JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

PREDICTION OF AUTISM SPECTRUM DISORDER USING MACHINE LEARNING

[1]Murugan R, [2]Senbagamalar L

[1]School of Computer Science and IT, [2]Research Scholar
[1]Jain (Deemed to be University), Bangalore, INDIA
[2]Bannari Amman Institute of Technology, Sathyamangalam, INDIA

Abstract: In current situation the growth of Autism Spectrum Disorder (ASD) is increasing day by day. Previously to detect ASD screening test took place but they ended up expensive and time consuming. Thus, using AI and ML technologies we can predict the disorder. There are number of studies carried out with different algorithms, dataset and other categories. But they haven't given effective result. Therefore, we have used three different algorithms i.e., Random Forest, SVM and Adaboost. They are compared between each other and the best of the three is been implemented and taken into consideration for an effective result.

Index Terms – Autism Spectrum Disorder, Machine Learning, Adaboost, Random Forest, Support Vector Machine

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurological and developmental disorder identified by a set of traits that affects person's social skills, their communication with others, sensory problems along with few repetitive disorders. It is also known as Asperger Syndrome. Almost 15 million people in a country have Autism. ASD is caused due to genetics or environmental factors, but there is no pathological proof to show how ASD is caused. Major symptoms are lacking in interaction, no proper eye contact and repetitive behavior. All these symptoms usually start before age of one and it can be detected in any age. But sooner it is treated the better. Because later it is detected it would affect the child life (i.e., a child can't interact with others thus ending up being bullied in his/her school and college life). Autistic children do not recognize activities, people or any object present in their environment. Thus, we focus to detect ASD with an age period of 12 to 36 months so that later the child future can be as normal as possible. In current time the explosion rate / growth of ASD is increasing abruptly. In earlier period, ASD diagnosis required effective/significant amount of time and cost. Now hence fast screening test tool is used which predicts ASD of an individual child. With the help of machine learning we will try to detect autism at a quite early stage. Though different studies are done on different aspects like only based on voice or considering only facial expressions or behaviors separately, but they all failed to predict autism at a quite early stage. We will develop a system which can detect autism before age of 3, with all possible symptoms and by allowing machine to learn more symptoms and predict whether the baby is autistic or not. This paper discusses about previous contribution done towards ASD, Methodology, Training phase and testing, Algorithms (Implementation of models), Conclusion and future enhancement, References.

II. LITERATURE SURVEY

In this section we discuss about previous contribution done towards ASD. Each paper has used different technique, algorithms, dataset and predicted ASD. [1] Benjamin Gesundheit and Joshua Rosenzweig have included the analysis of the current animal models for ASD and their suitability, reviewing, behavioral, immunogenic, and epigenetic research, reassessing clinical diagnostic tools. They have taken 12 adults diagnosed with ASD and age matched controls performing a visual target detection task. [2] Arodami Chorianopoulou, Efthymios Tzinis, Elias Losif, Asimenia Papoulidi, Christina Papoulidi and Alexandros Potamianos have investigated the degree of engagement of children in interactions with their parents. Features derived from both participants including acoustic, linguistic and dialogue act features are explored. They have considered the datasets of Video recordings and data labeling. They have experimented on the task of engagement detection using video-recorded sessions consisting of interactions of typically developing (TD) and ASD children. [3] Siriwan Sunsirikul and Tiranee Achalakul presents a technique to investigate the behavior factor associations, and to classify these relations using classification based on association (CBA). Their experiments used actual patient profiles from two hospitals in Thailand. This dataset was categorized by doctors in two types: Autism and Pervasive Developmental Disorder- Not otherwise Specified (PDD-NOS). In this paper an effective classification mining called associative classification (AC) has been proposed. The approach emphasized relations of attributes, which differed from traditional classification methods. [4] Beibin li, Sachin Mehta, Deepali Aneja and Claive Foster have introduced an end-to-end machine learning based system for classifying autism spectrum disorder (ASD) using facial attributes such as expressions, action units, arousal and valence. They trained CNN-based model that takes a facial image as input and outputs four facial attributes to be used for ASD prediction. They have taken brief review the existing work for facial attribute recognition and their application in autism. [5] Pratibha Vellanki used unsupervised learning methods in this task. [6] Paul Fergus used cartoon characters in the mobile application he developed to help children with autism. [7] V. Y Tittagalla used eye contact, responsiveness to stimulus, analysis of vocal behavioural patterns and questionnaire to predict autism. . [8] Daiki Mitsumoto used the features of speech to identify autism. [9] Tarannum Zaki used sensing keypad to provide an easy and flexible means of interaction for autistic kids. [10] Ardiana Sula proposed a system using JXTA-Overlay Platform based on peer-to-peer communication between children, parents or caretakers and therapists and they used Smartbox along with sensor for monitoring and controlling child's activities. But using sensors is expensive is the major drawback. [11] Haibin Cai, Yinfeng Fang, Zhaojie Ju and Honghai Liu have proposed a sensing system that automatically extracts and fuses sensory features such as body motion features, facial expressions, and gaze features, further assessing the children's behaviors by mapping them to therapist-specified behavioral classes. This paper made an attempt to improve the existing systems of both standard and robot assisted therapy for children with ASD via a sensing framework with multi-sensory configuration and fusion. [13] Sushama Rani Dutta and Sujoy Datta worked on detecting preliminary symptom using cogency and machine learning where they used CARMRMR algorithm for predicting next possible symptoms by training with old symptoms which failed to predict efficiently. [14] Che Zawiyah Che Hasan developed a system for identifying autism spectrum disorder using ANN and SVM classifiers based on Three-Dimensional ground reaction forces, here individual persons demonstrated various evidence of movement and gait imbalance and alteration in joint kinetics with clumsiness were observed to predict autism which failed to predict ASD in all cases. [15] Dally and his team used 15 algorithms to analyze autism genetic resource exchange. In ADI-R 93 questions were asked. Hence, it took a greater number of times to examine the disorder. [16] Bone used ADI-R and SRS algorithms. The limitation was it contained a wide age range dataset (4-55 years). [17] Kosmicki used logistic regression and SVM algorithms. The limitation was it had a larger dataset and used ADI-R technique. [18] LiYi and his team identified ASD using face detection. They used K-means algorithm. The limitation was that it was applicable for Chinese faces only. Eye tracking data was also not stable. [19] Omar and his team used decision tree and random forest technique. The limitation was that it was not applied at early age. The data set was large in number.

III. METHODOLOGY

The Research methodology mainly has two phases Training and Testing Phase.

1. TRAINING PHASES

The Training phase was carried out in six phases: Data collection, Data processing, Data visualization, Build the model, Training the model and test the model.

A. DATA COLLECTION

Data collection is the process where information gathered and measured on variables of interest, it establishes a systematic fashion where one answers to a stated research question, test the hypothesis and evaluates the outcomes. The data sets are collected to build an effective predictive model. This data set contains data of age group of 1 month to 36 months. The data set contains a set of questions which is used to identify whether the child is to be referred to autism assessment. We collected data sets which contains more than 1000 records. Domains focus on communication, age, behaviour, etc. So, we can train our model with different situations and experiences. Each question can be marked with 0 or 1.

B. DATA PROCESSING

Data processing are a series of actions which are performed on data to verify, organize transforms, integrate and extract data in an appropriate output form. Methods of processing must be rigorously documented to ensure the integrity and utility of the data. The data set that we have collected contains text format in few columns, so we first have to process this text to numerical format. We process the data because we need all the columns of the data set to contain a similar value so we can evaluate the dataset more efficiently. We also make use of null analysis to check if the data set consists of any null values or not. If the dataset contains any null value, it can be removed using the null analysis.

C. DATA VISUALIZATION

Data visualization is an easy way to represent more complex data in the form of graphics. We plot the graphs based on the dataset present to get a clear idea about which group is getting affected by autism. This helps to analyse the data collected. It is used to show the relationship among datasets. In our project we make use of three graphs plotting based on,

- Output versus age of month
- Output versus gender
- Output versus genetics

Data is split into prime and text format. The first graph plotting that is the output versus age of month tells us at which age the child exhibits more symptoms according to the dataset present so that we can take precautionary measure at that particular age. The second graph plotting that is the output versus gender tells us which gender group gets more affected by autism. The third plotting output versus genetics tells us whether the group which had autism in their family member is affected more or the group that did not have any autism cases in the past.

D. BUILD THE MODEL

After collecting all the necessary details about the model, we are interested in designing we start the process of building the model. Building the model has a few stages in which it is carried out. By building the model it makes it easier to communicate about it with the people and make them understand about the working of our predictive model. Our model is designed to predict if a child has autism or not with few features. The model accepts inputs in the binary format, which gives a clear split between child with ASD or not. Where 1 represents positive and 0 represents negative. Which means if it's a 1 than the child has no autism if its 0 than the child has autism. To develop the prediction of autism, the algorithms were built and their accuracy was tested.

E. *TRAINING THE MODEL*

Training the model is a important phase in ML. the result we obtain from the model depends on how well we train our model. The performance increases with more than 1000 records. So our model is well trained with all the possible cases. As we have more no of patients record the model is trained well with all the data possible. We make use of 70% of the dataset to train the model.

F. *TESTING THE MODEL*

After training the model with the data set, we then can test the model. We select few set of data and feed the input to the model and check if the model is working well. As we use 70% of the dataset to train we made use of the rest 30% to test the model. Where we can easily get to know if our built model is trained well or not as we already have the prediction to check the output obtained.

2. TESTING PHASES

It consists of loading the trained example model, get new patient record, feeding the record in trained model and display the result. A patient's record is been fed and stored and training the data takes place. After training the model with the data set, we then can test the model. We select few sets of data and feed the input to the model and check if the model is working well. As we use 70% of the dataset to train, we made use of the rest 30% to test the model. Where we can easily get to know if our built model is trained well or not as we already have the prediction to check the output obtained.

IV. IMPLEMENTING THE PREDICTION MODEL

We are using three algorithms for developing prediction model. Initially random forest is considered for classifying the dataset we have for predicting autism spectrum disorder, the wrong outputs got from RF we will classify again in SVM with full data set , whatever missed in RD will be covered her with best result . But still few wrong outputs will be predicted and if any wrong input is given output will not be accurate so we are combining these two algorithms with Adaboost which is boosting algorithm it will classify wrong output from previous algorithm and classify it correctly. So, we will be able to predict accurately.

A. USING RANDOM FOREST FOR PREDICTION MODEL

Random forest is a learning method used for classification, regression and other tasks. It works by constructing a collective of decision trees at training time and outputting the class that is the mode of the classes or mean prediction of individual trees. The basic quality of this algorithm is that to construct a small decision tree with few parameters. We can construct many small, weak decision trees parallel and then combine all that to form a single strong learner by considering average result from dataset or majority of result.

For gaining more accuracy and to work against over-fitting random forest is used, here we have split into two parts i.e., generating random forest [2-10] and classifying dataset [11-16].

The algorithm works as follows:

For each tree in the forest, we select a bootstrap sample S

(i) Where S denotes the ith bootstrap, we then learn a decision-tree using a modified decision-tree learning algorithm. The algorithm is modified as follows:

(ii) At each node of the tree, instead of examining all possible feature-splits, we randomly select some subsets of the features f $\underline{C}$ F, where F is the set of features.

(iii) Deciding on which feature to split is often computationally expensive aspect of decision tree learning.

(iv) By narrowing the set of features, we drastically speed the learning of the tree.

ALGORITHM 1: RANDOM FOREST

```
Precondition: A training set $S=(x_1,y_2),………., (x_n,y_n)$, features F and number of trees in forest B,
Function RANDOMFOREST(S,F)
        $H \leftarrow 0$
        for i in 1,……..,B do
            $S(i) \leftarrow$ A bootstrap sample from S
              $h_i \leftarrow$ RANDOMIZEDTREELEARN $(S^{(i)},F)$
                  $H \leftarrow HU\{h_i\}$
  end for
        Return H
end function
function RANDOMIZEDTREELEARN(S,F)
At each node:
        $f \leftarrow$ very small subset of F
          split on best failure in f
return The learned tree end function
```

A. USING SUPPORT VECTOR MACHINE FOR PREDICTION MODEL

SVM is a supervised learning method that looks at data and sorts it into one of two categories. It is a linear model for classification and regression. SVM algorithm creates a line or a hyper plane which separates the data into classes. We plot each data as point in n-dimensional space with particular co-ordinate. Initially it will identify the right hyper-plane and that hyper plane should have high margin then it will give correct classification. Then we classify two classes here we have considered two classes i.e., who has ASD and the other one who don't have ASD.

ALGORITHM 2 SUPPORT VECTOR MACHINE

```
CandidateSV = { closest pair from opposite classes}
    While there are violating points do
            Find a violator
            candidateSV = $U_{candidateSV}$
            If any $a_p < 0$ due to addition of C to S then
                candidateSV = candidateSV/p
                Repeat till all such points are pruned
            End if
    End while
```

USING ADABOOST FOR PREDICTION MODEL

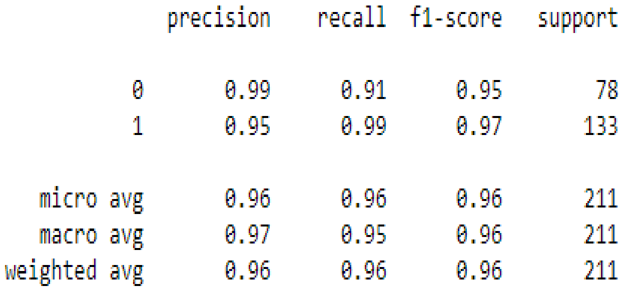
Adaboost is one of boosting algorithms; it will help in combining multiple weak classifiers into a one strong classifier. This algorithm will give us the best output because the wrong predicted outputs from random forest and SVM are combined into a weighted sum that represents the final outcome of the boosted classifier. The individual learners can be weak, but when combined with this the final model can be proven to converge to a strong learner. Adaboost is sensitive to noisy data and outliers. It gives high degree of precision.

ALGORITHM 3: ADABOOST

```
Given : $(x_1,y_1),……….,(x_m,y_m)$, $x_i \leftarrow X$, $y_i$ □ $Y = \{-1,1\}$.
    Initialize $D_i(i) = 1/m$
    For t = 1……..T:
        Train weak classifier using distribution $D_t$
        Get weak hypothesis $h_t$: $x \rightarrow \{-1,1\}$ with error $□_t = \sum_{i:ht(xi)!=yi} D_i(x_i)$
        Choose $□_t = 1/2 \log (1-□_t/□_t)$
        Update :
            $D_{t+1}(i) = D_t(i) / z_t$ =□$e^{-□□□}$if instance i is correctly classified or $e^{□t}$ where $z_t$ is normalization factor
    Output the final hypothesis: $H(x) = \text{sign} (\sum^{T}_{t=1} □_t h_t(x)$
```

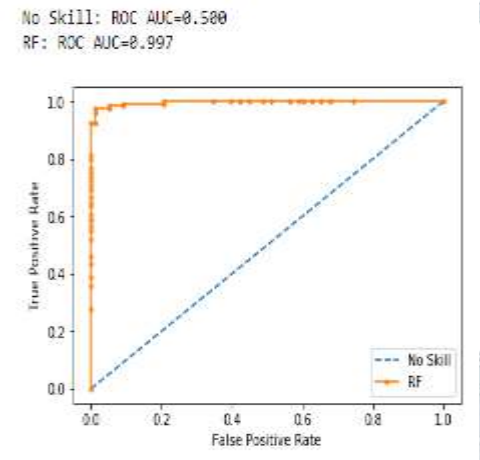
V. EVALUATION AN PREDICTION MODEL

RANDOM FOREST Classification



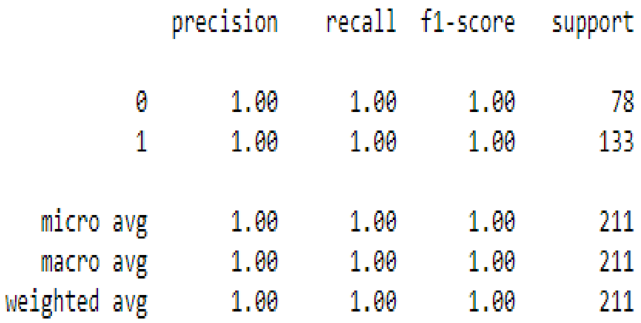
```
              precision    recall  f1-score   support

           0       0.99      0.91      0.95        78
           1       0.95      0.99      0.97       133

   micro avg       0.96      0.96      0.96       211
   macro avg       0.97      0.95      0.96       211
weighted avg       0.96      0.96      0.96       211
```

FPR and TPR Graph



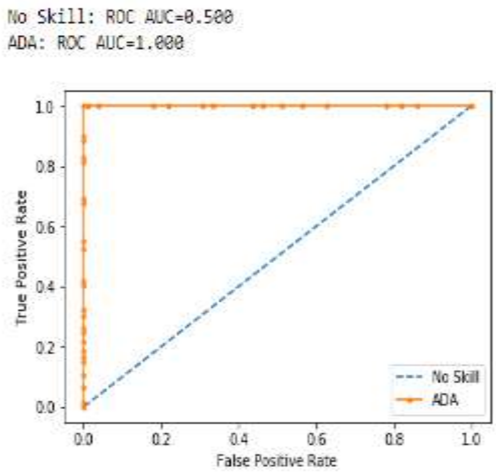
```
No Skill: ROC AUC=0.500
RF: ROC AUC=0.997
```

ADABOOST Classification



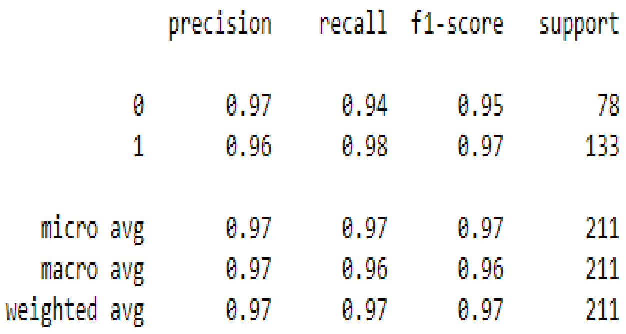
```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        78
           1       1.00      1.00      1.00       133

   micro avg       1.00      1.00      1.00       211
   macro avg       1.00      1.00      1.00       211
weighted avg       1.00      1.00      1.00       211
```

FPR and TPR Graph



```
No Skill: ROC AUC=0.500
ADA: ROC AUC=1.000
```

- **SVM Classification**



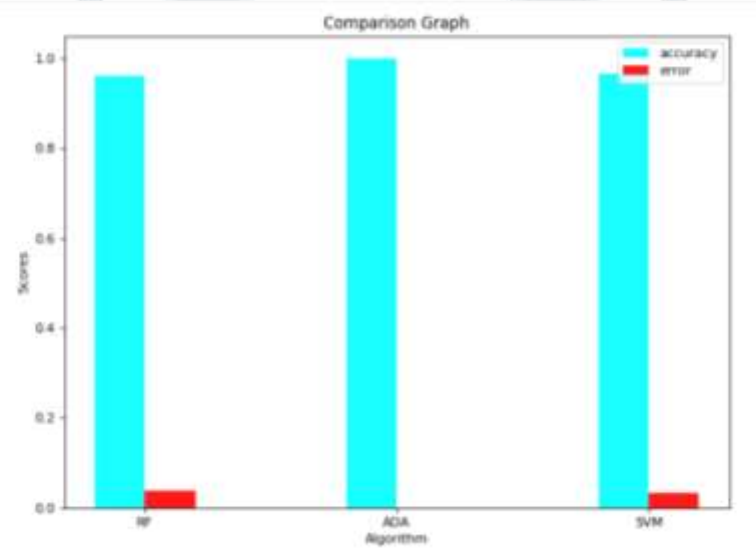
```
              precision    recall  f1-score   support

           0       0.97      0.94      0.95        78
           1       0.96      0.98      0.97       133

   micro avg       0.97      0.97      0.97       211
   macro avg       0.97      0.96      0.96       211
weighted avg       0.97      0.97      0.97       211
```

VI. RESULTS

A result is the final stage consequence of actions which is expressed qualitatively or quantitatively. Performance analysis is an operational analysis, is a set of basic quantitative relationship between the performance quantities. In our model we have used three different algorithms to classify whether child has ASD or not, the classifying property of adaboost, random forest and support vector machine algorithms has helped predicting desired output. We have predicted correctly for whatever dataset we have used to prepare this model.

Parameters	Random Forest	SVM	AdaBoost
Accuracy	96%	96%	100%
Loss	4%	4%	0%
Precision	0.96	0.97	0.1
Recall	0.96	0.97	0.1
f1-score	0.96	0.97	0.1

Data analytics table comparison between the three algorithms.



Accuracy comparison graph between the three algorithms.

VII. CONCLUSIONS

As identified through the literature review, we came to a conclusion that only a marginal success is achieved in the creation of predictive model for ASD patients. We mainly focussed on early ages of child so, that will be easy to make them cure. We investigated the parents by asking some questions, based on the parents reply we have made a classification to predict whether the child have autism spectrum disorder or not. Instead of using medicines child must be cured by regular counselling or by natural home remedies. The effects of ASD are often disastrous, thus families and schools have to adapt to provide the best for people with ASD to attain their potential. In real world example this can be implemented in many orphanages having the young children. Whether there are any changes in the growth of a child or the behaviour of the child is different when compared to other children then our approach is very useful. The approach we used in our experiment is more effective to classify different attributes. Our result will show the better performance comparing to other existing approach of screening autism. In future work, more features will be investigated and alternative machine learning algorithms will be evaluated for prediction.

REFERENCES

[1] Benjamin Gesundheit* and Joshua P. Rosenzweig, "Editorial: Autism Spectrum Disorders (ASD)-Searching for the Biological Basis for Behavioral Symptoms and New Therapeutic Targets, Published online 2017 Jan.

[2] Arodami Chorianopoulou, Efthymios Tzinis, Elias Iosif Asimenia Papoulidi, Christina Papailiou, Alexandros Potamianos, "Engagement detection for children with autism spectrum disorder", 2017.

[3] Siriwan Sunsirikul and Tiranee Achalakul, "Associative Classification Mining in the Behavior Study of Autism Spectrum Disorder", vol.3, 2010.

[4] Beibin Li ; Sachin Mehta ; Deepali Aneja ; ClaireFoster ; PamelaVentola ; Frederick Shic ; Linda Shapiro, "A Facial Affect Analysis System for Autism Spectrum Disorder", 2019.

[5] Pratibha Vellanki, Thi Duong, Svetha Venkatesh, Dinh Phung, "Nonparametric Discovery of Learning Patterns and Autism Subgroups from Therapeutic Data", 2014.

[6] Paul Fergus, Basma Abdulaimma, Chris Carter, Sheena Round, "Interactive Mobile Technology for Children with Autism Spectrum Condition (ASC)", 2011.

[7] V.Y Tittagalla, R. R. P Wickramarachchi, G. W. C. N. Chandrarathne, N.M. D. M. B. Nanayakkara, P. Samarasinghe, P. Rathnayake and M.G.N.M. Pemadasa, "Screening Tool for Autistic Children", 2019.

[8] Daiki Mitsumoto, Takeshi Hori, Shigeki Sagayama Hidenori Yamasue, Keiho Owada, Masaki Kojima, Keiko Ochi, Nobutaka Ono, "Autism Spectrum Disorder Discrimination Based on Voice Activities Related to Fillers and Laughter", 2019.

[9] Tarannum Zaki, Muhammad Nazrul Islam, Md. Sami Uddin, Sanjida Nasreen Tumpa, Md. Jubair Hossain, Maksuda Rahman Anti, Md. Mahedi Hasan, "Towards Developing a Learning Tool for Children with Autism", 2017.

[10] Ardiana Sula, Evjola Spaho, Keita Matsuo, Leonard Barolli, Rozeta Miho and Fatos Xhafa, "An IoT-based System for Supporting Children with Autism Spectrum Disorder", 2013.

[11] Haibin Cai, Yinfeng Fang, Zhaojie Ju, Cristina Costescu, Daniel David, Erik Billing, Tom Ziemke, Serge Thill, Tony Belpaeme, Bram Vanderborght, David Vernon, Kathleen Richardson and Honghai Liu, "Sensing-enhanced Therapy System for Assessing Children with Autism Spectrum Disorders: A Feasibility Study", 2018.

[12] Akshay Vijayan ; S Janmasree ; C Keerthana ; L Baby Syla, "A Framework for Intelligent Learning Assistant Platform Based on Cognitive Computing for Children with Autism Spectrum Disorder", July 2018.

[13] Sushama Rani Dutta ; Sujoy Datta ; Monideepa Roy, "Using Cogency and Machine Learning for Autism Detection from a Preliminary Symptom", July 2019.

[14] Che Zawiyah Che Hasan, Rozita Jailani and Nooritawati Md Tahir, "ANN and SVM Classifiers in Identifying Autism Spectrum Disorder Gait Based on Three-Dimensional Ground Reaction Forces", October 2018.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] Kazi Shahrukh Omar, Prodipta Mondal, Nabila Shahnaz Khan, “A Machine Learning Approach to Predict Autism Spectrum Disorder”, 7-9 February, 2019.