



# A Machine Learning Approach to Predict Autism Spectrum Disorder

<sup>1</sup>SNEHA N RAJ, <sup>2</sup>SINCHANA K N, <sup>3</sup>NISARGA T A, <sup>4</sup>MEGHANA G R, <sup>5</sup>MANU Y M

Student, Information science department, B G S Institute of technology B G Nagara, Mandya(di),Karnataka, India<sup>1-4</sup>

Assistant Professor, Dept., of Computer science & Engineering, B G S Institute of technology B G Nagara,

Mandya(di), Karnataka, India<sup>5</sup>

**Abstract:** Today, autism spectrum disorder (ASD) is developing faster than ever. Screening tests to detect the characteristics of autism is very expensive and time-consuming. with Advances in artificial intelligence and machine learning (ML), autism can It was predicted very early. Although a lot of research has been done to use Different technologies, these studies did not provide any clear conclusions about the prediction The characteristics of autism in different age groups. The proposed model is evaluated using the AQ 10 data set and 1000 real data sets collected from people with and without autism characteristics. In light of this, we are developing a model in machine learning algorithms namely SVM ,Adaboost and Random forest.

**IndexTerms** - SVM ,Adaboost and Random forest.

## I. INTRODUCTION

Autism spectrum disorder is a neurodevelopmental issue that influences an individual's association, correspondence and acquiring abilities. In spite of the fact that determination of mental imbalance should be possible at whatever stage in life, its side effects for the most part show up in the initial two years of life and creates through time. Chemical imbalance patients face various kinds of difficulties like challenges with fixation, learning handicaps, psychological well-being issues like tension, sadness and so on, engine troubles, tactile issues and numerous others.

Current blast pace of autism all throughout the planet is various and it is expanding at a high rate. As per WHO, around 1 out of each 160 kids has ASD. A few group with this problem can live autonomously, while others require long lasting consideration and backing.

Analysis of autism requires critical measure of time and cost. Prior discovery of autism can go to an extraordinary assistance by recommending patients with appropriate prescription at an early stage. It can keep the patient's condition from decaying further and would assist with decreasing long haul costs related with deferred determination. Subsequently a period productive, precise and simple screening test instrument is a lot of required which would foresee autism qualities in an individual and distinguish whether they require complete autism appraisal.

The target of this work is to propose a autism forecast model utilizing ML procedures and to foster a portable application that could adequately anticipate autism characteristics of a person of all ages. At the end of the day, this work centers around fostering a autism evaluating application for anticipating the ASD attributes among individuals old enough gatherings 4-11 years, 12-17 years and for individuals old enough 18 and more.

## II. ANOMALY DETECTION RECENT SURVEYS

Autism Spectrum Disorder (ASD) is a disorder joined by intellectual and social deficiencies (i.e., absence of visual eye to eye connection, helpless social mindfulness, and stereotypic practices), which thus prevents practical characteristics of freedom and social communications (American Mental Affiliation, 2013). Current reports gauge the quantity of kids determined to have ASD as 1 out of 68 (Community for Infectious prevention and Counteraction [CDC], 2018), with consistent rates expanding throughout the most recent 20 years (Negggers, 2014).

Additionally, there is an adequate examination pool to be show worry for personal satisfaction for guardians of kids with ASD (Khanna et al.,2011). For example, the way toward nurturing a kid with ASD has been found to emphatically correspond with higher nurturing pressure contrasted with those guardians of kids without ASD (Bread cook Ericzén, Brookman-Fraze, and Stahmer, 2005).



Parental pressure mediations can be conveyed in an assortment of ways, including however not restricted to: (a) parent support gatherings, (b) parent distributed tutoring, directing/treatment, or (c) web/electronic exploration and backing (Marshall, Noonan, McCartney, Marx, and Keefe, 2001; Artist et al., 1999). The reason for this paper is to investigate a particular type of help for guardians of youngsters with ASD, which is conveyed through The Autism Task in Rhode Island. In particular, this paper will depict and dissect the viability of The Autism Tasks' site and programming, with an essential zero in on parent discernment, fulfillment, and recurrence of utilization.

The Autism project was formed by a gathering of teachers, guardians of kids with autism, and experts in 1994 who perceived a requirement for Autism related assets and preparing for the two schools and the local area (The Autism Undertaking, n.d.). The Autism project got introductory subsidizing through the Rhode Island of Division of Schooling in 1997, in any case, this financing finished in 2001 (The Autism Task). Since 2001, The Autism Venture has been financed through government awards, repayments through Medicaid, or private payers. Likewise, through an association with Door Inc., a Life expectancy member, The Autism Venture is essential for Life expectancy and is straightforwardly associated with the Rhode Island Consortium Autism Exploration Group (RI-Truck) at Bradley Kids' Medical clinic (individual correspondence with Joanne Quinn, February 6, 2018).

### III. METHODOLOGY

**A. Data Acquisition and Preprocessing:-** In this module we will get the data from the online source. Further we will resize the data for future use. data resizing, or data scaling, is a geometric data transformation which modifies the data size based on an data interpolation algorithm. This data scaling process can increase or decrease the resolution of a target data so that the absolute size of image data is adjusted.

**B. Data Preparation and Model construction:-** Many a times, people first split their dataset into 2—Train and Test. After this, they keep aside the Test set, and randomly choose X% of their Train dataset to be the actual Train set and the remaining (100-X)% to be the Validation set, where X is a fixed number(say 80%), the model is then iteratively trained and validated on these different sets. So we will follow the same method to prepare data for training and testing phase.

**C. Model training:-** After model construction it is time for model training. We were able to build an Random forest , SVM, Adabooster can recognize autism. Split the dataset into train and test dataset. Finally we will build and train the model using training dataset.

**D. Model testing and evaluation:-** Once the model has been trained it is possible to carry out model testing. During this phase a test set of data is loaded. This data set has never been seen by the model and therefore its true accuracy will be verified. Finally, the saved model can be used in the real world. The name of this phase is model evaluation. This means that the model can be used to evaluate new data.

### IV. IMPLEMENTATION

This system is developed using Python programming with Anaconda framework. Data labeled according to their relevant classification then feed into the model. Data are trained and tested with Random forest, SVM and Adaboost.

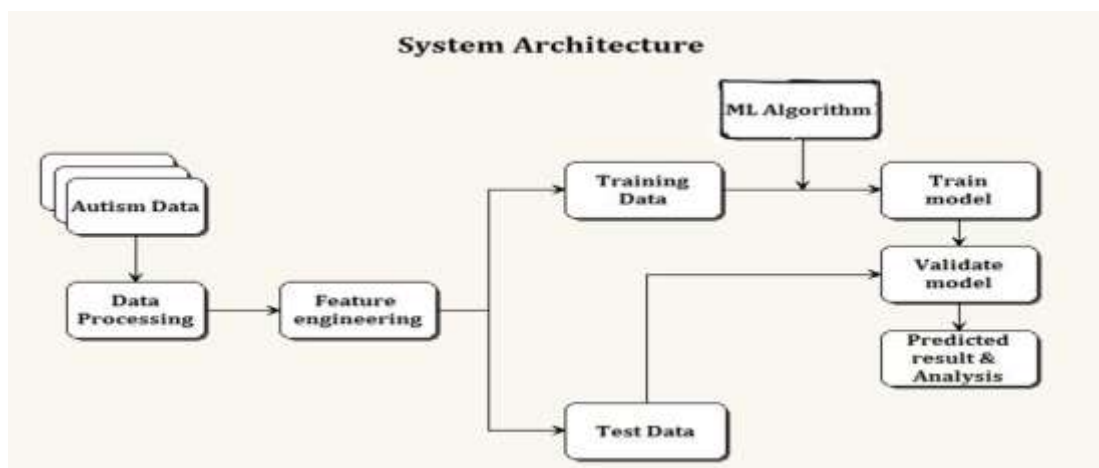


Fig 1: system architecture



#### 4.1 Implementation Steps

- Collect the data set
- Now build the algorithms Adaboost, SVM and Random forest.
- Split the data into training and testing.
- Using training data we are going to train our SVM, Random Forest and Adaboost.
- Once we trained the model utilizing the test data we are going to evaluate our model.

#### 4.2 ALGORITHMS

##### AdaBoost:

AdaBoost is one of the first boosting algorithms to be adapted in solving practices. Adaboost helps you **combine multiple “weak classifiers” into a single “strong classifier”**. Here are some (fun) facts about Adaboost!

- The weak learners in AdaBoost are decision trees with a single split, called decision stumps.
- AdaBoost works by putting more weight on difficult to classify instances and less on those already handled well.
- AdaBoost algorithms can be used for both classification and regression problem.

##### Random forests:

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.

- Step 1 – First, start with the selection of random samples from a given dataset.
- Step 2 – Next, this algorithm will construct a decision tree for every sample. Then it will get the prediction result from every decision tree.
- Step 3 – In this step, voting will be performed for every predicted result.

##### Support Vector Machines:

Support Vector Machines (SVMs, also support vector networks are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. It is formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning).

If we wish to categorize new unseen objects into two separate groups based on their properties and a set of known examples, which are already categorized. A good example of such a system is classifying a set of new documents into positive or negative sentiment groups, based on other documents which have already been classified as positive or negative. Similarly, we could classify new emails into spam or non-spam, based on a large corpus of documents that have already been marked as spam or non-spam by humans. SVMs are highly applicable to such situations.

## V. RESULTS & ANALYSIS

Using machine learning technique SVM, Adaboost and Random forest we build the model to predict autism spectrum disorder and the results shows 79% accuracy, 100% accuracy and 95% accuracy correspondingly with those three models.

### 5.1 Overall Accuracy for SVM , Adaboost and Random forest

- SVM gives us 79% accuracy.
- Adaboost gives us 100% accuracy.
- Random forest gives us 95% accuracy.

## VI. CONCLUSION AND FUTURE SCOPE

In this paper, a deep learning based models SVM, Adaboost and Random forest are proposed for detecting autism traits. To train and test the models 1000 data are used. The result shows Adaboost performing better than SVM model.

In conclusion, as identified through the literature review, we believe only a marginal success is achieved in the creation of predictive model for Autism spectrum disorder (ASD) patients and hence there is a need for combinational and more complex models to increase the accuracy of predicting the early onset of Autism spectrum disorder (ASD).



## REFERENCES

- [1] U. Frith and F. Happé, "Autism spectrum disorder," *Current biology*, vol. 15, no. 19, pp. R786–R790, 2005.
- [2] WHO, *Autism spectrum disorders, 2017* [Accessed August 22, 2018]. [Online]. Available: <http://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders>
- [3] J. A. Cruz and D. S. Wishart, "Applications of machine learning in cancer prediction and prognosis," *Cancer informatics*, vol. 2, 2006.
- [4] N. S. Khan, M. H. Muaz, A. Kabir, and M. N. Islam, "Diabetes predicting mhealth application using machine learning," in *2017 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*. IEEE, 2017, pp. 237–240.
- [5] D. P. Wall, R. Dally, R. Luyster, J.-Y. Jung, and T. F. DeLuca, "Use of artificial intelligence to shorten the behavioral diagnosis of autism," *PloS one*, vol. 7, no. 8, p. e43855, 2012.
- [6] D. Bone, S. L. Bishop, M. P. Black, M. S. Goodwin, C. Lord, and S. S. Narayanan, "Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion," *Journal of Child Psychology and Psychiatry*, vol. 57, 2016.
- [7] 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.
- [8] F. Thabtah, "Autism spectrum disorder screening: machine learning adaptation and dsm-5 fulfillment," in *Proceedings of the 1st International Conference on Medical and Health Informatics 2017*. ACM, 2017.
- [9] F. Hauck and N. Kliewer, "Machine learning for autism diagnostics: Applying support vector classification."
- [10] B. van den Bekerom, "Using machine learning for detection of autism spectrum disorder," 2017.
- [11] 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.
- [12] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identification of autism spectrum disorder using deep learning and the abide dataset," *NeuroImage: Clinical*, vol. 17, 2018.
- [13] 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.
- [14] 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.
- [15] J. Kosmicki, V. Sochat, M. Duda, and D. Wall, "Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning," *Translational psychiatry*, vol. 5, no. 2, p. e514, 2015.
- [16] F. Thabtah, "UCI machine learning repository," 2017. [Online]. Available: <https://archive.ics.uci.edu/ml>
- [17] T. Booth, A. L. Murray, K. McKenzie, R. Kuenssberg, M. O'Donnell, and H. Burnett, "An evaluation of the aq-10 as a brief screening instrument for asd in adults."
- [18] A. Liaw, M. Wiener et al., "Classification and regression by randomfor est," *R news*, vol. 2, no. 3, pp. 18–22, 20