International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

Engagement Detection with Autism Spectrum Disorder using Machine Learning

Aishwarya J[1], Akshatha N[2], Anusha H[3], Shishira J[4], Deepa Mahadev[5]

[1,2,3,4]UG Student, Dept. of Computer Science & Engineering, BNM Institute of Technology, Karnataka, India
[5]Assistant Professor Deepa Mahadev, Dept. of Computer Science & Engineering, BNM Institute of Technology, Bengaluru, Karnataka, India

---

Abstract - *Autism spectrum disorder (ASD) is a neurodevelopment disorder that affects a person's interaction, communication and learning skills. Although diagnosis of autism can be done at any age, its symptoms generally appear in the first two years of life and develop through time. Detecting autism traits through screening tests is very expensive and time consuming. With the advancement of artificial intelligence and machine learning (ML), autism can be predicted at quite early stage. Though a number of studies have been carried out using different techniques, these studies didn't provide any definitive conclusion about predicting autism traits among people of age group 3 years and below.*

Therefore the paper aims to propose an autism prediction model using ML techniques and to develop a web application that could effectively predict autism traits of an individual. In other words, this work focuses on developing an autism screening application for predicting the ASD traits among people of age group 3 years and below. The proposed model was evaluated with AQ-10 dataset (1054 datasets) and 50 real dataset collected from people with and without autistic traits. The evaluation results showed that the proposed prediction model provide better results in terms of accuracy, specificity, sensitivity, precision and f1-score for both kinds of datasets.

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

1. INTRODUCTION

Autism spectrum disorder (ASD) is a developmental disability that can cause significant social, communication and behavioral challenges and its symptoms generally appear in the first two years of life and develop through time. Autism patients face different types of challenges such as difficulties with concentration, learning disabilities, mental health problems such as anxiety, depression etc., motor difficulties, sensory problems and many others. Research suggests that both genes and environment play important roles.

Current explosion rate of autism around the world is numerous and it is increasing at a very high rate. According to WHO, about 1 out of every 160 has ASD. Some people with this disorder can live independently, while others require life-long care and support. Diagnosis of autism requires significant amount of time and cost. Earlier detection of autism can come to a great help by prescribing patients with proper medication at an early stage. It can prevent the patient's condition from deteriorating further and would help to reduce long term costs associated with delayed diagnosis.

Thus a time efficient, accurate and easy screening test tool is very much required which would predict autism traits in an individual and identify whether or not they require comprehensive autism assessment. In this research we use machine learning to determine a set of conditions that together prove to be predictive of Autism Spectrum Disorder. This will be a great use to physicians, helping them detect Autism Spectrum Disorder at a much earlier stage.

2. SYSTEM DESIGN

The proposed system makes use of Random Forest (RF), Support Vector Machine (SVM), Decision Tree and Ada Booster algorithms to predict autism spectrum disorder in an individual in terms of accuracy, specificity, sensitivity, precision and f1-score.

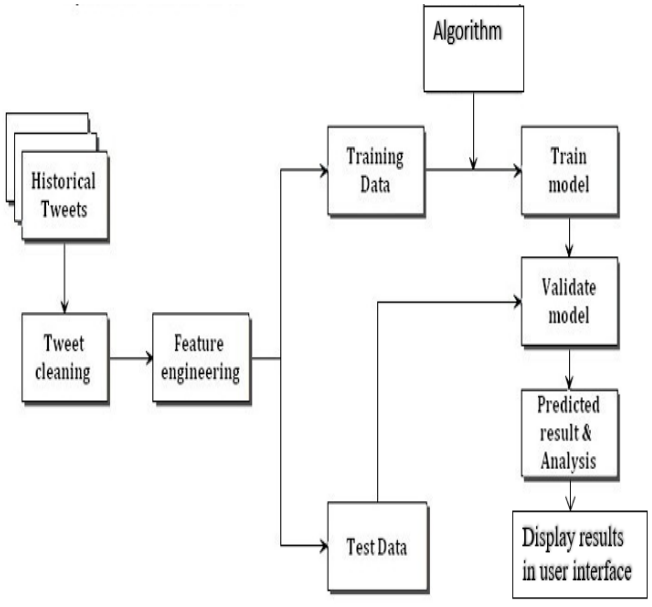

Fig -1: System Architecture

The System architecture design identifies the overall structure of the proposed model. The historical tweets i.e., the datasets that has been taken from the UCI repository. In tweet cleaning, the datasets are preprocessed by removing unwanted data and are converted into binary values for the machine to parse easily. Then feature extraction is performed by extracting the best features from the raw data via machine learning techniques. These features can improve the performance of the proposed model. The dataset is split into 80% of training dataset and 20% of testing dataset. The proposed model is developed by providing the supervised machine learning algorithms with the training dataset. The trained model is then evaluated with a testing dataset to analyze and predict the result based on accuracy, precision, recall and f1-score. The final result is displayed in user interface.

3. METHODOLOGY

The work is carried out in five phases: Data collection, Data synthesization, Developing the prediction model, Evaluating the prediction model and Developing a Web Application. The phases are briefly discussed in the following sub-sections:

A. Data collection

AQ-10 screening questions for toddlers focus on different domains such as- attention to detail, attention switching, communication, imagination and social interaction. Scoring method of the questions is that only 1 point can be scored for each of the 10 questions. User may score 0 or 1 point on each question based on their answer. Datasets of toddlers contain 1054 instances. The datasets contains fifteen attributes which are a mix of numerical and categorical data that includes: Age, Gender, if born with Jaundice, Family member with ASD, Who is completing the test.

B. Data Synthesization

The collected data are synthesized to remove irrelevant features. To handle null values, list wise deletion technique is applied where a particular observation was deleted if it had one or more missing values.

C. Developing a Prediction Model

To generate prediction of autism traits, algorithms will be implemented and their accuracy will be tested. The training data must contain the correct answer, which is known as a target or target attributes. The learning algorithm finds patterns in the training data that map the input data attributes to the target and it outputs an ML model that captures these patterns. Results from various types of supervised learning like SVM, Random Forest, AdaBoost, Decision tree will be compared.

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

SVM: A support vector machine (SVM) is machine learning algorithm that analyzes data for classification and regression analysis. SVM is a supervised learning method that looks at data and sorts it into one of two categories.

AdaBoost: AdaBoost is one of the first boosting algorithms to be adapted in solving practices. AdaBoost helps to combine multiple “weak classifiers” into a single “strong classifier”. AdaBoost works by putting more weight on difficult to classify instances and less on those already handled well. This algorithm can be used for both classification and regression problem.

Decision Tree: Decision tree algorithm falls under the category of supervised learning. They can be used to solve both regression and classification problems. Decision tree uses the tree representation to solve the problem in which each leaf node corresponds to a class label and attributes are represented on the internal node of the tree.

D. Evaluating the Prediction Model

The proposed predictive model will be tested with the AQ- 10 dataset and data collected from real-world in terms of the accuracy, specificity, precision and sensitivity.

E. Developing a Web Application

Finally, a web application is developed for predicting Autism spectrum disorder in toddlers. Web application discusses about the sign, causes and symptoms to give a brief description of ASD. By answering a set of closed ended questions, user will get a result of having or not having autism traits.

4. ASD DATASET AND ATTRIBUTES ANALYSIS

The autism dataset contains 15 attributes that will be utilized for feature analysis especially in determining influential autistic traits and improving the classification of ASD cases.

The dataset contains 10 behavioral characteristics and few individual characteristics. The attributes are combination of 10 binary attributes, 1 continuous attributes, 4 categorical attributes and 1 binary class variable. These attributes has proved to be effective in detecting Autism

spectrum disorder in toddler of age group 3 years and below.

TABLE I
ASD TODDLER DATASET ATTRIBUTE DESCRIPTION

Sl. No	Attribute Name	Description	Possible Values
1.	A1 to A10	Diagnosis Questionnaire result.	Yes-1,No-0
2.	Age in months	Age of the diagnostic patient. Continuous values	minimum value: 12 months and maximum value : 36 months
3.	Sex	Gender of the patient.	Male-m, Female-f
4.	Jaundice	Jaundice effected immediate after delivery or not	Yes-1,No-0
5.	Family with ASD	ASD existence in family hierarchy	Yes-1,No-0
6.	Who completed the test	Person who took the test	Family members, Health care professional, Self, Others
7.	Class	Class variable. Classified under ASD or not	Yes-1, No-0

5. RESULTS

The classification performance is calculated by counting the True Positive (TP), False Negative (FN), False Positive (FP) and True Negative (TN).
Recall: It is the ratio of actual positive to the positives classified correctly by the model.

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

Precision: It is the percentage of correctly classified positives by the model.

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

Classification report:The classification report shows a representation of the main classification metrics on a per-class basis.

Accuracy: It tells how close the predicted value is to the act ual value.

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

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

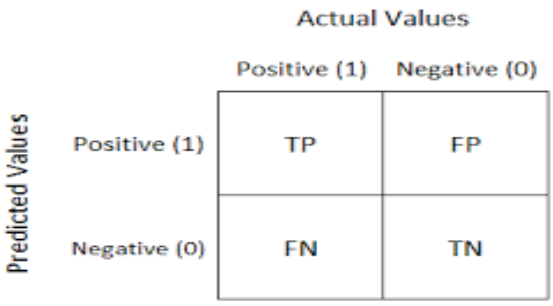

Fig -2: Confusion Matrix

5.1 Decision Tree Model

Overall accuracy of Decision Tree model using test-set is 92.89%

	precision	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

5.2 Random Forest Model

Overall accuracy of Random Forest model using test-set is 96.20%

	precision	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	0.97	0.97	0.97	211

Fig-5: Classification Report of Random Forest Model

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

Fig-6: Confusion matrix of Random Forest Model

5.3 AdaBoost Model

Overall accuracy of AdaBoost model using test-set is 100.00%

	precision	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

5.4 SVM Model

Overall accuracy of SVM model using test-set is 79.14%

	precision	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
PERFORMANCE ANALYSIS FOR ASD PREDICTION USING SUPERVISED MACHINE LEARNING APPROACHES

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
AdaBoost	100.00	100.00	100.00	0
SVM	75.14	100.00	79.14	20.85

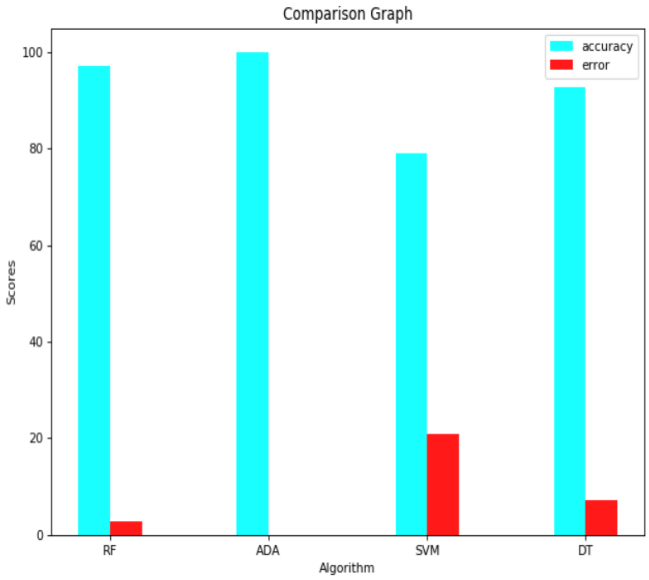

Fig-11: Comparison of Algorithms

6. CONCLUSION

This research provides a prediction model that is developed to predict autism traits. Using the AQ-10 dataset, the proposed model can predict autism with 92.89%, 96.20%, 100.00% and 79.14% accuracy in case of Decision tree, Random forest, Adaboost and SVM algorithms respectively. Comparing all four supervised machine learning algorithms, AdaBoost and Random Forest algorithm are efficient algorithms for better prediction of ASD. This result showed better performance comparing to the other existing approach of screening autism. Moreover, the proposed model can predict autism traits for age groups below 3 years, while many other existing approaches missed this feature.

A user-friendly Web application has been developed for end users based on the proposed prediction model so that individual can use the application to predict the autism traits easily.

The outcome of this research provides an effective and efficient approach to detect autism traits for age groups 3 years and below. Since diagnosing the autism traits is quite a costly and lengthy process, it's often delayed because of the difficulty of detecting autism in toddlers. With the help of autism screening application, an individual can be guided at an early stage that will prevent the situation from getting any worse and reduce costs associated with delayed diagnosis.

Limitations and Future Work:

The primary limitation of the study is lack of sufficiently large data to train the prediction model. Another limitation is that, the screening application is not designed for age group above 3 years.

Our future work will focus to collect more data from various sources and to improve the proposed machine learning classifier to enhance its accuracy and also develop a mobile application for all age groups.

REFERENCES

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

[2] Milan N. Parikh, Hailong Li and Lili He "Enhancing Diagnosis of Autism With Optimized Machine Learning Models and Personal Characteristic Data ", 2019, Perinatal Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States, Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, United States.

[3] Fatiha Nur Büyükoflaz Bilgisayar Mühendisligi Bölümü. KTO Karatay Üniversitesi, "Early Autism Diagnosis of Children with Machine Learning Algorithms", 2018, IEEE

[4] T. Lakshmi Praveena and N. V. Muthu Lakshmi, "Prediction of Autism Spectrum Disorder Using Supervised Machine Learning Algorithms", 2019, Asian Journal of Computer Science and Technology.

[5] Bram van den Bekerom, "Using Machine Learning for Detection of Autism Spectrum Disorder", 2018, University of Twente.

[6] Scott H. LeeID, Matthew J. Maenner, Charles M. Heilig, "A comparison of machine learning algorithms for the surveillance of autism spectrum disorder", 2019, Indiana University Purdue, University at Indianapolis, UNITED STATES.

[7] Anurekha G and Geetha P, "Performance Analysis of Supervised Approaches for Autism Spectrum Disorder Detection", International Journal of Trend in Research and Development (IJTRD) 2017.

BIOGRAPHIES

Aishwarya J,
BE final year,
Dept. of Computer Science & Engineering,
BNM Institute of Technology

Akshatha N,
BE final year,
Dept. of Computer Science & Engineering,
BNM Institute of Technology

Anusha H,
BE final year,
Dept. of Computer Science & Engineering,
BNM Institute of Technology

Shishira J,
BE final year,
Dept. of Computer Science & Engineering,
BNM Institute of Technology

Mrs. Deepa Mahadev,
Assistant Professor,
Dept. of Computer Science & Engineering,
BNM Institute of Technology