International Journal of Scientific Research & Engineering Trends
Volume 7, Issue 4, July-Aug-2020, ISSN (Online): 2395-566X

Detection of Autism Spectrum Disorder Using Machine Learning

Kruthi C H, Tejashwini H N , Poojitha G S, Shreelakshmi H S, Asst. Prof. Shobha Chandra K
Department of Computer Science and Engineering,
Malnad College of Engineering,
Karnataka, India.

***Abstract*- Autism spectrum Disorder (ASD) also known as Neuro-developmental disorder that affects people's interaction, learning and communication skills. Despite the fact that identification of this syndrome can be done at any age, its symptoms typically appear in the first 2 years of life. Determining the autism traits through screening tests is extravagant and extensive. With the improvement of artificial intelligence and machine learning, autism can be detected at quite an early stage. However a number of studies have been done using different techniques, these studies did not provide any final conclusion about detecting autism traits among the people of age group 3 years and below. Hence the paper aims to present an autism prediction model using machine learning techniques and to develop a web application that could successfully predict autism traits of a person. In other words, this paper focuses on developing an autism screening application for detecting the autism spectrum disorder traits among the people of age 3 years and below. This project model was evaluated with AQ 10 dataset (1054 datasets) and 50 actual dataset collected from people with and without autism traits. The Analysis results showed that the proposed prediction model gives better results in terms of specificity, precision, sensitivity, accuracy and f1 score for both datasets.**

***Keywords*- AQ-10 dataset, ASD, machine learning, specificity, accuracy, sensitivity, f1-score, precision.**

I.INTRODUCTION

Autism spectrum disorder (ASD) is a cognitive disability that can cause serious social, communication and behavioural challenges and its traits generally seen in the first two years of life and gradually develop through time.

People with autism face different kinds of struggles with learning disabilities, concentration, mental health issues like depression, anxiety and so on, motor difficulties, sensory issues and others. Experimentation suggests that both the genes and environment play significant roles.

According to the WHO's statistical data, 0.63% of children are diagnosed with Autistic Spectrum Disorder (ASD). ASD occurs in children and continues to expand to adolescence and adults. In most cases, symptoms appear within the first 5 years of life. Present day explosion rate of autism around the world is plentiful and it is growing at a very high rate. As stated by WHO about 1 out of every 160 individuals has autism spectrum disorder.

Few people with this disorder can live without depending on anyone while others require lifelong support and care. Identification of autism needs a significant amount of cost and time. Detection of autism earlier can come to much help by advising people with proper medication at an early stage. This could prevent the patient's condition from getting worse and would decrease long term costs related to delayed diagnosis.

Thus an effective, accurate and easy screening check tool is highly required which would detect the traits in a person and recognize whether the person requires thorough autism syndrome assessment or not. In this paper we use machine learning to find out a set of conditions that are put together to be predictive of autism spectrum disorder.

This can be immensely useful to physicians, assisting them to detect autism spectrum disorder at a very early stage.

II. LITERATURE SURVEY

D. P Wall [1] used Alternating Decision Tree (ADTree) for faster detection and reducing the screening time of Autism traits. Wall used the Autism Diagnostic Interview-Revised (ADI-R) method and achieved a high level of accuracy for the data of 891. The test failed to predict Autism Spectrum Disorder for different age groups as it was limited to the age of 5 to 17.

Daniel Bone [2] applied ML for the same purpose and used Support Vector Machine (SVM) and achieved 89.2% sensitivity and 59% specificity. His research included 462 individuals with NON-ASD and 1264 with ASD traits. As a screening approach their research was not accepted for all the age groups, due to wide range of age (4-55 years).

C. Allison [3] used the tool 'Red Flags' for screening Autism with Autism Spectrum Quotient for adults and children, then shortlisted them to AQ-10 with over 90% accuracy.

Van D Bekerom [4] used many of the ML techniques such as Random Forest, Naive Bayes and SVM algorithms to detect Autism traits in children like delay in development, less physical activity, obesity and compared the result.

III. SYSTEM DESIGN

The projected system use Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT) and Ada Boost (ADA) algorithms to predict the autism spectrum disorder (ASD) in terms of specificity, accuracy, sensitivity, precision and f1-score.

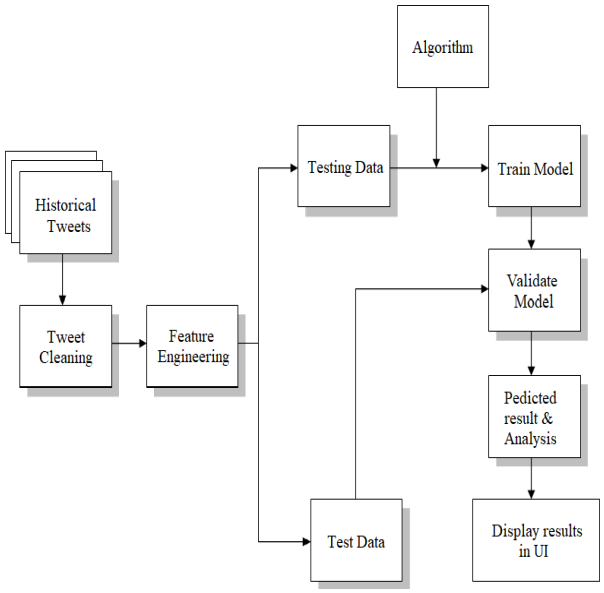

Fig 1. System Architecture.

The System architecture identifies the structure of the projected model. The Historical tweets i.e., the datasets taken from the UC repository. In Tweet cleansing, by removing unwanted data information the data sets are pre processed and converted into binary values to analyse.

Then the feature extraction is carried out by extracting the best features from the raw data using machine learning techniques. These features can improve the performance of the projected model. The dataset is split into 80% of the training dataset and 20% of testing dataset.

The projected model is developed by using the supervised machine learning algorithms with the training dataset. Then the trained model is evaluated with a testing dataset to analyse and predict the result supported by accuracy, precision, recall and f1-score and the final result is displayed in UI.

IV. PROPOSED METHODOLOGY

The work has five different phases and the same is briefly discussed below.

1. Data Collection:
AQ-10 screening queries for toddlers focus on totally different domains such as attention to detail, attention shift, communication, imagination and social interaction. Analysis methodology of the queries is that just one purpose scored for each of the 10 queries. Users may score zero or one on each question supporting their answer.

Datasets of toddlers contain 1054 instances. The data sets contain fifteen attributes that are a mixture of numerical and categorical data that includes: Age, Gender, if born with Jaundice, family members with ASD, who is taking the test.

2.Data Synthesization:
The collected data is synthesized to induce to eliminate unsuitable features. To handle null values, list wise deletion technique is applied where a selected observation was deleted if it had one or tons of missing values.

3. Developing a Prediction Model:
To generate prediction of ASD, algorithms are implemented and then their accuracy will be tested. The training data must contain the correct answers that are known as a target attribute. The learning algorithm finds patterns in the training data that map the input data attributes to the target and its output is an ML model that captures these patterns.

Results from various types of supervised learning like SVM, Random Forest, Ada Boost, Decision tree will be compared.

3.1 Random Forest: Random forests or random decision forests are an ensemble learning methodology for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class is mean prediction or mode of the classes of the individual trees.

3.2 SVM: A support vector machine (SVM) is a machine learning formula that analyses data for classification and regression methods. SVM is a supervised learning methodology that looks at data and sorts it into any one of two categories.

3.3 AdaBoost: AdaBoost was one of the first boosting algorithms to be adopted in solving practices. AdaBoost helps to combine two or more "weak classifiers" into one "strong classifier". AdaBoost works by putting more weight on difficult to classify instances and fewer on those which are already handled well. Adaboost can be used for regression as well as classification problems.

3.4 Decision Tree: Decision tree algorithm falls under the category of supervised learning. It is used to solve both

regression and classification problems. Decision trees use the tree illustration to resolve the matter through which each and every leaf node corresponds to a class label and attributes are represented on the internal node of the tree.

4. Evaluating the Prediction Model:
The projected predictive model will be tested with the AQ-10 data set and data collected from the real world in terms of accuracy, specificity, sensitivity, precision and f1 score.

5. Developing a web Application:
Finally, a web application is developed for predicting ASD in toddlers. Web application discusses matters related to the sign, causes and symptoms to give a brief description of ASD. By answering a set of closed ended queries, users will get the results of having or not having autism traits.

V. ASD DATASET AND ATTRIBUTE ANALYSIS

The dataset of autism has fifteen attributes that are utilized for feature analysis particularly to determine influential traits and improvise the classification of ASD. The dataset has a combination of ten binary attributes, one continuous attribute, and also includes four categorical attributes and one binary class variable.

The dataset contains ten behavioural and few individual characteristics. These attributes are verified effectively in detecting autism spectrum disorder in infants of three years and below. Table I describes the attributes of ASD toddler dataset.

Table 1. Attribute Description.

S.N	Attribute	Description	PossibleValues
1.	A1-A10 Score	Result of Diagnosis Questionaire	1 = Yes, 0 = No
2.	Age	Age of diagnostic patient, Continuous Values	Age below3 years
3.	Gender	Patient gender	m - male,f - female
4.	Jaundice	Jaundice effected immediately after birth or not	1 = Yes, 0 = No
5.	Family with ASD	Existence of ASD in familyhierarchy	1 = Yes, 0 = No
6.	Who completed thetest	Person who tookthe test	Family members, Medical staff, Others
7.	Class	Class variable, Classified underASD or not	1 = Yes, 0 = No

VI. RESULTS

The classification performance is calculated by True Positive (TP), False Negative (FN), False Positive (FP) and True Negative (TN).

1. Precision:
The percentage of properly classified positives by the model.

$$\text{Precision} = \frac{TP}{TP + FP}$$

2. Recall:
The ratio of actual positive to the positives classified by the model.

$$\text{Recall} = \frac{TP}{TP + FN}$$

3. Classification Report:
The classification report shows an illustration of the main classification metrics on a per- class basis.

4. Accuracy:
It tells how close the predicted value is to the actual value.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

5. Confusion Matrix:
A confusion matrix is a table that is used to show the performance of a classification model on a set of data whose true values are known.

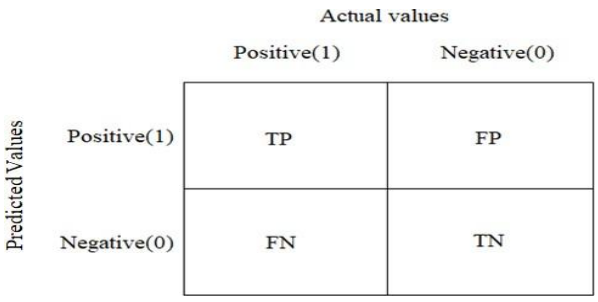

Fig 2. Confusion Matrix.

6. Decision Tree Model:
Accuracy of Decision Tree model using test-set is 92.89%

	Prediction	recall	f1-score	support
0	0.93	0.87	0.90	78
1	0.93	0.96	0.94	133
accuracy			0.93	211
macro avg	0.93	0.92	0.92	211
weighted avg	0.93	0.93	0.93	211

Fig 3. Classification Report of decision Tree Model.

```
[ [ 68  10]
  [  5 128] ]
```

Fig 4. Confusion matrix of Decision Tree Model.

7. Random Forest Model:
Accuracy of Random Forest model using test-set is 96.20%

	Prediction	recall	f1-score	support
0	0.99	0.94	0.96	78
1	0.96	0.99	0.98	133
accuracy			0.97	211
macro avg	0.97	0.96	0.97	211
weighted avg	0.99	0.99	0.99	211

Fig 5. Classification Report of Random Forest Model.

```
[ [ 73   5]
  [  1 132] ]
```

Fig 6. Confusion matrix of Random Forest Model.

8. AdaBoost Model:
Accuracy of AdaBoost model using test-set is 100%

	Prediction	recall	f1-score	support
0	1.00	1.00	1.00	78
1	1.00	1.00	1.00	133
accuracy			1.00	211
macro avg	1.00	1.00	1.00	211
weighted avg	1.00	1.00	1.00	211

Fig 7. Classification Report of AdaBoost Model.

```
[ [ 78   0]
  [  0 133] ]
```

Fig 8. Confusion matrix of AdaBoost Model.

9. SVM Model:
Accuracy of SVM model using test-set is 79.14%

	Prediction	recall	f1-score	support
0	1.00	0.44	0.61	78
1	0.75	1.00	0.86	133
accuracy			0.79	211
macro avg	0.88	0.72	0.73	211
weighted avg	0.84	0.79	0.77	211

Fig.9. Classification Report of SVM Model.

```
[ [ 34  44]
  [  0 133] ]
```

Fig 10. Confusion matrix of SVM Model.

Table II show the performance analysis of ASD prediction using supervised machine learning techniques.

Table 2. Performance Analysis.

Algorithm	Precision	Recall	Accuracy	Error Rate
Decision Tree	92.75	96.24	92.89	7.10
Random Forest	94.96	99.24	96.20	3.79
Ada Boost	100.00	100.00	100.00	0
SVM	75.14	100.00	79.14	20.85

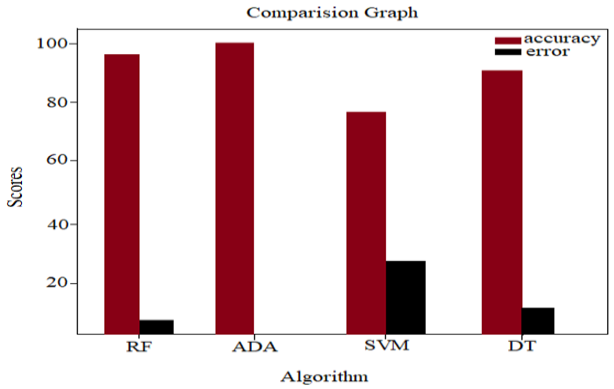

Fig 11. Comparison of Algorithms.

VII. CONCLUSION

This paper gives a prediction model that is developed to detect autism traits. By using AQ-10 dataset, the developed model predicts the Autism Spectrum Disorder with 92.89%, 96.20%, 100% and 79.14% accuracy in case of Decision tree, Random forest, Adaboost and SVM algorithms for better detection of autism spectrum disorder. Comparing all four supervised machine learning algorithms, Adaboost and Random forest algorithms are effective for better detection of autism spectrum disorder.

This result showed better performance compared to the previously existing approach of screening autism. Moreover, this model will predict autism traits for age groups of 3 years and below, but several other existing systems missed this feature. A user-friendly Web application has been developed for the end users based on this prediction model so that people can use the application to predict autism spectrum disorder easily.

The end result of this research gives an effective approach to detect autism for the age group of 3 years and below. As diagnosing autism traits is sort of an expensive and long process, it is usually delayed due to issues faced in detecting autism in toddlers. With the assistance of this autism screening web application a person can be guided at an early stage which will prevent the case from becoming much worse and cut back on expenses associated with delayed diagnosis.

VIII. LIMITATIONS AND FUTURE WORK

The main limitation of the work is insufficient data to train the prediction model. Another limitation is, the screening application is not designed for the age group of three years and above. Our future work focuses on collecting additional data from various different sources and to improvise the developed machine learning classifier to enhance its accuracy. And to develop a mobile application for all age groups.

REFERENCE

[1] Daniss P. Wall, R. Dally, R. Luyster, J.-Y. Jung, and T. F. DeLuca, “Use of artificial intelligence to shorten the behavioural diagnosis of autism” PloS one, vol. 7, no. 8, p. e43855, 2012.

[2] Daniel.Bone, S.L.Bishop, M. P. Black, M. S. Goodwin, C. Lord, and “Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrumentfusion,” Journal of Child Psychology and Psychiatry, vol. 57, 2016.

[3] C. Allison, B. Auyeung, and S. Baron-Cohen, “Toward brief “red flags”for autism screening: the short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls,” Journal of the American Academy of Child & Adolescent Psychiatry, vol. 51, 2012.

[4] Van den Bekerom, “Using machine learning for detection of autism spectrum disorder,” 2017.

[5] Kazi Shahrukh Omar, Prodipta Mondal, Nabila Shahnaz Khan, Md. Rezaul Karim Rizvi, Md Nazrul Islam, “A Machine Learning Approach to Predict syndrome Spectrum Disorder”, International Conference on Electrical, pc and Communication Engineering (ECCE), February, 2019.

[6] Milan N. Parikh, Hailong Li and Lili He “Enhancing identification of syndrome With Optimized Machine Learning Models and private Characteristic information ”, Perinatal Institute, Children’s Hospital heart, Cincinnati, OH, u.s., Department of pediatric medicine, University of city faculty of drugs, Cincinnati, OH, u. s..

[7] Fatiha Nur Büyükoflaz Bilgisayar Mühendis ligi Bölümü. KTO Karatay Üniversitesi, “Early syndrome identification of youngsters with Machine Learning Algorithms”, 2018, IEEE

[8] T. Hindu deity Praveena and N. V. Muthu Hindu deity, “Prediction of syndrome Spectrum Disorder victimization supervised Machine Learning Algorithms”, 2019, Asian Journal of applied science and Technology.

[9] D. Wall, J. Kosmicki, T. Deluca, E. Harstad, and V. Fusaro, “Use of machine learning to shorten observation- based screening and diagnosis of autism,” Translational psychiatry, vol. 2, no. 4, p. e100, 2012.

[10] A.S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, andF. Meneguzzi, “Identification of autism spectrum disorder using deep learning and the abide dataset,” Neuro Image: Clinical, vol. 17, 2018.

[11] W. Liu, M. Li, and L. Yi, “Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework,” Autism Research, vol. 9, no. 8, pp. 888–898, 2016.

[12] D. Bone, M. S. Goodwin, M. P. Black, C.-C. Lee, K. Audhkhasi,and S. Narayanan, “Applying machine learning to facilitate autism diagnostics: pitfalls and promises,” Journal of autism and developmental disorders, vol. 45, no. 5, pp. 1121–1136, 2015

[13] Bram van den Bekerom,“Using Machine Learning for Detection of syndrome Spectrum Disorder”, 2018, University of Twente.

[14] Scott H. LeeID, Matthew J. Maenner, Charles M. Heilig, “A comparison of machine learning algorithms for the police investigation of syndrome spectrum disorder”, 2019, American state University Purdue, University at state capital, u.s.

[15] Anurekha G and Geetha P, “Performance Analysis of supervised Approaches for syndrome Spectrum Disorder Detection”, International Journal of Trend in analysis and Development (IJTRD) 2017.