



IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2320-6710



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 7, July 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.569



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

Early-Stage Prediction of Autism Spectrum Disorder by Machine Learning Approaches

M C Ranjitha^{*1}, Dr. Jalaja G^{*2}

PG Student, Department of CSE, BNM Institute of technology, Bangalore, Karnataka, India^{*1}

Associate Professor, Department of CSE, BNM Institute of technology, Bangalore, Karnataka, India^{*2}

ABSTRACT: Presently, Autism Spectrum Disorder (ASD) is acquiring additional traction than ever. It is indeed complex and potentially tedious to diagnose autistic traits by screening tests. Autism can be predicted at reasonably at an early stage with the implementation of artificial intelligence and machine learning (ML). While multiple studies using various techniques have been carried out on these and did not provide any definite conclusion on the prediction of description of autism in demographic variables. The proposed model is going to be AQ- 10 dataset and 1054 real dataset both with autistic tendencies it was used to hypothesise. The evaluation results may show by high accuracy, the proposed prediction model trumps competitors for both kinds of datasets. This paper therefore aims to develop an effective prediction model. The objective of this work is to propose an autism prediction model using ML techniques and to develop a mobile application that could effectively predict autism traits of an individual of any age. 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 strategy focuses on a machine learning procedure for autism spectrum disorder (ASD) classification and prediction, thus overcoming the existing problem. By utilizing Random Forest (RF), Support Vector Machine (SVM), AdaBooster algorithms will make the model to increase the performance and accuracy, precision, recall and f1-score.

KEYWORDS: Autism spectrum disorder, AdaBooster, Machine learning, Random Forest, Support Vector Machine.

I. INTRODUCTION

Autism spectrum disorder is a neurological condition that affects a person's ability to connect, interact, and learn. Autism may be diagnosed at any age, although signs often appear in the first two years of life and evolve with time. Autism is on the growing all over the world, and it's getting higher. Increasing at an incredible rate According to the WHO, ASD affects approximately one out of every 160 children. Any individuals with this condition are capable of surviving independently, while others will need lifelong care and assistance. Autism diagnosis takes a long time and gets expensive. Early diagnosis of autism can be very beneficial in terms of providing appropriate treatment to patients. It has the potential to prevent the patient's health from deteriorating further, as well as reduce the long-term costs of delayed diagnosis. As a result, a fast, precise, and simple screening test method is desperately needed to predict autism traits in individuals and determine whether or not they require a rigorous autism evaluation.

A dataset containing influential features that can be used for further analysis, especially in examining autistic traits and improving the classification of ASD cases, as well as autism screening of toddlers. This dataset includes ten behavioural traits (Q-Chat-10) as well as other human features that have been found to be important in separating ASD situations from behavioural science controls. The objective of this work is to build a smartphone application which could accurately predict autism traits in people of any age using machine learning techniques and to propose an autism prediction model. In other words, the purpose of this work is to build an autism screening app that can predict ASD traits in children aged above 36 months. The model's efficiency and accuracy can be enhanced by using the Random Forest, Support Vector Machine algorithms.

II. RELATED WORK

This section 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. The different research implements different methods in each of the steps namely pre-processing, Model, feature extraction and classification. The data samples used for training are very less in many of the models implemented. A considerable amount of time and cost is needed for the diagnosis of autism. Earlier autism diagnosis can be a great benefit by administering proper treatment to patients at an early stage.

Machine learning strategies to predict autism spectrum disorder are presented by Bhawana Tyagi [1] et al., autism spectrum disorder (ASD) is a major developmental abnormality that severely affects an individual's behaviour and expression. K-Nearest Neighbour, Linear regression, Linear discriminant analysis, Naïve Bayes, CART (Classification and regression Tree), and Support Vector Machine (SVM) are the machine learning techniques used here for data classification. The adult data set has 19 attributes and 702 occurrences, with one of the 19 attributes, score, consisting of the sum of all the values of those ten questions. The dataset's performance consists of two values: 1 indicates that the person has ASD, and 2 indicates that the person does not have ASD. The LDA algorithm, as a result of implementation, has the best test result, 72.2024%, and is the most accurate of the algorithms.

In this analysis, ASD adolescent scan data from the UCI machine learning repository is used. Demirhan, A [2], the author, presents the Success of machine learning methods in evaluating the autism spectrum disorder cases. The dataset includes 20 characteristics from 104 adolescents who were screened for autism. There are ten behavioural features and ten human features in this dataset that have been shown to be effective in separating ASD cases from controls. To detect ASD instances, support vector machine (SVM), K-Nearest Neighbour, and Random Forest machine learning methods were used. Binary classification using support vector machine (SVM), K-Nearest Neighbour, and Random Forest methods yielded accuracy rates of 95%, 89%, and 100%, respectively. The K-Nearest Neighbour algorithm achieves the lowest efficiency in the classification of ASD cases. The findings of the random forest approach show that it is fully efficient in classifying ASD instances.

Prediction of the Autism Spectrum Disorder Diagnosis with Linear Discriminant Analysis Classifier and K-Nearest Neighbour in Children was presented by Osman Altay [3] et al., in which the classification system for ASD diagnosis was used in children aged 4-11 years. The analysis makes use of a dataset with a wide range of questions. Classification approaches include linear discriminant analysis (LDA) and the K-Nearest Neighbour (KNN) algorithm. There are 292 samples in the dataset, each with 19 different attributes. There are 10 questions in the dataset that are specifically linked to ASD, as well as a score attribute that is the sum of the questions. The accuracy value was estimated as 0.9080% using the LDA algorithm. The accuracy value was estimated as 0.8851% using the KNN algorithm. As a consequence of the implementation, the LDA algorithm outperformed the KNN algorithm in terms of accuracy.

There have been several studies using machine learning approaches, where datasets are obtained from online, as presented by Dadang Eman [4] et al. in Machine Learning Classifiers for Autism Spectrum Disorder a study. Decision tree, Random Forest, Support vector machine (SVM), Nave bayes, generalized linear model, logistic regression, and K-Nearest Neighbour are the machine algorithms used here (KNN). With the use of machine learning in the case of ASD, it is supposed to be possible to speed up and enhance the precision of diagnosis. Manual evaluation (ADDS) and (ADI-R) are still the gold standard; however, the treatment takes a long time and is relatively costly to diagnose. Five classification models (nave Bayes, SVM RBF, SVM linear, decision tree, and random forest) have varying degrees of accuracy, but SVM has the highest accuracy. The use of the SVM classifier resulted in a classification and comparison of classification accuracy of 86.7%.

Azian Azamimi Abdullah [5] et.al., presents the Evaluation on Machine Learning Algorithms for Classification of Autism Spectrum Disorder (ASD), the methodology used here is logistic regression, random forest, K-Nearest Neighbour. In this data consists of 704 instances and 20 features with 1 output. Features consists of 10 shortened version AQ, self-administered questionnaires which assessed on behaviour traits which were represented in binary data (0 and 1). The removal of null data giving a total of 609 data left with 20 features consisting 19 features and 1 output. Six Random Forest models, five Logistic Regression models, and five K-Nearest Neighbour models have been developed as a result of the selection methods. The LR4-CV model from Logistic Regression had the best results, with 97.541 %, 96.591 % precision, and 100 percent sensitivity. The model RF4-CV scored 96.721 % and 98.864 % for Random Forest.

Detecting autism traits by screening tests is very costly and time consuming, according to Kazi Shahrukh Omar [6] et al., paper. A Machine Learning Approach to Predict Autism Spectrum Disorder. Autism can now be predicted at an early stage thanks to developments in artificial intelligence and machine learning (ML). An autism prediction model was developed by combining Random Forest-CART (Classification and Regression Trees) and Random Forest-ID3 (Iterative Dichotomiser 3). In other words, the purpose of this work is to build an autism screening app that can predict ASD traits in people aged 4-11 years (child), 12-17 years (adolescent), and 18 years (adult) and older. The AQ-10, or Autism Spectrum Quotient tool, is used to determine if a person should be referred for a full autism evaluation. Using

the AQ-10 dataset, the proposed model will reliably predict autism in infants, teenagers, and adults with 92.26 %, 93.78 %, and 97.10 %, respectively. The results revealed that the Random Forest-CART algorithm outperformed the Decision Tree-CART algorithm.

III. SYSTEM DESIGN

The aim of this model is to create an effective and accurate framework for predicting disease risk using multimodal risk prediction. To process the data, several steps are taken. The model was created to aid doctors in verifying their findings, which are based solely on visualizing the patient. Figure1 depicts the steps involved in the proposed architecture. The model being proposed for prediction of autism spectrum disorder. To predict autism spectrum disorder in an individual in terms of accuracy, precision, recall and f1-score.

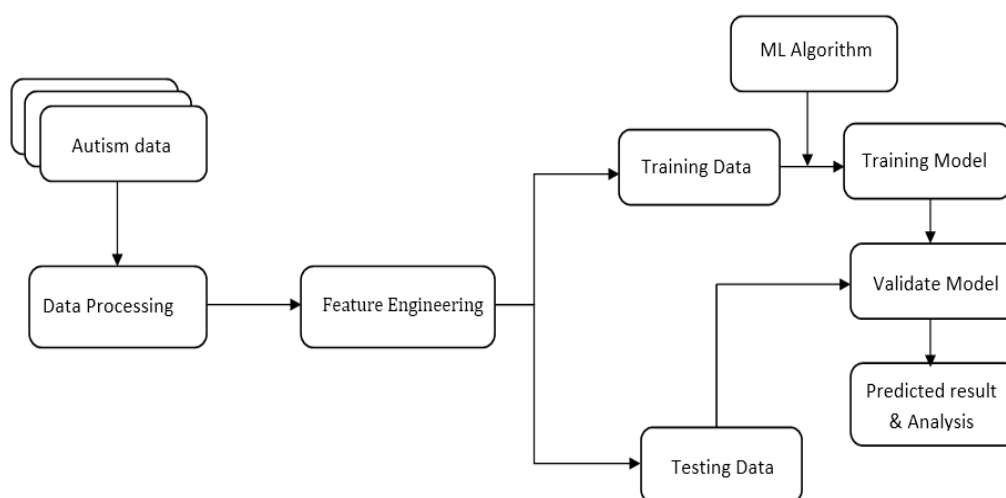


Figure 1: system architecture

Datasets are processed in data processing by extracting unwanted data and labelling the data so that the computer can decipher it easily. Feature extraction is performed using machine learning techniques to extract the best feature from the raw data. These features will help the proposed model perform better. The datasets are split into two parts: 80 percent training data and 20% testing data. The proposed model is created by feeding training datasets to machine learning algorithms. The testing datasets are also used to validate the model by predicting and analysing the outcomes based on accuracy, precision, recall and f1-score. The user is shown the final result.

IV. 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, used in the study, are detailed below.

4.1 Data collection

The data consists 16 attributes and 1054 instances. The Autism Spectrum Quotient, or AQ-10, is a tool that is used to determine if a person should be referred for a full autism evaluation. Sixteen numerical and categorical attributes are included in the datasets as shown in the table 1, including age, gender, and whether or not they were born with jaundice, as well as a family member with PDD, who is taking the test, Question 1-10 and class.

4.2 Data Pre-processing

By removing unnecessary data and labelling the necessary data so that the computer can easily decipher it. Machine learning techniques are used to extract the best feature from raw data during feature extraction. These characteristics will aid the proposed model's performance. The datasets are divided into two parts: 80% training data and 20% testing data. Machine learning algorithms are fed training datasets to create the proposed model.

The testing datasets are also used to validate the model by predicting and analysing the results using accuracy, precision, recall, and the f1-score. The final result is displayed to the user. Comparison of various types machine learning techniques are described below Figure 9.

4.3 Implementing a Model for Prediction

Details of various algorithms or strategies that are suitable while performing the procedure for the provided disorder dataset are mentioned in the Implementation section. The technique handles the problem by relying on a revolutionary machine learning process for autism spectrum disorder (ASD) classification and prediction. Ensemble learning algorithms combine numerous machine learning algorithms into a unified predictive model to diminish volatility, bias, and improve predictions. The model's performance and accuracy will be improved by using Random Forest (RF), Support Vector Machine (SVM), and AdaBooster techniques.

4.3.1 Random forests:

Random forests, also known as random choice forests, are an ensemble classifier, prediction, and perhaps other problems which works by developing a large random forest and then generating a category which is the mean forecast of the individual trees.

Step 1 – Begin by selecting random samples out of a training sample.

Step 2 – Thereafter, within each instance, the technique will generate a decision tree.

Step 3 – In this step, voting will take place in this phase to every expected outcome.

4.3.2 AdaBoost:

One of several boosting algorithms being used in solving activities being AdaBoost. AdaBooster aims to combine numerous weak classifiers into a centralized control class. Decision trees with a single split, also recognised as decision stumps, are the weak learners in AdaBoost.

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.

4.3.3 Support Vector Machines:

Support Vector Machines (SVMs), described to as support vector platforms, being supervised learning models which really evaluate information for classification and regression evaluation utilizing related learning methods. A splitting hyper - plane specifies it. In several other words, the model achieves an ideal hyperplane that identifies fresh samples offered labelled data (supervised learning). Categorizing a batch of fresh transcripts into positive or negative emotion clusters based on existing papers that have previously become categorized as positive or negative is an instance of a class. A Support Vector Machine (SVM) simulates the circumstance by generating a feature vectors, which is a symmetric plane curve with each dimension indicating an attribute of a certain item. Each feature in the sense of spam or document classification is the frequency or value of a specific term. The data points closest to the hyperplane, or the points of a data set that, if removed, will change the direction of the dividing hyperplane, are called support vectors.

Hyperplane: A hyperplane is a system that separates and identifies a number of observations in a linear manner. **Margin:** The margin is the separation between both the hyperplane and the adjacent piece of data by either collection. The categories are not differentiated in H1. H2 has a minor impact, but only by a slight margin. H3 separates them by the broadest margin.

4.4 Predicting and validating Model

The proposed model will be evaluated in terms of accuracy, precision, recall, and f1-score using the AQ-10 dataset and a collection of real-world datasets.

Table 1: Description of the Dataset

SI No	Attribute Name	Description	Values possible
1	A1 to A10	Corresponding Questions to patients	No-0, Yes-1
2	Age_Mons	Age in months	Value between 12 to 36
3	Sex	Patient's gender	Male-m, Female-f
4	Jaundice	Whether the patient was born with jaundice	No-0, Yes-1
5	Family with ASD	Whether any immediate family member has ASD	No-0, Yes-1
6	Who completed the test	Person who completed the test	'Family member', 'Health care professional', 'Self', 'Others'
7	Class	ASD traits or non-ASD traits	No-0, Yes-1

V. EXPERIMENTAL RESULTS

Evaluating the power of model prediction on a fresh instance after model development is a critical topic. It is indeed natural to wonder how a testable theory developed on existing records will fare on data it hasn't seen before. One may even try several supermodels for just that target variable and instead compare their prediction evidence to determine which model to use in a legitimate decision-making situation. Accuracy, recall, and other performance indicators are widely used to assess a predictor's performance. A confusion distribution is a tool frequently used to define the class label on a validation dataset for which the real analysis was done. Although the confusion matrices are indeed simple, the associated terms could be confusing. Accuracy, recall, and other performance indicators are widely used to assess a predictor's performance. The most often used performance measures will be described Early-stage prediction of autism spectrum disorder by machine learning approaches first, followed by an explanation and comparison of certain well-known estimating approaches.

Precision: Precision refers to a classifier's ability to avoid labelling a negative instance as positive. It is calculated as the ratio of true positives to the sum of true positives and false positives for each class.

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{-----} \quad 1$$

Recall: The capacity of a classifier to identify all positive cases is known as recall. It is calculated as the ratio of true positives to the sum of true positives and false negatives for each class.

$$\text{Recall} = \frac{TP}{TP + FN} \quad \text{-----} \quad 2$$

Accuracy: The proportion of accurately identified entities points based on the number of points is the simple ratio of accuracy.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{-----} \quad 3$$

Confusion matrix: A confusion distribution is a tool frequently used to define the class label on a validation dataset for which the real analysis was done. Although the confusion matrices are indeed simple, the associated terms could be confusing.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figure 2: Confusion matrix

5.1 Random forest Model

Overall accuracy of RF model using test-set is : 96.68%

	precision	recall	f1-score	support
0	0.99	0.92	0.95	78
1	0.96	0.99	0.97	133
accuracy			0.97	211
macro avg	0.97	0.96	0.96	211
weighted avg	0.97	0.97	0.97	211

Figure 3: classificationreport of Random Forest model

```
[[ 72  6]
 [  1 132]]
```

Figure 4: Confusion matrix of Random Forest model

5.2 AdaBoost Model

Overall accuracy of ADA 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

Figure 5: classification report of AdaBoost Model

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

Figure 6: Confusion matrix of AdaBoost Model

5.3 Support Vector Machines Model

Overall accuracy of SVM model using test-set is : 79.62%

	precision	recall	f1-score	support
0	1.00	0.45	0.62	78
1	0.76	1.00	0.86	133
accuracy			0.80	211
macro avg	0.88	0.72	0.74	211
weighted avg	0.85	0.80	0.77	211

Figure 7: Classification report of Support Vector Machines Model

[[35 43]
[0 133]]

Figure 8: Confusion matrix of Support Vector Machines Model

Table 2: Performance Analysis for prediction model using machine learning

Algorithm	Precision	Recall	F1-Score	Accuracy
Random forest	96	99	97	96.68
AdaBoost	100	100	100	100
SVM	76	100	86	79.62

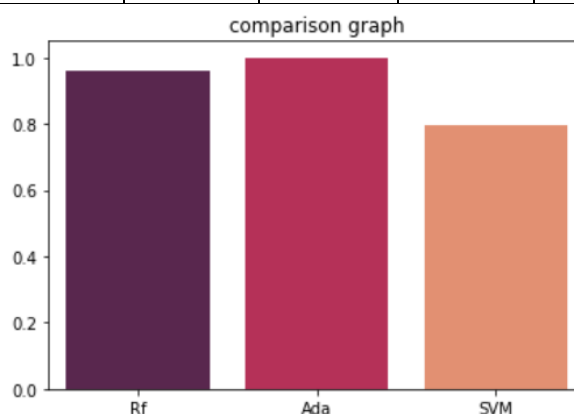


Figure 9: comparison graph of algorithms

VI. 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 1054 real the sample which the proposed approach can use identify autism with 95.73%, 100%, 79.62% accuracy 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] Bhawana Tyagi, Rahul Mishra and Neha Bajpai, "Machine Learning Techniques to Predict Autism Spectrum Disorder" School of Information Technology C-DAC Noida, India, [2018 IEEE Punecon](#).
- [2] Demirhan, A., (2018)., "Performance of machine learning methods in determining the autism spectrum disorder cases", Mugla Journal of Science and Technology, 4(1), 79-84.
- [3] Osman Altay and Mustafa Ulas., "Prediction of the Autism Spectrum Disorder Diagnosis with Linear Discriminant Analysis Classifier and K-Nearest Neighbour." 2018 in 6th International Symposium on Digital Forensic and Security (ISDFS).
- [4] Dadang Eman and Andi W.R Emanuel., "Machine Learning Classifiers for Autism Spectrum Disorder: A Review", Magister Teknik Informatika Universitas Atma Jaya Yogyakarta, Indonesia, 2019 4th ICITISEE, IEEE.
- [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.
- [6] Muhammad Nazrul Islam, Kazi Shahrukh Omar, Prodipta Mondal, Nabila Shahnaz Khan, "A Machine Learning Approach to Predict Autism Spectrum Disorder," International Conference on Electrical, Computer and Communication Engineering, Feb 2019.



- [7] 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), 2019, pp. 1-7, doi: 10.1109/ic-ETITE47903.2020.186.
- [8] 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.
- [9] 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, 2019.
- [10] 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.



**Impact Factor:
7.569**



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details