 children. Some individuals with this condition are self - reliant, while others will need lifelong care and assistance. Autism diagnosis consumes a lot of time and costs a lot of money. Early detection of autism can be quite beneficial in terms of providing effective medication to patients. It has the ability to inhibit the patient's health from deteriorating further, along with reducing the long-term expenses of delayed diagnosis. As a result, a simple, reliable, and inexpensive screening test tool is definitely needed to predict autistic features in individuals and determine whether or the person require a rigorous autism examination.

2. LITERATURE REVIEW

This subsection briefly summarizes the research on ASD prediction techniques. The accuracy of machine learning in predicting various types of diseases based on syndrome is very remarkable. The review enables to critically summarize the current knowledge under investigation. Many researches have been done on ASD prediction.

Suman Raj [1] et.al., presents the Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques for predicting and analyzing ASD problem in children, teenagers and adults that used a network and a convolutional neural network. The proposed strategies are tested using three publicly available non-clinically ASD datasets. The first dataset contains 292 incidences and 21 features that are relevant to ASD screening in youngsters. The second dataset for ASD screening Adult individuals has 704 instances and 21 attributes. The third dataset contains 104 cases and 21 attributes associated with teenage ASD screening. In this study, both the SVM and CNN dependent models show a prediction an 98.30% accuracy score besides the ASD child dataset while dealing with missing values.

Noora Saleem Jumaa [2] et.al., presents The Autism Spectrum Disorder Diagnosis Based on Machine Learning Techniques where the main aim is Another mission of project is to examine any tactic which do diagnose autism in children aged 4 to 17. The client, which is responsible for collecting patient data and transferring it to the server, is the initial portion of the ASD disease prediction technique in this research. The system has a 0.044% average mistake rate and can detect ASD with a 98% accuracy rate. A mechanism has been postulated for predicting the onset of autism. As a tool for testing the suggested scheme, it was tested using two types of machine-learning sets of collections of information collected from the (UCI) warehouse. The output of many classifying algorithms that may be employed for this type of

classification problem can be used here. Three of the most common algorithms are Nave Bayes, Logistic Regression, and K-Nearest Neighbors. Using those several algorithms, one can figure out which one has the best accuracy and lowest error rate, and which one will be utilized as the classifier in the suggested technique.

Muhammad Faiz Misman [3] et.al., presents the Classification of Adults with Autism Spectrum Disorder using Deep Neural Network they used have, researchers used the Deep Neural Network (DNN) architecture, and it has become a popular tool in past decades and has been found to enhance prediction performance. Considering multiple variables of adult ASD screening findings, this study will explore the achievement of the DNN model in the wide range Of symptoms in terms of solution quality. The findings are comparable to Support Vector Machine (SVM), a prior machine learning system produced either by developer. On the first dataset, the DNN model correctly diagnosed ASD diagnoses with 99.40 %, while on the validation set, it did so with 96.08 %. Conversely, considering both first and subsequent data, the SVM model average state of 95.24 % and 95.08 %, respectively. The findings suggest that ASD patients may be appropriately classified utilizing ASD adult screening data and the DNN classification methodology. Meanwhile, missing values caused 95 rows of data in the second dataset to be eliminated, resulting in a considerable reduction in the number of instances in the dataset.

Astha Baranwal [4] et.al., presents the Autistic Spectrum Disorder Screening: Prediction with Machine Learning Models were the ASD screening dataset is used in this study to analyze and forecast probable cases in adults, infants and adolescents. Each age group's data is analyzed, and conclusions are generated as a result. For estimation and analysis, machine learning techniques such as Artificial Neural Networks (ANN), Random Forest, Logistic Regression, Decision Tree, and Support Vector Machines (SVM) were implemented. The Author as used the Autism Screening Datasets for adults, adolescents, and infants. Each dataset has 20 properties with ongoing segmented, and true or false values. Whether or not a person has ASD (1) is determined by the dependent attributes Class ASD (0). For Decision Tree, Artificial, Random Forest, Logistic Regression, Support Vector Machines (SVM), and Neural Networks, the accuracy is 80 %, 88 %, 92 %, 80%, and 76%, respectively (ANN). Apparently, the information available is inadequate. The information needed to build a successful statistical method are extremely limited. Across all collections, the Decision Tree produces an overcomplete architecture.

Azian Azamimi Abdullah [5] et.al., presents the Machine Learning Algorithms for Autism Diagnosis Identification are just being investigated (ASD), Logistic regression, random forest, and K-Nearest Neighbor are the methods employed in this study. There are 704 instances and 20 features in this data set, with one output. There are ten shorter versions of the AQ, self-administered questionnaires that examine behavior variables reflected in binary data (0 and 1). After information collection, a phase of the data extraction phase is data immediate post. It contains missing values, which necessitates the use of appropriate techniques.

3. METHODOLOGY

The approach concentrates on a novel machine learning system for classification and prediction of Autism Spectrum Disorder (ASD) to overcome an emerging problem. AdaBooster algorithms allow the model to improve efficiency and accuracy by using Random Forest (RF), Support Vector Machine (SVM). These algorithms, as well as other methods used in the study, are detailed below.

3.1. Random forests

Random forests, also based on random decision forests, are an ensemble learning method for classification, regression, and other tasks that functions by constructing a large number of decision trees during training and then transmitting the class which is the mean prediction of the individual trees.
Step 1 – Begin by selecting random samples from a given dataset.

Step 2 – Regarding that, this algorithm will generate a decision tree for each sample. The prediction result from each decision tree will then be obtained.

Step 3 – Voting will take place in this phase for each predicted outcome.

3.2. Support Vector Machines

Support Vector Machines (SVMs), also defined as support vector networks, are supervised learning models that analyses data for classification and regression analysis utilizing associated learning algorithms. A separating hyperplane formally defines it. In other words, the algorithm produces an optimal hyperplane that classifies new examples given labelled training data (supervised learning). Similarly, based on a massive corpus of documents that have already been identified as spam or non-spam by humans, may identify new emails as spam or non-spam. In such scenarios, SVMs are extremely useful. A Support Vector Machine (SVM) simulates the situation by generating a feature space, which is a finite-dimensional vector space with each dimension representing a feature of a specific entity. Each feature in the sense of spam or document classification is the frequency or value of a specific term.

3.3. AdaBoost

AdaBoost is one of several boosting algorithms being used in solving activities being AdaBoost. AdaBooster aims to combine numerous weak classifiers into a centralized control class. On this are a few important tidbits regarding AdaBooster:

• In AdaBoost, the weak learners are decision trees with a complete and unbiased, popularly known as decision stumps.
• AdaBoost effectively gives more weight to instances those are hard to characterize and less to anyone who is already well classified.
• AdaBoost methods is used to tackle issues in classification and regression.

4. RESULTS

A result is the qualitative or quantitative expression of the outcome of actions or events. Performance analysis is a type of operational analysis that involves establishing a set of basic quantitative relationships between performance variables. There are a wide range of performance metrics that can be used to assess a predictor's performance, including accuracy, recall, and others.

Table 1: Comparison Analysis for prediction model using machine learning

Algorithm	Precision	Recall	F1-Score	Accuracy
Random forest	96	99	96	96.68
AdaBoost	100	100	100	100
SVM	76	99	86	79.62

5. CONCLUSION

This study presents a model for predicting autism features that has been developed. The developed approach has the potential to detect autism side effects in children under the age of three, which is a feature that many other existing models lack. Using the AQ-10 dataset and real the sample which the proposed approach can use identify autism in case of random forest classifier, AdaBooster and support vector machine algorithms correspondingly. When comparing the three machine learning techniques, AdaBooster and Random forest classifier are the most effective for better ASD detection. When compared to other established autism screening methods this solution significantly improved.

REFERENCES

[1] Raj, Suman; Masood, Sarfaraz (2018). Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques. Procedia Computer Science, 167(), 994–1004. doi:10.1016/j.procs.2020.03.399.
[2] Noora Saleem Jumaa, Aymen Dawood Salman, Rafah Al-Hamdani, "The Autism Spectrum Disorder Diagnosis Based on Machine Learning Techniques" Journal of Xi'an University of Architecture & Technology, Volume XII, Issue V, 2020.
[3] M. F. Misman et al., "Classification of Adults with Autism Spectrum Disorder using Deep Neural Network," 2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS), 2019, pp. 29-34, doi: 10.1109/AiDAS47888.2019.8970823.
[4] A. Baranwal and M. Vanitha, "Autistic Spectrum Disorder Screening: Prediction with Machine Learning Models," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, pp. 1-7, doi: 10.1109/ic-ETITE47903.2020.186.
[5] Azian Azamimi Abdullah, Saroja Rijal and Satya Ranjan Dash., "Evaluation on Machine Learning Algorithms for Classification of Autism Spectrum Disorder (ASD)", School of Mechatronic Engineering, Universiti Malaysia Perlis, Pauh Putra Campus, 02600 Arau, Perlis, Malaysia, School of Computer Applications, KIIT University, India, 2019 J. Phys.: Conf. Ser. 1372 012052.