

# An Effective Model Of Autism Spectrum Disorder Using Machine Learning

Razieh Asgarnezhad<sup>1</sup>, Karrar Ali Mohsin Alhameedawi<sup>2</sup>, Hani Akram Mahfoud<sup>3</sup>

<sup>1</sup>Department of Computer Engineering, Aghigh Institute of Higher Education Shahinshahr, Iran

<sup>2</sup>Department of Computer Engineering, Iran University of Science and Technology, Tehran

<sup>3</sup>Department of Computer Engineering, University of Isfahan, Iran

---

## Article Info

### Article history:

Received Aug 6, 2022

Revised Mar 8, 2023

Accepted Mar 24, 2023

---

### Keyword:

Data Mining

Pre-processing

Autism spectrum disorder

Ensembles

Machine learning techniques

---

## ABSTRACT

Autism spectrum disorder (ASD) is one of the most common diseases that affect human nerves and cause a decrease in the intelligence and comprehension of the person. This disease is a group of various disorders that are characterized by poor social behavior and communication. It affects all age groups, including adults, adolescents, children, and the elderly, but the symptoms of this disease always appear in their early years. ASD suffer from problems, the most important of which are data loss, low quality, and extreme values. This makes the process of diagnosing the ASD early. Our goals in this research is to solve the ASD problems. The cussent authors proposed a technical model that solves all data problems. We used ensemble techniques that include Bayesian Boosting, Classification by Regression, Polynomial by Binominal Classification. We also used classification techniques that include CHAID, Decision Stump, Decision Tree (Weight-Based), Gradient Boosted Trees, ID3. It is proven that the proposed model solves data problems, and has obtained the highest search accuracy that has reached 100% as well as we have obtained the highest f1 measurement that has reached 100%. This proves that our work is superior to its peers.

*Copyright © 2023 Institute of Advanced Engineering and Science.  
All rights reserved.*

---

## Corresponding Author:

Razieh Asgarnezhad

Department of Computer Engineering,

Aghigh Institute of Higher Education Shahinshahr,

Isfahan, Iran

Email: r.asgarnezhad@aghigh.ac.ir

---

## 1. INTRODUCTION

Autism spectrum disorder (ASD) suffers from many problems, and this is what makes the diagnosis of this disease late. ASD is a condition related to human thinking and mental development, as it causes psychological conditions and obsessive-compulsive disorder in people with this disease, affecting all age groups [1] [2]. ASD over time, some sufferers may enjoy an independent life far from problems and some disabilities that hinder their work and practice their daily lives without problems. But others suffer from ASD and need health care and someone to guide and help them in their requirements with their matters ASD on the field of work and jobs. Once they know that he has ASD, they do not accept to work with them, and this causes many problems for people with this disease. The most important problems of ASD are the problems of data and records related to this disease [3] [4]. For example, missing values and outliers cause problems in diagnosing this disease. The biggest challenge in this research is to apply a technical model that solves repeated data problems and improves the performance of the chronic autism spectrum [5].

In 2017, the authors proposed a technical model based on supervised machine learning (ML) methods. They improved the ASD performance, reduced the incidence, and found the best solutions to address the problems facing this disease. They presented a study on determining knowledge of treatment progress among children with autism spectrum disorder based on the most important behavioral data. The authors also noted the most important things in patients who use video games, computer techniques, and computer science. Most of the patients do not have a high sense of distinguishing. They applied classification algorithms where they

got values that are considered good with tree algorithm. The decision tree reached the highest accuracy of 80%. The precision, recall, and F1 were 81.05%, 86.12%, and 83.51%, respectively. These values are considered satisfactory and convincing [6]. In 2016, the authors developed and evaluated a supervised ML approach to classify cases in ADDM using words and phrases from developmental awareness assessments for children. They trained and taught a random forest classifier using a dataset from the 2008 Georgia ADDM website. It included 1,162 children with 5,396 assessments. They used words and phrases to predict the best outcomes and values for improvement in the ASD [7]. They monitored their performance on the 2010 Georgia ADDM observational data. The authors found a satisfactory outcome of 86.5% by the specialist. With these injuries also reached 84.0% sensitivity, 89.4% positive predictive value [8].

In this study, we downloaded a dataset for autism from the UCI website. This data suffers from many problems, such as missing values and outliers [9] [10]. It makes the diagnosis of this disease be slow and not early. We proposed a technical model using supervised ML techniques [11] [12] [13] [14]. We used two experiments. In the first experiment, we applied Ensemble techniques include Bayesian Boosting, Classification by Regression, and Polynomial by Binominal Classification. With the application of preprocessing techniques such as replace the missing value with mean and detect outliers with K-nearest neighbor (KNN), we obtained the highest accuracy of with Bayesian Boosting, Polynomial by Binominal Classification, and Classification by Regression techniques. We also applied another experiment with supervised ML techniques. But in comparison with the first experiment, we used classification techniques in conjunction with five algorithms including CHAID, Decision Stump, Decision Tree (Weight-Based), Gradient Boosted Trees, and ID3. With the preprocessing techniques that include replacing missing value with mean and detect outliers with the KNN. The highest accuracy obtained in this experiment equal 100% with both Decision Stump, Gradient Boosted Trees. Through the results obtained, proves that the proposed model will address the problems of data to solve the problems of missing and extreme values. The model improves the performance of the ASD rather than its counterparts.

Our innovation in this study is a technical model that compares classification algorithms and ensemble algorithms to get good results on the ASD with many problems. Our proposal address data problems to improve diagnosing the ASD to predict the best results. Our innovation has proven in terms of the results obtained that it addresses data problems and outperforms previous works.

This article is organized as follows: Sect. 2, shows a summary of the related works followed by a description of data collection in Sect. 3. The proposed method and evaluation by the experiment explained in Sect. 4. Finally, the article presents a conclusion in Sect. 5.

## 2. RELATED WORK

Several works have been proposed to predict the ASD diagnoses in the years between 2015 to 2021 [15] [16] [17]. The current authors summarize some important works here. We used supervised ML techniques that include bayesian boosting, classification by regression, polynomial by binominal, classification categories include chaid, decision stump, decision tree (weight-based), gradient boosted trees, ID3. We also applied preprocessing techniques, after applying these techniques [18] [19]. We got the highest accuracy in our article reaching 100%, this proves that our work is superior to peers it improves and diagnoses ASD and solves data problems for the ASD [20].

In 2015, the authors presented a study to show a relationship between therapy power and time for supervision. They found age and gender to master training results specific to children's ABA systems on the spectrum. Their used methods are incorporated into the clinical. The preparation also indicates the influence of ML and Artificial Intelligent (AI) in this area [21]. They randomly divided their data as follows: 65% for training and 30% for testing, and 5% for validation. They used a backpropagation network with Bayesian adjustment [22].

One year later, in 2016, the authors conducted a deep study to analyze gene expression in the prefrontal cortex of 63 patients with the ASD to obtain an acceptable result and improve the ASD performance. Some methods give good results and accelerated the diagnosis of the ASD. In addition, there are 63 chimpanzees and monkeys from birth to adulthood and adolescence, where they appear among the many abnormal groups. The authors observed in the brains of the injured. Techniques used have satisfactory results for improving the diagnosis of the ASD. Several members of the family of early developmental response transcribers may be involved in regulating this human developmental change. As a result, the presented method is good for improving and resolving part of the ASD and its problems [23]. Other authors, in 2017, conducted a comprehensive lower extremity analysis study among children aged 5-12 years with autism and age- and gender-matched samples with TD. By conducting the model statistical procedure for the purpose of testing the statistical significance between similar pairs, the autistic children showed large numbers of the most important differences. They obtained satisfactory results [8]. Leroy, G., et al., in 2018, given the importance of the ASD

and required conducting several experiments to improve the ASD performance. The authors suggested designing and configuring a rule-based programming language performance that could be specified by DSM BED. The authors' goal was to extract a description of mood and behavior in the records. The results obtained by these authors showed good and satisfactory. They obtained an accuracy of 76%, a precision of 76%, and a recall of 42%, a F1 of 54.1%. These values were considered somewhat satisfactory since the data on the ASD suffer from many problems such as missing values, extreme values, and other problems that delayed the diagnosis of the ASD. But they were able to improve the ASD [24].

Raju, NG, et al., in 2019, proposed a new technical model for improving the ASD. They used supervised ML methods that include SVM technology. They also used reinforcement and decision tree to improve the ASD. These authors applied these techniques to the data on the ASD to obtain values for improving the performance of this disease. They obtained an accuracy of 95%, precision of 97%, recall of 95%, f1 of 95.98% [25].

In 2020, the authors proposed a technical model based on supervised ML techniques. They used ensemble techniques include AdaBoost [26]. ASD suffers from the many problems and this proposed model was used to solve them. Obtained results showed after applying these techniques, an accuracy of 97.1%, precision of 97%, recall of 98%, and f1 of 98% were reached. They found that this proposal improved the performance of the ASD and predicts the best results [27].

The current authors studied many works in this scope and seemed that the best result was obtained through the application of ML techniques, including classification techniques, ensemble techniques, and pre-processing techniques. The highest accuracy in the current article reached 100%. It proves that the proposed model outperforms peers. The obtained results also compared with related works.

### 3. THE METHODOLOGY

In this section, the current authors used two experiments with supervised ML methods [28] [29]. Using mean and detect outliers with KNN, the highest accuracy and f1 reached 100%. It proves that the first experiment is good for improving ASD and improving its performance. In the second experiment, other techniques of ML include CHAID, Decision Stump, Decision Tree (Weight-Based), Gradient Boosted Trees, and ID3 used. The highest accuracy in this experiment obtained 100% using Decision Stump and Gradient Boosted Trees.

It proved that the proposed model obtains good results and improves the performance of the ASD and solves its problems. Figure 1 shows the proposed methodology.

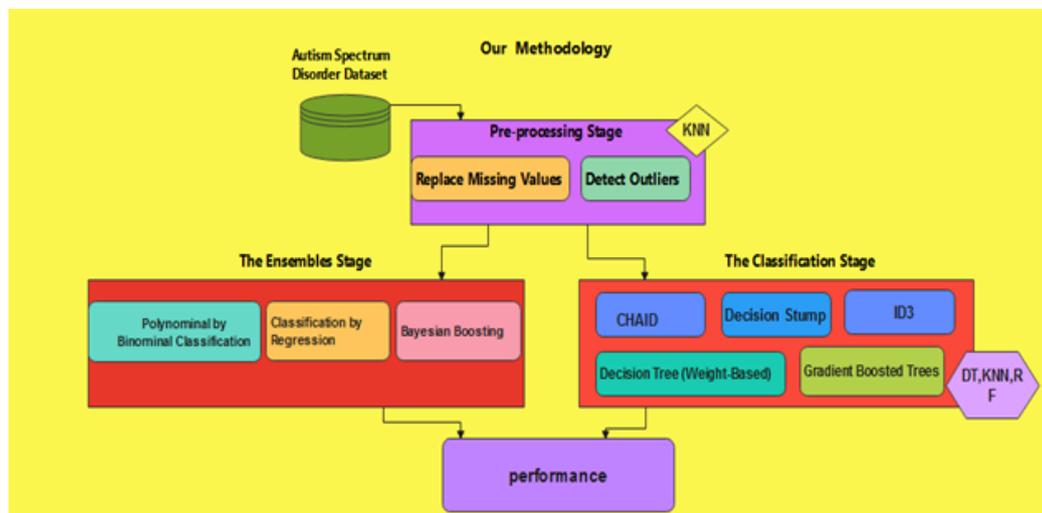


Figure 1. The proposed methodology

#### 3.1. Pre-processing Stage For Replace Missing Values And Detect Outliers

In this section, we explain the techniques that were used to handle missing values and outliers. After downloading data and records that contain low data quality, missing values, and outliers, we fed these data into data mining programming, and in the processing phase, we used transposition techniques. The missing values with the arithmetic mean and outliers with the nearest neighbors. Then, using ML algorithms such as ensemble and classification, the highest accuracy in the two experiments obtained.

To solve current data problems that suffer from missing values, the Mean measurement is used,

$$\text{Mean}(A_i) = \frac{A_1 + A_2 + \dots + A_i}{N}$$

where  $A_i$  is the values of the  $i^{\text{th}}$  column and  $N$  is the number of records in the dataset.

### 3.2. The Ensembles Stage

In this section, we explain ensemble techniques include Bayesian Boosting, Classification by Regression, Polynomial by Binominal Classification. We applied this technique with pre-processing techniques and obtained good enough values to predict the best results. We obtained the highest accuracy equal 100% in this study.

#### 3.2.1. Bayesian Boosting

The Bayesian Boosting operator is an embeded operator. The subprocess have a learner to expect an ExampleSet and generate a model. In this section, we explain an algorithm of supervised ML, Bayesian Boosting. The used dataset divided into two parts, the first part is the training part with 60% examples, while the other part is the testing part with 40% examples. This technique applied with the pre-processing techniques that included both the missing value with mean and detect outlier with KNN. After applying these techniques, the highest accuracy obtained equal 100%. This proves that this proposed model could improved the performance of the ASD.

#### 3.2.2. Classification by Regression

The Classification by Regression operator is a nested method. For the subprocess, a regression learner used that generates as supervised ML methods. It is one of the types of ensemble techniques and creates a classification model in the subprocesses. This technique is applied with pre-processing techniques include replacing the missing values according to the average rate. And also the detection outlier used with the nearest neighbor algorithm. An accuracy of 96.55% was obtained. It can consider good enough for improving the ASD and addressing missing and extreme values.

#### 3.2.3. Polynomial by Binominal Classification

Here we explain one of the supervised ML methods. It is one of the ensemble types and is considered good for improving medical conditions. The Polynomial by Binomial Classification operator is a nested operator. Subprocesses have a learner to perform a binomial classification. Pre-processing techniques were also applied with these algorithms to solve the problems of missing data, outliers and strays. We have obtained an accuracy of 100%. This confirms that the proposed model could improved the performance of the classification and outperforms its peers in terms of performance and results.

### 3.3. The Classification Stage

In this section, we explain classification techniques was one of the supervised ML methods. Data of the ASD suffers from many problems affect on the performance. The data has been entered into program and classification techniques include CHAID, Decision Stump, Decision have been applied. Tree (Weight-Based), Gradient Boosted Trees, and ID3 with pre-processing techniques applied to process missing and outliers. The highest accuracy was obtained at this stage. With classification techniques, the highest accuracy reached 100%. This proved that classification techniques improve the performance of the ASD and address data problems. The proposed model has proven superior to its peers.

#### 3.3.1. CHAID

Here we explain one of the supervised ML methods, CHAID. This operator is a decision tree operator with one exception that is similar to the currency of the decision tree. This technique was applied with pre-processing techniques that include replacing the missing values and the values nearest neighbors. After applying these techniques, we obtained an accuracy of 87.36% The highest accuracy is not the value, but it is considered good for improving the performance of the ASD and overcoming data problems. The advantages of CHAID are that its output is highly visible and easy to interpret. Because it uses omnidirectional partitions by default.

#### 3.3.2. Decision Stump

Another supervised ML technique is the Decision Stump operator to generate a decision tree with only one partitioning. It is one of the classification types. When you can boost it with operators like AdaBoost,

this operator can be effective. With the use of pre-processing techniques, an accuracy of 100% was obtained. It is considered the highest value in this article. It proves that our work is good and is superior to its peers.

### 3.3.3. Decision Tree (Weight-Based)

In this section, we explain another classification technique which is one of the supervised ML methods. The (weight-based) decision tree operator is a nested operator with a subprocess. This technique has been applied with pre-processing techniques to solve problems of the ASD. We have obtained an accuracy of 74.36%, and it is not the highest value in this research. But it is good for solving data problems Missing and extremes and improving the performance of the ASD diagnosis.

### 3.3.4. Gradient Boosted Trees

In this section, we applied a tree technique of classification types. The enhanced gradient model is a group Regression or classification tree models. Enhancement is an elastic non-linear regression procedure to improve the printing of trees. Next, we obtained the highest accuracy equal 100%. The obtained results outperformed all previous work in this scope.

### 3.3.5. ID3

ID3 is one of the types of supervised ML methods. This type of classification is an algorithm used in retailing a decision tree invented by Ross Quinlan. This algorithm takes all unused attributes and calculate their selection criterion e.g. information gain. It chooses the attribute for which the selection criterion has the best value e.g. minimum entropy or maximum information gain. It also makes the node contain that attribute. This algorithm is implemented with preprocessing techniques. An accuracy of 87.36% obtained. This value not the highest in this article, but it is good for improving the diagnosis of the ASD.

## 4. RESULTS AND ANALYSIS

### 4.1. Experiments

In the first experiment, we employed data on the ASD from the UCI website. This data suffers from many problems. For example, there are missing values, outliers and low quality data. So, we applied it to learn algorithms supervised automated. We used ensemble techniques that include Bayesian Boosting, Classification by Regression, Polynomial by Binominal Classification but without pre-processing, whereby the working area of the algorithms used was divided into 60% of the training section as well as 40% of the test section, where we obtained the highest accuracy and precision and recall and f1 with the ensemble algorithms, reached 100% with both Bayesian Boosting as well as Polynomial by Binominal Classification and these are the highest values in the study. The accuracy of the Classification by Regression algorithm reached 96.55%. It confirms that the proposed model and the algorithms used will improve the diagnosis and improve the performance of the ASD.

According to Table 1, the results with the ensemble algorithms that include Bayesian Boosting, Classification by Regression, Polynomial by Binominal Classification, showed that this model improves the performance of the autism spectrum and predicts the best results. The highest accuracy in this experiment reached 100% with both Bayesian Boosting algorithms as well as Polynomial by Binominal Classification, these results confirm that the proposed model improves the performance of the autism spectrum and outperforms its peers.

Table 1. The obtained best results through ensembles without pre-processing in conjunction with KNN and RF through (%)

Method	Precision	Recall	Accuracy	F1
Bayesian Boosting	100	100	100	100
Classification by Regression	96.66	96.66	96.55	97.25
Polynomial by Binominal Classification	100	100	100	100

In the second experiment, we employed the data that we downloaded for ASD, these data contain many problems, and to treat this data we applied pre-processing techniques include replace missing value with mean and detect outlier with KNN. By applying supervised ML techniques, we used ensemble techniques that include Bayesian Boosting, Classification by Regression, Polynomial by Binominal. We have obtained high values that predict the best results and improve the performance of the ASD and solve the problems of missing data and outliers. According to Table 2, after applying the Bayesian Boosting, Classification by Regression, Polynomial by Binominal techniques. We obtained the highest value of 100% with accuracy and precision and recall and f1. This confirms that the proposed model improves the performance of the ASD classification and

solves data problems, and we demonstrated a great performance that outperformed its peers and their previous experiences.

Table 2. The obtained best results through ensembles with pre-processing in conjunction with Decision Tree and Random Forest (%)

Method	Precision	Recall	Accuracy	F1
Bayesian Boosting	100	100	100	100
Classificationby Regression	100	100	100	100
Polynominal by Binominal Classification	100	100	100	100

In the third experiment, other techniques were used to obtain good results with other techniques, and even to be compared with me, the techniques that were used in the first experiment and the second. One of the methods of ML is classification algorithms including CHAID, Decision Stump, Decision Tree (Weight-Based), Gradient Boosted Trees, ID3 The data set was entered program, and then we applied the pre-processing techniques that include replacing the missing value with mean and detect outliers with KNN. These techniques were used with classification algorithms, in this experiment we obtained high values that predict the best results and improve the performance of the ASD. According to Table 3, we applied the classification techniques that include CHAID, Decision Stump, Decision Tree (Weight-Based), Gradient Boosted Trees, ID3, and we got the highest value in the article reaching 100% with accuracy and recall and precision and f1 with Decision Stump, Gradient Boosted Trees techniques. This confirms that the proposed model, using different techniques, adopts the best results, improves autism, and solves data problems.

Table 3 illustrates the best obtained performance of the proposed model in terms of precision, recall, accuracy, and f1. Table 4 shows the confusion matrix for an example. Table 5 shows Gains/Lift parameter with average depicts rate equal 48.29% and average score equal 48.29%. Table 6 illustrates Gains/Lift parameter with average response rate equal 48.29% and with average score equal 48.29% with Gradient Boosted model. Table 7 illustrates variable importances for records of the ASD.

Table 3. The best obtained results through classification with pre-processing in conjunction with Decision Tree and Random Forest (%)

Method	Precision	Recall	Accuracy	F1
CHAID	80.80	87.07	87.36	84.50
Decision Stump	100	100	100	100
Gradient Boosted Trees	100	100	100	100
Decision Tree (Weight-Based)	77.50	73.30	74.36	74.94
ID3	87.74	87.54	87.36	88.28

Table 4. CM: Confusion Matrix (Row labels: Actual class; Column labels: Predicted class)

	NO	YES	ERROR	RATE
NO	151	0	0.0000	0 / 151
YES	0	141	0.0000	0 / 141
Totals	151	141	0.0000	0 / 292

Table 5. Gains/Lift Table (Avg response rate: 48.29 %, avg score: 48.29 %)

Group	Cumulative Data Fraction	Lower Threshold	Lift	Cumulative Lift	Score
1	0.48287671	0.686903	2.070922	2.070922	0.686
2	1.00000000	0.292419	0.000000	1.000000	0.292

Table 6. Gains/Lift Table (Avg response rate: 48.29 %, avg score: 48.29 %) with Gradient Boosted model

Group	Cumulative Response Rate	Cumulative Score	Capture Rate	Cumulative Capture Rate	Gain
1	1.000000	0.686903	1.000000	1.000000	107.09
2	0.482877	0.482906	0.000000	1.000000	-100.00

Table 7. Variable importances for the ASD Records

Variable Relative	Importance Scaled	Importance	Percentage
Result	2323.939697	1.000000	1.000000
A2_Score	0.000000	0.000000	0.000000
A3_Score	0.000000	0.000000	0.000000
A4_Score	0.000000	0.000000	0.000000
A5_Score	0.000000	0.000000	0.000000
A6_Score	0.000000	0.000000	0.000000
A7_Score	0.000000	0.000000	0.000000
A8_Score	0.000000	0.000000	0.000000
A9_Score	0.000000	0.000000	0.000000
A10_Score	0.000000	0.000000	0.000000
Age	0.000000	0.000000	0.000000
Gender	0.000000	0.000000	0.000000
Austim	0.000000	0.000000	0.000000
Relation	0.000000	0.000000	0.000000

#### 4.2. Evaluation Metrics

We have applied evaluation parameters such as accuracy, precision, recall, and F1. We obtained very high values that outperformed their peers in terms of results, as the highest value with these criteria reached 100% with accuracy, precision, recall and f1 with ensemble and classification. These procedures were calculated and implemented through these formulas and the requirements in Table 8.

Table 8. Evaluation parameters

Parameter	Equation
Accuracy	$(TP + TN) / (P + N)$
Precision	$(TP) / (TP + FP)$
Recall	$TP / P$
F1	$\frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$

#### 4.3. Data Collection

In this article from among the many datasets for the ASD, we downloaded data and selected a children's autism set, which was downloaded from the UCI ML data repository. The dataset contains a set of 292 autism-related records as well as 21 traits. This data suffers from many problems such as data loss, outliers, and low-quality values. It makes the level of the ASD lower and makes the diagnosis not early. So, we applied ML techniques to this data with pre-processing techniques. The dataset contains all the numeric and nominal features [30]. These data are well known, but here are missing and extreme values, and we will solve these problems with our proposal.

#### 4.4. Discussion

In this article, the data mining tool was used to do data analysis and processing and solve data problems. We did three practical experiments. In the first experiment, we only applied ensemble techniques such as Bayesian, Boosting, Classification by Regression, Polynomial by Binominal Classification with Decision Tree and Random Forest, without applying pre-processing techniques. We got high values through reaching the highest accuracy of 100% with Bayesian Boosting and Polynomial by Binominal Classification techniques. In the second experiment, with pre-processing techniques such as Bayesian Boosting, Classification by Regression, Polynomial by Binominal Classification with Decision Tree and Random Forest, and with pre-processing such as replace the missing value with mean and detect outlier with KNN, we have

obtained high values. In the article, precision and recall and accuracy and f1 equal 100% with all ensemble techniques. In the third experiment, we used ensemble techniques such as Bayesian Boosting, Classification by Regression, Polynomial by Binominal Classification. With other techniques such as CHAID, Decision Stump, Decision Tree (Weight-Based), Gradient Boosted Trees, and ID3, we obtained high results, as the highest accuracy of 100%. This confirms that our work is good enough and our proposal improves the ASD performance rather than previous works. Four important evaluation parameters including recall, accuracy, F1, and accuracy were also used. Since precision equals the number of true positives over the total number Negative, true, and false positives, as the first table, shows the first experience with the application of supervised ML refinements. This means that the best results through ensembles without pre-processing in conjunction with KNN. And Random Forest, where the highest accuracy in Table 1 reached 100 %, as through the results of a table. We proved that our work is good and improves ASD. Table 2 also shows the application of ensemble techniques such as Bayesian. Boosting, Classification by Regression, Polynomial by Binominal Classification with Decision Tree and Random Forest with pre-processing techniques, where we obtained high values, the highest accuracy in the first table reached 100%. This proves that our model is good and outperforms its peers. Table 3 illustrates the third experiment where we used techniques from other Supervised ML methods namely CHAID, Decision Stump, Decision Tree (Weight-Based), Gradient Boosted Trees, ID3, with pre-processing techniques to improve ASD performance. We applied these techniques to get the best results. We will also explain in Table 7 how our work surpassed the previous work and gave more than wonderful results. Also, Tables 4, 5, 6, and 7 show the most important results that we obtained after applying the Gradient Boosted Model technology with pretreatment techniques to improve ASD performance. Table 4 shows the confusion matrix. Tables 5, 6, and 7 show the Gains/Lift and Demonstrates Variable Importances.

Table 9 shows a comprehensive comparison between the best obtained results and the previous works. The criteria in our work were precision, recall and accuracy and f1 obtained equal 100%. The value of the criteria in [24] were 76%, 42%, 76%, 54.1%. Also, another work developed to be compared to our work, where the values of the criteria in [25] were 97%. Accuracy, precision, recall, and f1 reached 97%, 95%, 95%, and 95.98% in this reference. Another previous work was also added to be compared to our work [27]. The criteria values for this reference were 97%, 98%, 97.1%, and 98%, respectively. Our work was also compared with another previous one where the criteria for accuracy, precision, recall, and f1 in reference [6] reached 80.5%, 86.12%, 80%, 83.51%, as through these results that appear in the seventh table, it is clear that our work outperforms all previous works and the proposed model is considered good for improving the ASD performance.

Figure 1 in the article shows the steps of the proposed model and the most important materials that we used to obtain improvement values from the ASD. Figures 2 to 13 illustrate the work of algorithms. Figure 2 and 3 illustrate the comparison between our current work with the two experiments, an experiment with classification techniques and an experiment with ensemble techniques. And comparing it with previous works, it was found that our work is superior in the two experiments, and in the two forms, our work is superior to its counterparts. Figure 4 and 5 show the best results that we obtained during the two experiments with classification techniques and ensemble techniques. The work shows that the values we obtained are high values and outperform their counterparts.

Note that Figures 6 to 13 show the ROC for the best classification and ensemble of the current authors Where the figures show the works that we obtained with and without pre-treatment and that the Messenger is the best evidence that confirms the superiority of our work in all respects. Thus, the proposed model has proven that it improves autism spectrum disorder, improves the diagnosis of the ASD, solves data problems, missing and extreme values, and solves problems of low data quality. This confirms that our work is superior to peers.

Table 9. A comparison among the obtained results through ensembles with pre-processing and other works (%)

Works	Precision	Recall	Accuracy	F1
[24]	76	42	76	54.10
[25]	97	95	95	95.98
[27]	97	98	97.10	98
[6]	80.50	86.12	80	83.51
Our best results of the current authors with ensemble and classification	100	100	100	100



Figure 2. A comparison among our best results and other works with ensembles stage

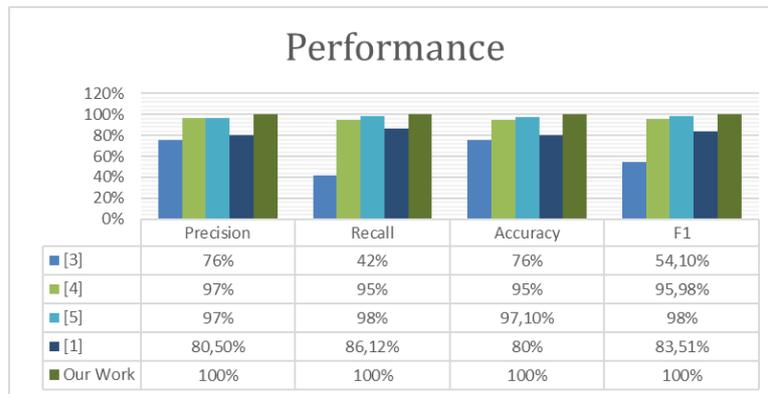


Figure 3. A comparison among our best results and other works with classification stage

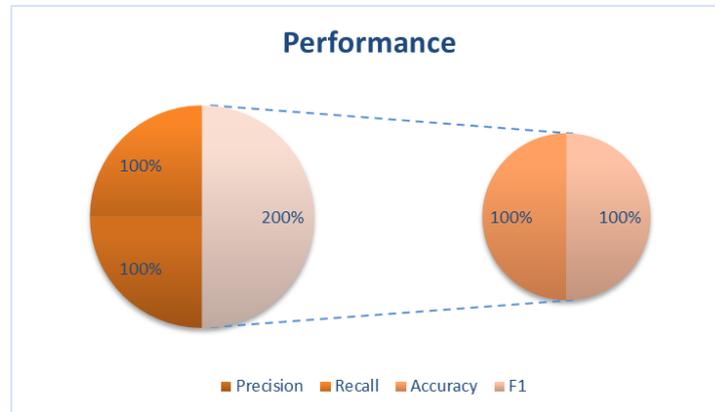


Figure 4. The best result in our work is the first experiment with ensemble techniques

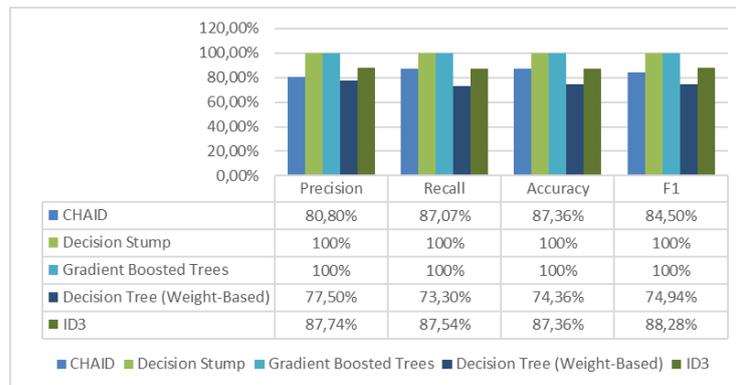


Figure 5. The best result in our work is the second experience with the classification

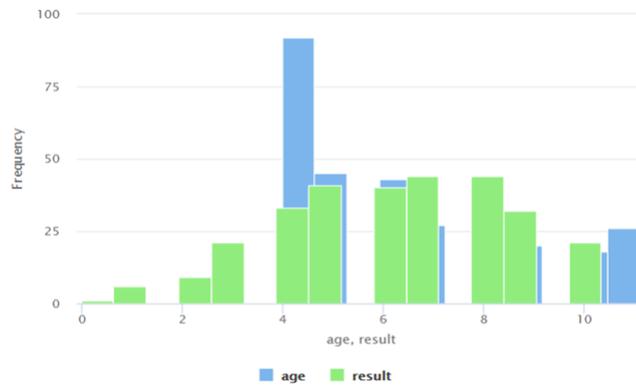


Figure 6. The ROC for Histogram of Polynomial by Binominal Classification

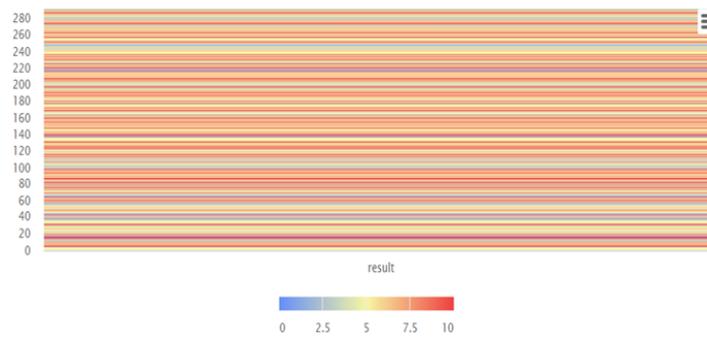


Figure 7. The ROC for Heatmap Polynomial by Binominal Classification

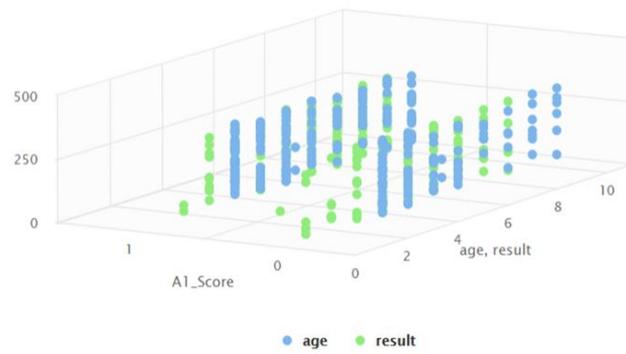


Figure 8. The ROC for scatter 3D of Bayesian Boosting

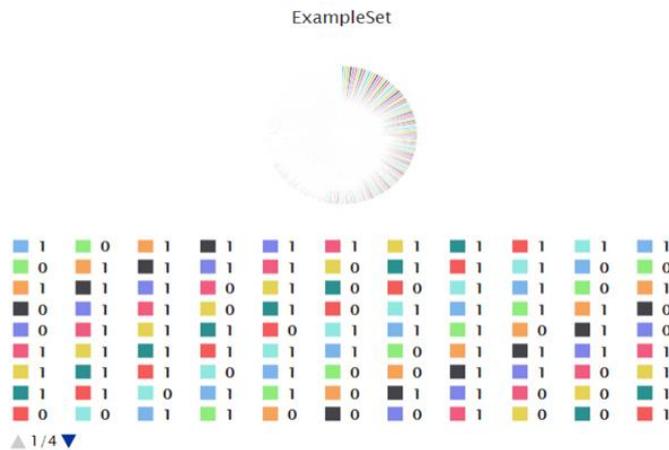


Figure 9. The ROC for pie/donut of Classification by Regression

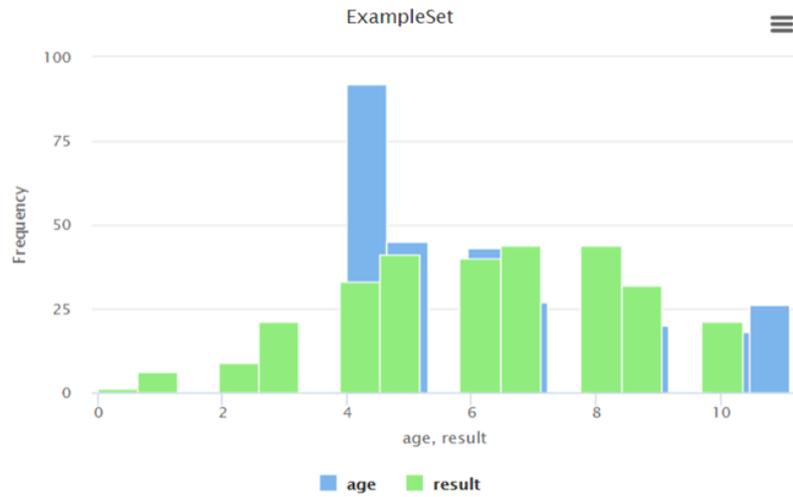


Figure 10. The ROC for Histogram of Classification by Regression

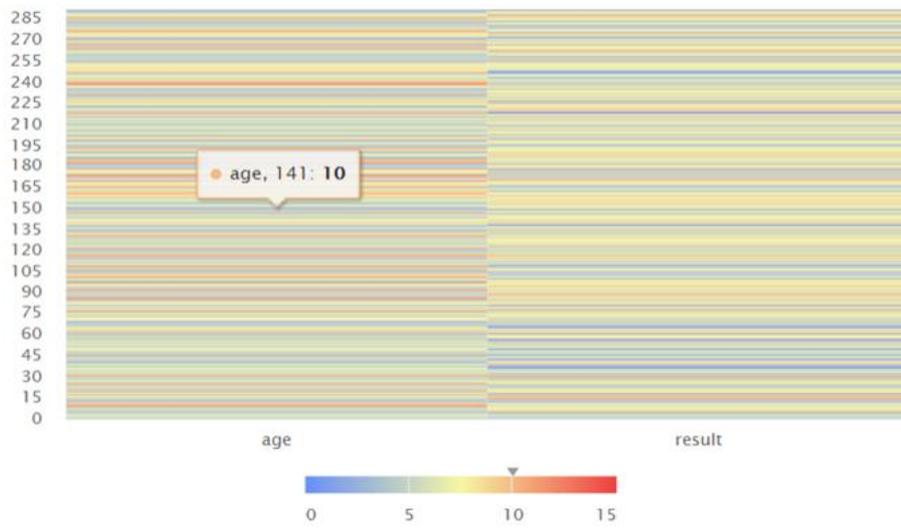


Figure 11. The ROC for Heatmap of Decision Stump

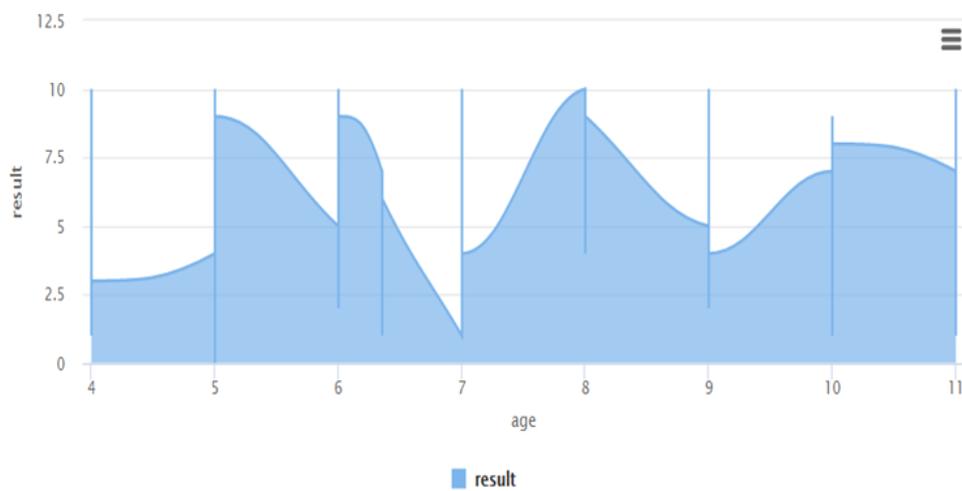


Figure 12. The ROC for spline Area with Gradient Boosted Tree

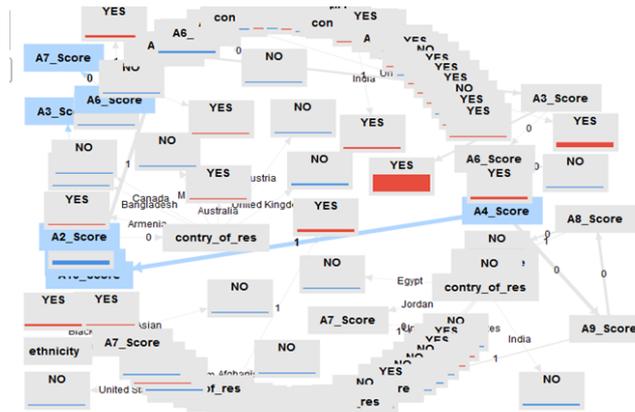


Figure 13. The ROC for Tree ID3

## 5. CONCLUSION

ASD is a group of various disorders that describe impaired behavior and continuous communication among people. ASD suffers from some problems that negatively affect the early diagnosis of this disease, and this makes the ASD a very complex disorder. There are many problems in the ASD records such as missing values, outliers, and low data quality. This complicates the diagnosis of this disease and prediction becomes very small. There are two types of people with the ASD, the first type of sufferers can communicate and do practical, social, and cultural work without confusion. But the second type of people with this disease suffer from many problems. For example, they suffer from severe disabilities and need special care and need. Those who help them spend their work and general needs, and often autism affects education, job opportunities, and social, scientific and cultural matters. Early childhood, but is usually diagnosed at a later stage. People with autism spectrum disorder often suffer from neurological and psychological diseases, the majority of whom are very tired and suffer from comorbidities that include epilepsy, depression, anxiety, and other psychological matters that negatively affect the patient's condition, attention deficit disorder with hyperactivity, in addition to intractable behaviors such as sleep difficulties, self-harm and thinking. The increased negativity and intolerance of bullying from the majority of people make their condition worse. The biggest challenge in this research is to solve and treat the problems that autism spectrum disorder suffers from, and this is the problem of the current research. Our model presented in this research, we downloaded a set of data and records from the UCI website for the ASD. But these data suffer from many problems for example they contain missing values, outliers, outliers and contain low-quality data. It makes autism difficult when we want to explore or diagnose this disease. To solve these problems in our study, we proposed a technical model consisting of two groups. The first group is ensemble techniques such as Bayesian Boosting, Classification by Regression, Polynomial by Binominal Classification with Decision Tree and Random Forest, and KNN with pre-processing techniques. We entered this data and applied it to algorithms as well as techniques Pretreatment such as replace with missing value with mean and detect outlier with KNN. After applying the techniques of the first group and the first experiment, we reached results that far outperform the previous works. The highest accuracy in the first experiment and reached 100%, and it improved the performance of the ASD. At the second experiment, we used other ML techniques to be compared with the first experiment. We also used classification techniques that include CHAID, Decision Stump, Decision Tree (Weight-Based), Gradient Boosted Trees, and ID3. We also obtained very high results to outperform their peers. The highest accuracy in the second experiment reached 100%. It proved that the proposed model is good and can improve the ASD performance to predict excellent results.

For future work, we seek and search for the best ways to provide the best results. We want to use other techniques such as unsupervised ML techniques such as clustering techniques. We also want to use fuzzy algorithms to apply on other medical and engineering research to get results that give improvement.

## REFERENCES

- [1] F. Zhu, S. Wang, W. Liu, H. Zhu, M. Li, and X. Zou, "Multi-modal machine learning system in early screening for toddlers with autism spectrum disorders based on response to name," *Frontiers in Psychiatry*, vol. 14, p. 34, 2023.
- [2] N. A. Mashudi, N. Ahmad, and N. M. Noor, "Classification of adult autistic spectrum disorder using machine learning approach," *IAES International Journal of Artificial Intelligence*, vol. 10, p. 743, 2021.
- [3] I. A. Ahmed, E. M. Senan, T. H. Rassem, M. A. Ali, H. S. A. Shatnawi, S. M. Alwazer, *et al.*, "Eye tracking-based diagnosis and early detection of autism spectrum disorder using machine learning and deep learning techniques," *Electronics*, vol. 11, p. 530, 2022.

- [4] M. E. Minissi, I. A. Chicchi Giglioli, F. Mantovani, and M. Alcaniz Raya, "Assessment of the autism spectrum disorder based on machine learning and social visual attention: A systematic review," *Journal of Autism and Developmental Disorders*, vol. 52, pp. 2187-2202, 2022.
- [5] S. Karim, N. Akter, M. J. Patwary, and M. R. Islam, "A review on predicting autism spectrum disorder (asd) meltdown using machine learning algorithms," in *2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, 2021, pp. 1-6.
- [6] A. Kołakowska, A. Landowska, A. Anzulewicz, and K. Sobota, "Automatic recognition of therapy progress among children with autism," *Scientific reports*, vol. 7, pp. 1-14, 2017.
- [7] N. Chaitra, P. Vijaya, and G. Deshpande, "Diagnostic prediction of autism spectrum disorder using complex network measures in a machine learning framework," *Biomedical Signal Processing and Control*, vol. 62, p. 102099, 2020.
- [8] M. J. Maenner, M. Yeargin-Allsopp, K. Van Naarden Braun, D. L. Christensen, and L. A. Schieve, "Development of a machine learning algorithm for the surveillance of autism spectrum disorder," *PloS one*, vol. 11, p. e0168224, 2016.
- [9] J. Bhola, R. Jeet, M. M. M. Jawarneh, and S. A. Pattekari, "Machine learning techniques for analysing and identifying autism spectrum disorder," in *Artificial Intelligence for Accurate Analysis and Detection of Autism Spectrum Disorder*, ed: IGI Global, 2021, pp. 69-81.
- [10] K.-F. Kollias, C. K. Syriopoulou-Delli, P. Sarigiannidis, and G. F. Fragulis, "The contribution of machine learning and eye-tracking technology in autism spectrum disorder research: A systematic review," *Electronics*, vol. 10, p. 2982, 2021.
- [11] I. Voinsky, O. Y. Fridland, A. Aran, R. E. Frye, and D. Gurwitz, "Machine Learning-Based Blood RNA Signature for Diagnosis of Autism Spectrum Disorder," *International Journal of Molecular Sciences*, vol. 24, p. 2082, 2023.
- [12] H. Zhang, X. Tang, C. Feng, Y. Gao, H. Qi, X. Zhang, *et al.*, "The use of data independent acquisition based proteomic analysis and machine learning to reveal potential biomarkers for autism spectrum disorder," *Journal of Proteomics*, p. 104872, 2023.
- [13] S. Peketi and S. B. Dhok, "Machine Learning Enabled P300 Classifier for Autism Spectrum Disorder Using Adaptive Signal Decomposition," *Brain Sciences*, vol. 13, p. 315, 2023.
- [14] H. S. Nogay and H. Adeli, "Machine learning (ML) for the diagnosis of autism spectrum disorder (ASD) using brain imaging," *Reviews in the Neurosciences*, vol. 31, pp. 825-841, 2020.
- [15] G. Devika Varshini and R. Chinnaiyan, "Optimized machine learning classification approaches for prediction of autism spectrum disorder," *Ann Autism Dev Disord. 2020; 1 (1)*, vol. 1001, 2020.
- [16] K. Chowdhury and M. A. Iraj, "Predicting autism spectrum disorder using machine learning classifiers," in *2020 International conference on recent trends on electronics, information, communication & technology (RTEICT)*, 2020, pp. 324-327.
- [17] A. Baranwal and M. Vanitha, "Autistic spectrum disorder screening: prediction with machine learning models," in *2020 International conference on emerging trends in information technology and engineering (ic-ETITE)*, 2020, pp. 1-7.
- [18] R. Asgarnezhad and S. A. Monadjemi, "An effective combined method for data aggregation in WSNs," *Iran Journal of Computer Science*, vol. 5, pp. 167-185, 2022.
- [19] R. Asgarnezhad and A. Monadjemi, "NSE: An effective model for investigating the role of pre-processing using ensembles in sentiment classification," *Journal of Advances in Computer Research*, vol. 12, pp. 27-41, 2021.
- [20] R. Asgarnezhad and K. Ali Mohsin Alhameedawi, "PEML-E: EEG eye state classification using ensembles and machine learning methods," *Journal of Advances in Computer Engineering and Technology*, vol. 7, pp. 147-156, 2021.
- [21] Z. T. Zoana, M. W. Shafeen, N. Akter, and T. Rahman, "Application of machine learning in identification of best teaching method for children with autism spectrum disorder," *arXiv preprint arXiv:2302.05035*, 2023.
- [22] X. Liu, D. Han, M. Somel, X. Jiang, H. Hu, P. Guijarro, *et al.*, "Disruption of an evolutionarily novel synaptic expression pattern in autism," *PLoS biology*, vol. 14, p. e1002558, 2016.
- [23] J. S. Dufek, J. D. Eggleston, J. R. Harry, and R. A. Hickman, "A comparative evaluation of gait between children with autism and typically developing matched controls," *Medical Sciences*, vol. 5, p. 1, 2017.
- [24] G. Leroy, Y. Gu, S. Pettygrove, M. K. Galindo, A. Arora, and M. Kurzius-Spencer, "Automated extraction of diagnostic criteria from electronic health records for autism spectrum disorders: development, evaluation, and application," *Journal of medical Internet research*, vol. 20, p. e10497, 2018.
- [25] N. G. Raju, K. Madhavi, G. S. Kumar, G. V. Reddy, K. Latha, and K. L. Sushma, "Prognostication of autism spectrum disorder (ASD) using supervised machine learning models," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 8, pp. 2249-8958, 2019.
- [26] R. Asgarnezhad and S. A. Monadjemi, "NB VS. SVM: A contrastive study for sentiment classification on two text domains," *Journal of applied intelligent systems & information sciences*, vol. 2, pp. 1-12, 2021.
- [27] M. S. J. D. S. Kumar, "Prediction and Comparison using AdaBoost and ML Algorithms with Autistic Children Dataset."
- [28] M. D. Hossain, M. A. Kabir, A. Anwar, and M. Z. Islam, "Detecting autism spectrum disorder using machine learning techniques: An experimental analysis on toddler, child, adolescent and adult datasets," *Health Information Science and Systems*, vol. 9, pp. 1-13, 2021.
- [29] R. Sujatha, S. Aarthy, J. Chatterjee, A. Alaboudi, and N. Jhanjhi, "A machine learning way to classify autism spectrum disorder," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, pp. 182-200, 2021.
- [30] C. Küpper, S. Stroth, N. Wolff, F. Hauck, N. Kliewer, T. Schad-Hansjosten, *et al.*, "Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning," *Scientific reports*, vol. 10, p. 4805, 2020.