International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V70I4P230

Volume 70 Issue 4, 343-359, April 2022

Original Article

Effective Autism Spectrum Disorder Prediction to Improve the Clinical Traits using Machine Learning Techniques

Dr. R. Surendiran[1], Dr. M. Thangamani[2], C. Narmatha[3], M. Iswarya[4]

[1]*School of Information Science, Annai College of Arts and Science, Kumbakonam, India, ORCID: 0000-0003-1596-7874*
[2]*Department of Information Technology, Kongu Engineering College, Erode, India. ORCID: 0000-0001-8864-3315*
[3,4]*Department of Computer Technology, Kongu Engineering College, Erode, India.*

[1]Corresponding author: surendiranmca@gmail.com

Received: 17 November 2021 Revised: 03 March 2022 Accepted: 10 April 2022 Published: 30 April 2022

Abstract **-** *Autism spectrum disorder (ASD) is a neuro-developmental complaint that influences an individual's communication, announcement, and knowledge talents. Analysis of Autism can be completed at any age-group level. Autism patients look at diverse kinds of disputes learning disabilities, and complexity with meditation. Mental health problems, motor difficulties, and sensory problems are some of the problems faced by Autism patients. Earlier diagnosis and proper medication at the early stage are essential to control ASD. The ASD prediction framework is built to support a behavioral aspect-based analysis model without any device in this research. The ASD prediction process is focused on the childhood and adolescent analysis model utilized in the system. The behavioral parameters are collected with the support of the Autism Query collections. The decision tree (DT) and Support Vector Machine (SVM) techniques, K-Nearest Neighbors (KNN), and Artificial Neural Network (ANN) are applied for the ASD prediction process. The Correlated Feature selection based Random Forest (CFS-RT) algorithm is applied for the ASD prediction process, giving an accuracy of 93.03%, and ANN produces 97.68% and outperformance other methods.*

Keywords **-** *Autism Spectrum Disorder, Decision Tree, Machine Learning, Data Mining, Support Vector Machine.*

1. Introduction

Autism spectrum disorder (ASD) is a progressive disability that can benefit important basis community, contact, and behavioral issues. There are frequent tells how a community with ASD looks that sets them together from new persons. Still, individuals with ASD could converse, interrelate, function, perform and hear in the approach part from mainly individual people. The knowledge, thoughts, and difficulty-resolving facility of persons with ASD can provide a choice from talented to more confronted. An analysis of ASD contains a lot of provisions that applied to be analyzed alone as autistic disorder syndrome. These situations in named ASD.

ASD is a state connected to mind growth that forces how a person recognizes and meet peoples, obtaining difficulty in public contact and dealings. The disorder also contains genetic factors, environmental aspects, biological factors, and inadequate and recurring behavior patterns. In the United States, more children's affected by Autism. Each person with Autism has different issues, like anxiety, seizures, and depression, as shown in Figure 1. Autism can be obtained at two or three and diagnosed as early as before 18 months. Early detection gives a good impact on the life of persons. Persons with ASD may perform work, interconnect, interrelate, and acquire behaviors that are changed from person to person. The skills of a person with ASD can differ significantly. The person with ASD are issues with the social announcement and limited behaviors. The children have varying habits of culture and give less attention. Diagnosing ASD is also a problem because there is no proper medical test or blood test to identify the disorder. Doctors see the child's behavior and can start a diagnosis to increase the quality of life.

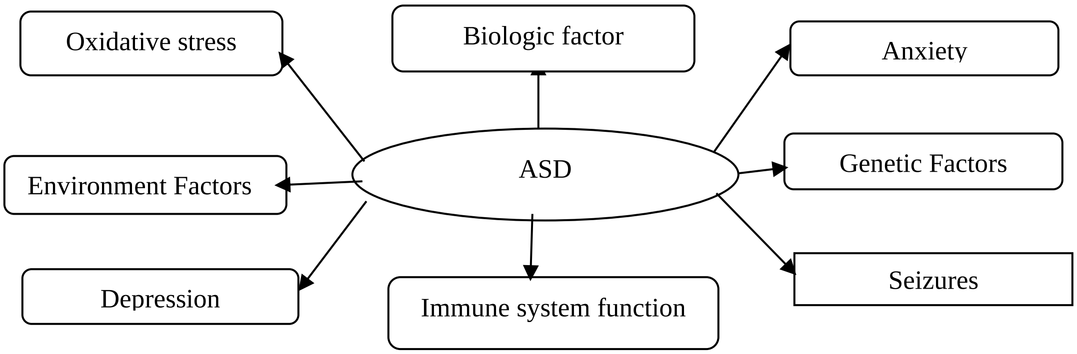


Fig. 1 Outline the Factors for obtaining ASD

ASD disturbs how an individual reacts, study, and states themselves. For parents of a child with ASD, significantly which level the child has can help prepare them for issues their child might face in daily life. Figure 2 shows the level of ASD. The first level needs help for every work, and the second level indicates the need for sustainable help. Three-level indicate the need for more sustainable help and very problem to changing focus and actions.

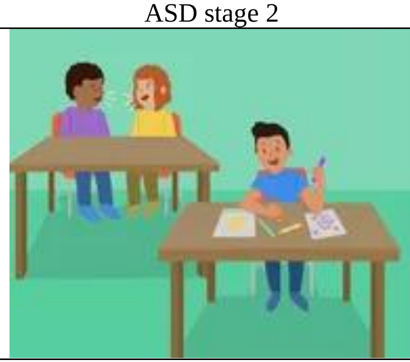
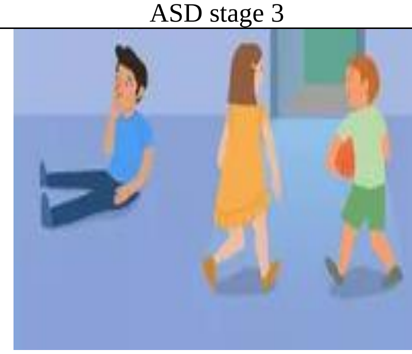
ASD stage 1	ASD stage 2	ASD stage 3
		

Fig. 2 Type of ASD

1.1 Supervised Machine Learning

Machine Learning (ML) is a division of Artificial Intelligence (AI) that studies and determines useful models in data to formulate diagnoses in the medical field. It is put up on mathematical opinions of likelihood and data and information. This supervised learning is an ML approach that studies by plotting the input features and output mutable for prediction. It is shown in Figure 3 and learning through training and test data set.

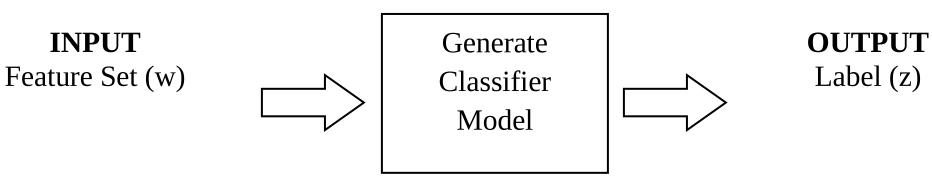


Fig. 3 Classifier Model

As shown in Figure 1, a representative controlled learning difficulty includes attribute space W called features in the data set, a label space Z, and a prediction space Z'. Classifier map each feature w to one of the already defined class labels z.

A training sample set is illustrated as $Y = ((w_1,z_1),\ldots,(w_m,z_m)) \in (W \times Z)^m$ that consists of prearranged class labels, where each $w_i \in Z$. The output is a model H_s: $W \rightarrow Z'$ that learns from the sample set These approaches find the class of new instances, $Z' = Z = \{\pm 1\}$ and input samples are a map with two labels as $Z = \{\pm 1\} = \{-1, +1\}$ in binary classification.

1.2 Classification Algorithms

The ID3 algorithm is an approximation supervised learning algorithm through the sample for various labels. The decision tree root and internal nodes hold the feature test uniqueness, whereas the leaf node includes the class label. The decision tree holds numeric and categorical variables from the sample and produces a visual representation. Many open-source tools contain a build-in decision tree algorithm. For example, WEKA produces good data visualization by visualize the tree option.

ID3 identifies features that discriminate one label from another. The features should be continuous or chosen from a group of identified values, and the feature should be well-known in progress. At the same time, this algorithm is more responsive to features when a huge number of data. Limitations in the ID3 are overcome through the C4.5 algorithm with the help of information gain measures.

Naïve Bayes is a supervised learning classifier for data mining and the field of machine learning that solve the classification issues by generating the Bayes rule. This classifier holds any volume of data and also holds missing values effectively. The target is to forecast the class l as the label, fi as features, and select a significance of l that exploit $P(l|F_1, F_2,\ldots, F_n)$ given a sample with attributes (m1, m2,…, m_n) is shown in equation (1)

$$P(c|F) - \frac{P(m|l)P(l)}{P(m)} \quad (1)$$

K-Nearest Neighbors is an instance-based ML method for classifier, regression approach, and it's lazy learning because there is no explicit training. This algorithm stores the training observations and evaluates against the new opinions with the help of a similarity assessment as Euclidean distance according to real and binary values as shown in equations (2) and (3)and positions the stored data at some point in the test instant to formulate calculations. This algorithm also holds the missing values effectively.

Euclidean Distance:

$$f(w_i,z_j) = \sqrt{\sum_{im=1}^{D}(w_{im} - w_{jm})^2} \quad (2)$$

Hamming Distance:

$$f(w_i,z_j) = \sum_{im=1}^{D} II(w_{im} \neq w_{jm})^2 \quad (3)$$

The sample data j fields to frame a j-dimensional vector, $w=(w_1,w_2,\ldots,w_j)$, and the training data is characterized by $F=\{(w_1,z_1),\ldots(w_j,z_j)\}$ where x_i is the content of the input and y_i is the named as the label. The w_i is a vector that contains F attributes, whereas x_{im} represents the m-th attributes of w. That means the result is $z_i \in \{1,..., C\}$

2. Related Work

Thangamani et al.[1-3] investigated different models for disease prediction using ML approaches. Early detection and follow-up healings have the most important blow on the autistic group. Dilantha et al. [4] constructed an EEG approach control improved classifier approach for ASD. The present diagnostic follow-up is behavior needy and delays the prediction at a premature age. This research used the feature extraction technique to extract the relevant attributes. The authors [35] applied ML and data mining algorithms to predict ASD in toddlers and children. The researchers [6] developed the system with the help of ML techniques for autism grading and assessment for the diagnosis of autistic children. It produces 98% of accuracy. Inon Wiratsin et al. [7] proposed a feature selection algorithm to detect the relevant attributes and suggest treatment for various age groups. Bekerom et al.[33] investigated the ML algorithm for early detection of ASD to develop good mental health of the child. Daniel Bone [9] planned to use the ML approach to get a better result in Autism and diagnosis with multi-level instrument fusion for human behavior analysis and clinical progress. Lucia Billeci [10-11] deliberated an incorporated eye way techniques with EEG for learning of replying and beginning joint consideration in ASD. The correlation between eye and brain features is measured in [12-13]. The authors [14-17] illustrated the study of EEG signals for ASD detection using ML approaches. William J. Bosl [18] suggested a data drive method for EEG diagnostics for a timely finding of ASD. Zhong Zhao et al.[19] addressed restricted features are used according to patient behavior patterns, which is used in ML algorithms to detect the ASD prediction and produce an accuracy of 88.7%. J. Baio et al. [20] developed a surveillance network for ASD prediction, approximating the popularity of ASD and further unique children and evaluating the system by skilled clinicians.

Thabtah et al. [34] designed an ASD screening method with ML variation, which perform classification task and generate a predictive model to detect the ASD effectively. The authors [22] focused on ML algorithms with behavioral features to identify the ASD at the earliest. N. Jassimet al. [23] quantitatively compressed discovery from fMRI learning of non-community sensory observation in autistic evaluated to direct representative contributors by the behavior of a series of conservatively- threshold Activation Likelihood Estimation meta-analyses. Suman Raj et al. [24] Suggested ML Algorithm predicts and diagnoses ASD problems for adults, adolescents, and children. J. Charpentier et al. [25] diagnosed ASD by the mutual presence of societal destructions and warning, recurring performance models. Antoine Frigaux et al. [36] proposed a new technique to detect autism spectrum disorder. Jennie Hayes et al. [27] reviewed all the approaches used and suggested clinical guidelines estimate ASD in adults and children in the UK. Mengyi Liao et al.[28] investigate the ASD by facial and EEG data of children with the help of the ML approach and obtain 87.50%. This model used the naïve Bayes algorithm and did not train more than one model. Chelsea et al. [29] reviewed the state of MLCand ASD with the help of some questions and answered for quires. The authors use the ML model with MRI and fMRI attributes to know the neuronal behavior changes in ASD to predict the disease earlier [30-31].

From existing research, Autism screening tests are a huge cost and not a time-saving one. The behavioral differences are not measured in the device-dependent ASD prediction process. Technical people are required to handle the data collection operations. Age group-based analysis is not supported in the ASD prediction process.

3. Problem Definition

The AI and ML techniques are used in autism detection. The neuroimaging and kinematic data are used to discover the Autism Spectrum ASD levels. The ASD classification is carried out with electroencephalogram (EEG) signal dispensation and knowledge approaches. The EEG is a consistent biomarker for diagnosing ASD abnormalities. The electrodes attached to the scalp are used to detention the EEG signals. The mobile applications collect answers from the people for the Autism Questions. The MLg-based methods are applied to predict ASD at any age-group level. The Random Forest-CART (Classification and Regression Trees) algorithm predicts ASD traits. The Random Forest-ID3 (Iterative Dichotomiser 3) algorithm is applied to perform the ASD prediction process. The Correlation Feature Selection based Random Forest (CFS-RF) technique is built to discover ASD with behavioral features from different age levels.

Specific Objectives

- To design a system for Autism Spectrum Disorder (ASD) discovery with behavioral analysis.
- To perform age group based analysis on the ASD discovery process
- To design the Correlation-based Feature Selection with Random Forest (CFS-RT) technique for the ASD prediction process
- To eliminate device and technician dependency in the ASD detection process
- To minimize cost and time in the ASD prediction operations

4. Proposed System Methodology

4.1 ASD Detection through Behavioral Features using Decision Trees

Adult autism detection is designed to analyze the behavioral attributes of patients above 18 years. The behavioral features are gathered with the help of a questionnaire. The Adult Questioner (AQ10) data set is get from the University of California, Irwin ML repository. Each person is diagnosed with 10 questions. The ASD is identified using the responses given by the individual. The decision tree classification is applied to discover the disease levels. Figure 4 shows the proposed system architecture, and a detailed view is shown in Fig. 5.

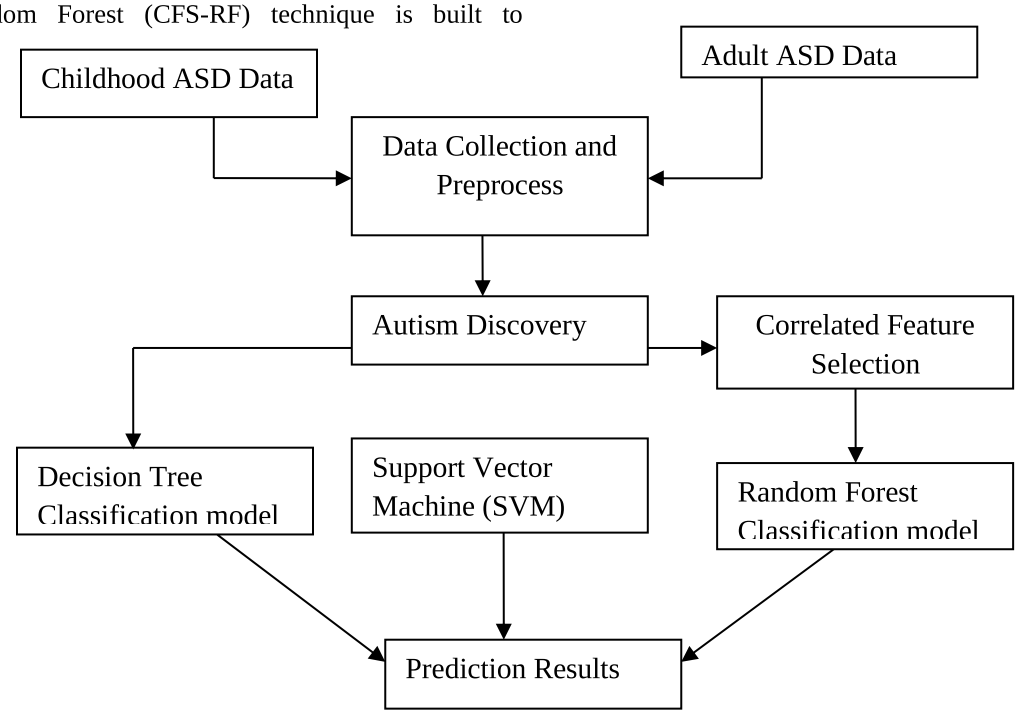


Fig. 4 System Architecture Diagram

The children dataset and the adolescent dataset is stored in the data repository. It contains missing and incorrect values. This can be preprocessed and get relevant data to build the ML classification models after preprocessing, splitting the operation as testing and training to generate the classifier to detect the Autism of children and adolescents. The proposed system uses DT, SVM, and Random Forest models to generate a good model to predict Autism in the early stage. In classification, 10-fold cross-validation is applied to estimate the model.

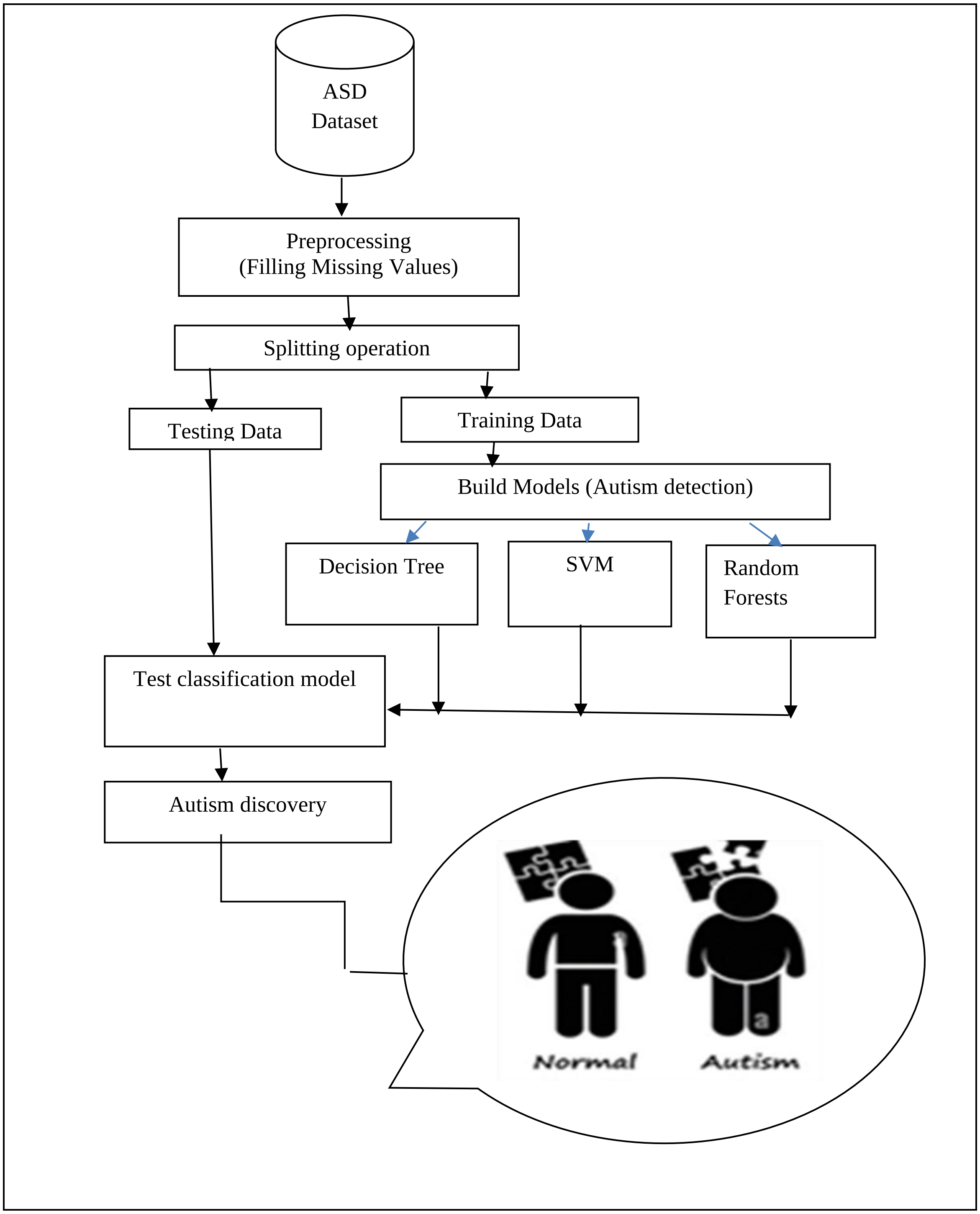


Fig. 5 Proposed Autism discovery model

Algorithm steps to create the decision tree in R-Environment

For creating decision tree, first load from database usin R tool as follows

```
# Load the ASD Data from datasets
# Attribute Relationship File Format (ARFF) data
ASDData <- read.csv("ASDData.arff", header = FALSE,
comment.char = "@")
print("All ASD data")
print(paste("Number of Rows    : ",nrow(ASDData)))
print(paste("Number of Columns : ",ncol(ASDData)))
print("       Attribute Details")
print("1. Answer for Question - 1")
      print("2. Answer for Question - 2")
      ..................................................
    ....................................................
print("19. Used the screening app befosinre")
print("20. Screening Method Type")
("21. Screening Score")
```

View the data by the command as summary(ASDData)

Missing value analyzed using the following coding

```
ASDData[, 13][ASDData[, 13] == '?'] <- NA
ASDData[, 20][ASDData[, 20] == '?'] <- NA
missmap(ASDData)
```

Next step is data cleaning process to remove inconsistent data

```
 # Data cleaning process
 Assign alternate values to missing values
 <- sum(is.na(ASDData[1]))
(paste("Missing values in Variable 1 : ",mcount))
mcount <- sum(is.na(ASDData[2]))
................
-----------------
```

Then data optimization process by the following command

```
ASDData <- ASDData[c(1:12,14,32,21)]
# Data Factorization - Categorical data coversion for class attribute
# Data spliting for training and testing process
DataMat <- createDataPartition(y = ASDData$V21, p = 0.5, list = FALSE)
TrainData <- ASDData[DataMat, ]
TestData <- ASDData[-DataMat, ]

# Dimensions of the split
print("Class probability rate analysis for ASD data set")
prop.table(table(ASDData$V21)) * 100
print("Class probability rate analysis for ASD training data set")
prop.table(table(TrainData$V21)) * 100
print("Class probability rate analysis for ASD testing data set")
prop.table(table(TestData$V21)) * 100
TrainData[["V21"]] <- factor(TrainData[["V21"]])
#Matrix construction with penalty
penalty.matrix <- matrix(c(0,1,10,0), byrow=TRUE, nrow=2)
# Building the decision tree
#Classification Tree construction
tree <- rpart(TrainData$V21~., data=TrainData, parms = list(loss = penalty.matrix), method = "class")
# visualization Tree process
rpart.plot(tree, nn=TRUE)
# ASD prdiction using the tree model
TestPredict <- predict(object=tree,TestData,type="class")
TestPredict
# Performance of the aid
# Build Confussion Matrix
confusionMatrix(table(TestPredict, TestData$V21))
print(paste("Time : ",Sys.time()))
Options (warn=0)
```

Figure 6 shows the workflow of the decision tree to discover Autism in children and adults data. The input dataset is given to the classifier for modeling. Split the dataset using information gain. The highest information gain gives the best solution. Its result is positive, and it predicts that the patient may have ASD based on reading the given questionnaires. Otherwise patient doesn't suffer from any autistic spectrum disorder. Figure 7 shows the flow of the SVM classifier for autism discovery. In this model, "C " is the number of misclassified samples. It contains a small value and no weightage for misclassification. Compute the loss function using equation 4. Figure 8 shows the autism detection by the Random Forest classifier of the machine learning model. It combines all the results and aggregate results based on majority voting.

$$L=1/2||W||^2+C \tag{4}$$

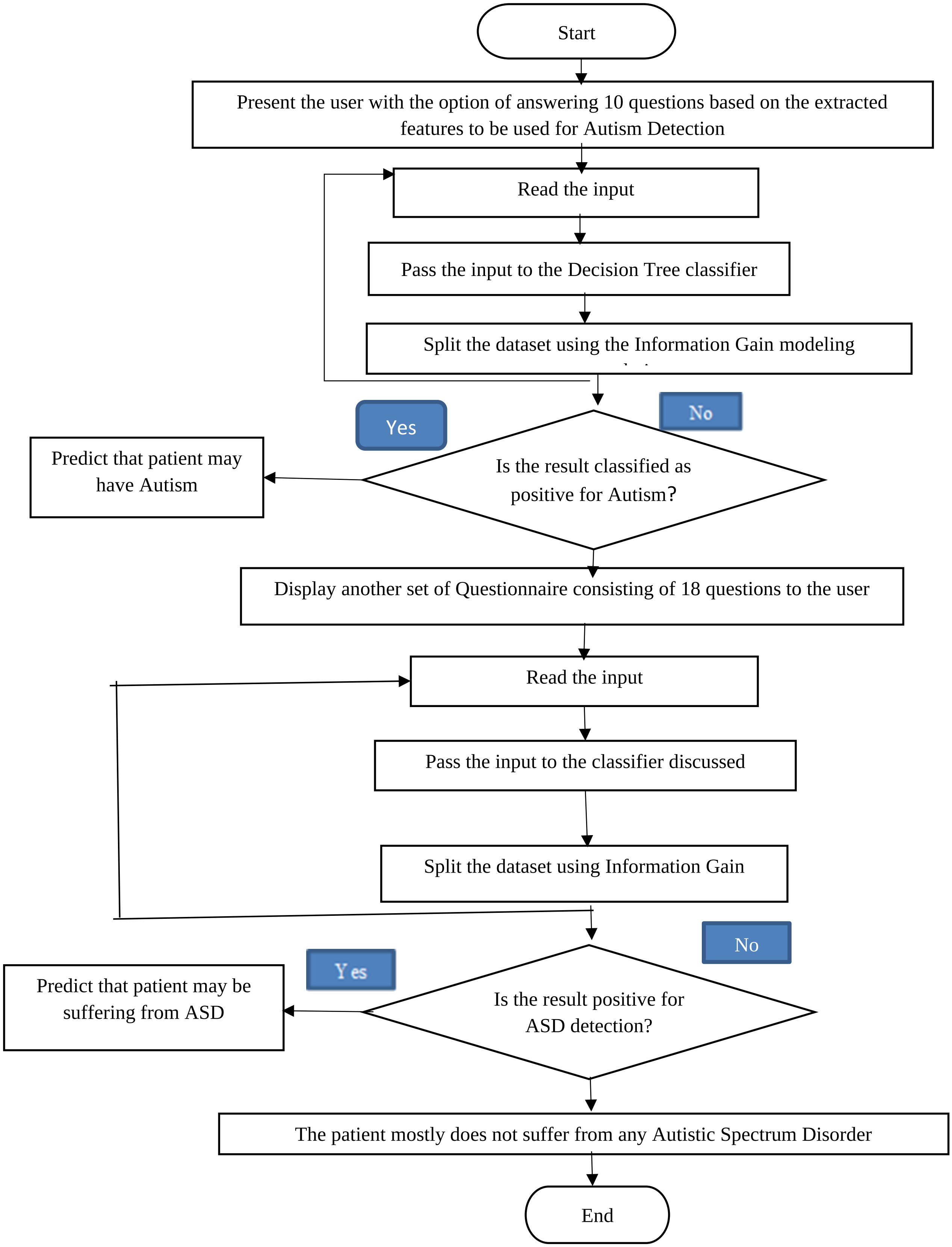


Fig. 6 Workflow of Decision Tree classification model

4.2 Support Vector Machine

Algorithm step to building SVM in R environment

```
# Load the CARET library
library('caret')
# Dispaly the structure of the dataset in the DataFrame
str(ASDData)
# Dispaly the observations
View(ASDData)
# Display the dimension of the data set
print("ASD data dimension : ")
dim(ASDData)
# Displays the Summary of the ASD data values
summary(ASDData)
Data preprocessing: Replacing missing values with suitable values
# Data splitting process
DataMat <- createDataPartition(y = ASDData$V21, p = 0.02, list = FALSE)
TrainData <- ASDData[DataMat, ]
TestData <- ASDData[-DataMat, ]
# Display the dimension of the TrainData
dim(TrainData)
# Display the dimension of the TestData
dim(TestData)
# Data Factorization - Categorical data coversion for class attribute
# Training process for SVM classification
# trctrl <- trainControl(method = "repeatedcv", number = 10, repeats=3)
trctrl <- trainControl(method = "cv", number = 10)
SVM_Linear <- train(V21 ~., data = TrainData, method = "svmLinear", trControl = trctrl, preProcess = c("center", "scale"), tuneLength = 10)
TrainData[["V21"]] <- factor(TrainData[["V21"]])
# Autism Spectrum Disorder (ASD) prediction process with trained model
TestPredict <- predict(SVM_Linear, newdata = TestData)
TestPredict
Accuracy analysis process
```

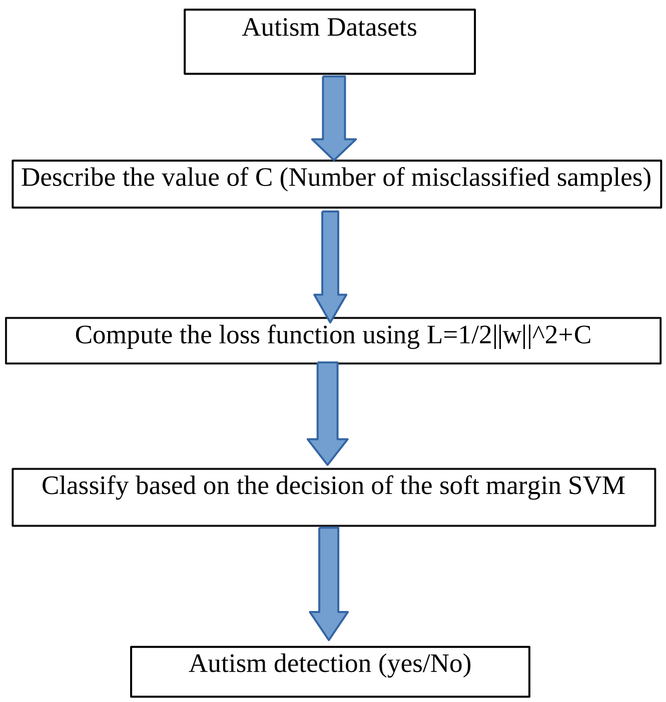


Fig. 7 Generating the Model using SVM classifier

4.3Autism Spectrum Disorder Prediction using Random Forest

The Autism Spectrum Disorder (ASD) prediction framework is built to support a behavioral aspect-based analysis model without any device. The behavioral parameters are collected with the support of the Autism Query collections. The ASD prediction process is focused on the age boundary-based analysis process. The childhood and Adolescents analysis model is utilized in the system. Random Forests is an ensemble classification that creates multiple decision trees during the training period and produces individual trees by means of mode or means finding. Random Forest of this machine learning algorithm avoids the overfitting problem in the Decision Tree.

Step to build Random Forest
The Random Forests relate to the bagging techniques. Suppose training set W = w1 and with reply Z = z1 and selects a sample with replacement of the training sample.
1. Sample, with replacement, n training examples from W, Z,
2. Train a classification
Data splitting process
Categorial data conversion for class attributes
Autism Spectrum Disorder (ASD) prediction process with trained model
TestPredict <- predict(RFModel, newdata = TestData)
TestPredict
# Accuracy analysis process
# Build Confussion Matrix
confusionMatrix(table(TestPredict, TestData$V21))
print(paste("Time : ",Sys.time()))
options(warn=0)

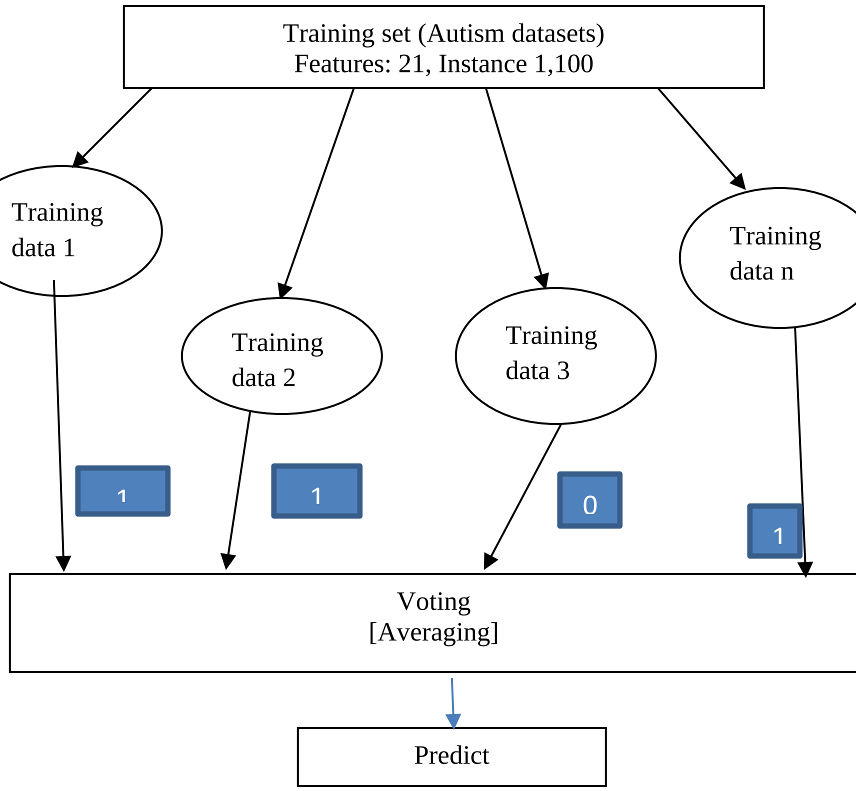


Fig. 8 Autism detection using Random Forest Machine learning Model

The ML algorithms are practiced to perform the ASD prediction process. The Correlated Feature selection (CFS) based Random Forest algorithm is applied for the disease prediction process. The prediction scheme is associated with the Naïve Bayes and Support Vector Machine (SVM) classifiers.

4.4 Other Machine Learning Models

K- Nearest Neighbor (KNN) model used for regression and classification issues. The 'K' denotes the number of kernel theme that is to be nominated to minimize the error. KNN is constructed on the knowledge of similarity, which can relate to Euclidean distance, nearness, or closeness. The artificial Neural Network (ANN) model is a neural network that has a joining with numerous neurons. Each neuron cell has a group of input values and associated weights shown in Figures 9 and 10.

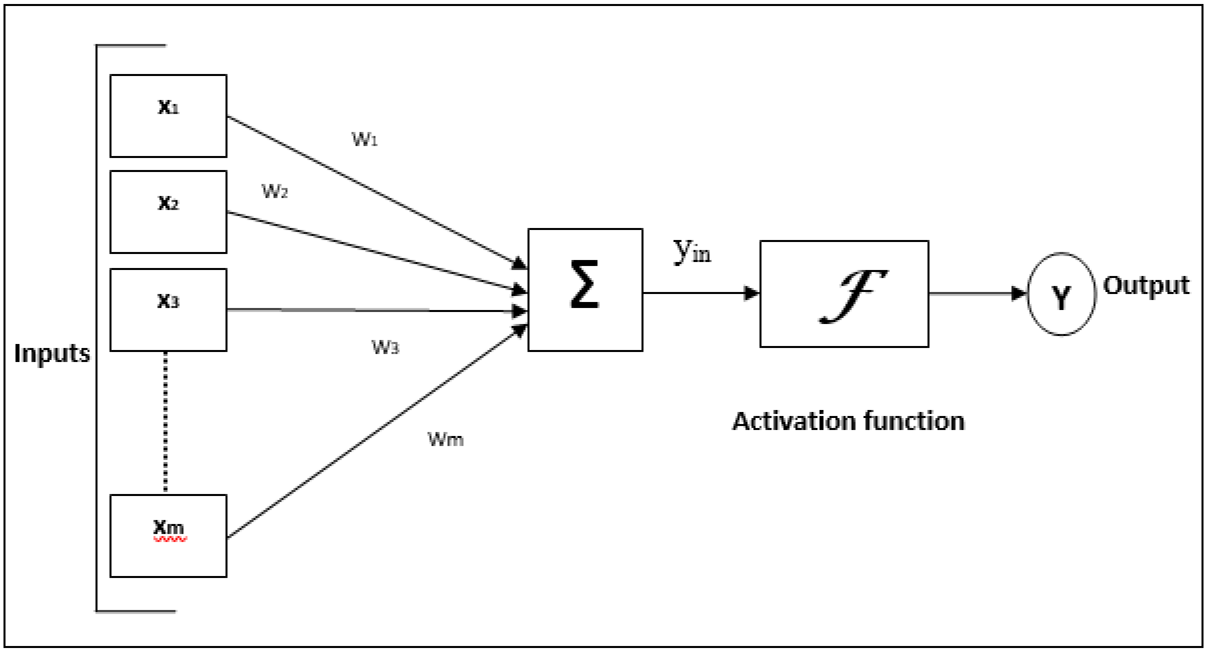


Fig. 9 ANN with associate weight

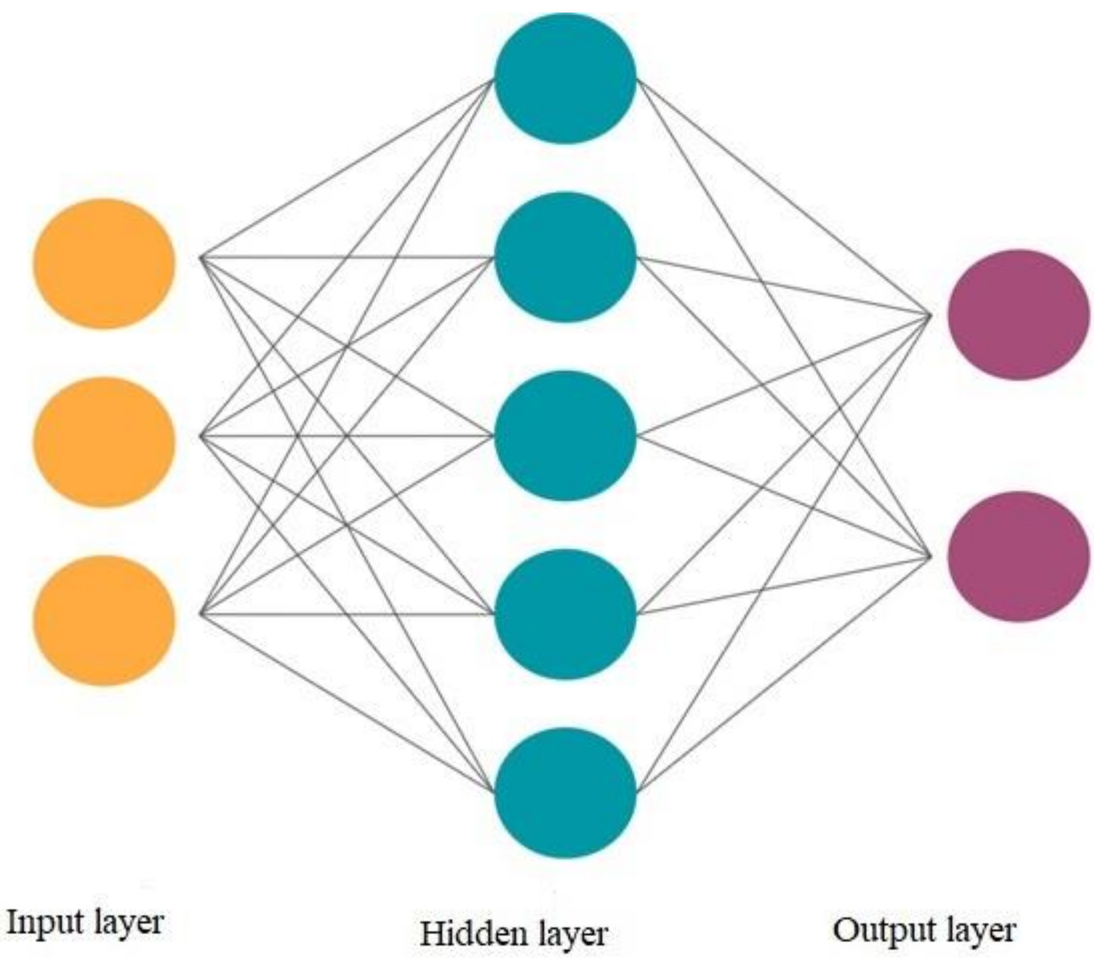


Fig. 10 Architecture of ANN

5. Experiment Result and Discussion

5.1 Environmental setup

The autism prediction models are implemented and verified using the R language. The prediction process is evaluated with four factors. They are False Positive Rate (FPR), False Negative Rate (FNR), prediction accuracy, and prediction time measures. The FPR rate and FNR rate parameters are used to estimate the fault level of the ASD detection process. The FPR and the FNR are the findings of error criteria. The error stage should be reduced to increase the detection accuracy levels. The prediction accuracy is applied to estimate the ASD discovery result accuracy levels. The processing time is calculated for the prediction process.

5.2 Data set description

The ASD detection process is implemented to detect the autism stages in childhood, adolescence, and adulthood. The AG10 dataset is taken from the University of California machine learning repository. The children's data set has 292 instances. The adolescent data set has 104 instances. The adult data set is collected for 704 patients. The entire data set has 1100 instances with 21 categorical, continuous, and binary attributes. The ML-based ASD forecast process uses the Decision Tree Classification (DTC), SVM, and CFS-RF techniques. Table 1 and Table 2 show the ASD data set representation.

Table 1. Data set representation

S.No	Name of the dataset	Instances in dataset
1	Adult-ASD screening dataset	704
2	Children-ASD screening dataset	292
3	Adolescent-ASD screening dataset	104

Table 2. Features in ASD screening data

Features	Features explanation
1	Age of the patient
2	Sex
3	Nationality
4	By birth patient suffered from jaundice
5	Patient family numbers anyone affected by progress disorders
6	Who is fulfilling the experiment
7	Country
8	Screening application used by the user before or not?
9	Test type
10-19	Based on the screening method, answer the questions
20	Screening Score

5.3 Measures used for evaluation

The verified criteria are used to assess the worth of a classifier. For verifying a classifier by contingency matrix, precision and recall, Accuracy, ROC, and time analysis. The contingency matrix contains True Positive (tp), False Positive (fp), False Positive (fp), and False Negative (fn)

- tp – Quantity of optimistic instances in the trial or experimental dataset that are categorized as positive
- fp – Quantity of undesirable instances in the trial or experimental dataset that are categorized as positive
- tn – Quantity of undesirable instances in the trial or experimental dataset that are categorized as negative
- fn – Quantity of optimistic instances in the trial or experimental dataset that are categorized as negative

Accuracy of the classifier: Accuracy is repeatedly used result metric for prediction. Obtaining more accuracy in the model in the ML approach is the highest aspect of education (5).

$$\text{Classification Error} = \frac{fp+fn}{tp+tn+fp+fn} \quad (5)$$

5.4 Performance Analysis

5.4.1 False Positive Rate (FPR) Analysis

The relation knows the FPR of falsely assigned positive ASD results. The FPR evaluated between the DTC, SVM, and CFS-RF techniques are shown in Table 3 and Figure 11. The SVM-based technique decreases the FPR rate by 20% more than the DTC technique. The CFS-RF) decreases the FPR rate by 35% more than the Support Vector Machine (SVM) technique.

Table 3. FPR Analysis between Decision Tree Classifier (DTC), SVM, and CFS-RF Techniques

Records	DTC (%)	SVM (%)	CFS-RF (%)
220	9.48	7.92	4.74
440	8.88	7.64	4.44
660	8.36	7.14	4.18
880	7.64	6.56	3.82
1100	6.92	6.28	3.46

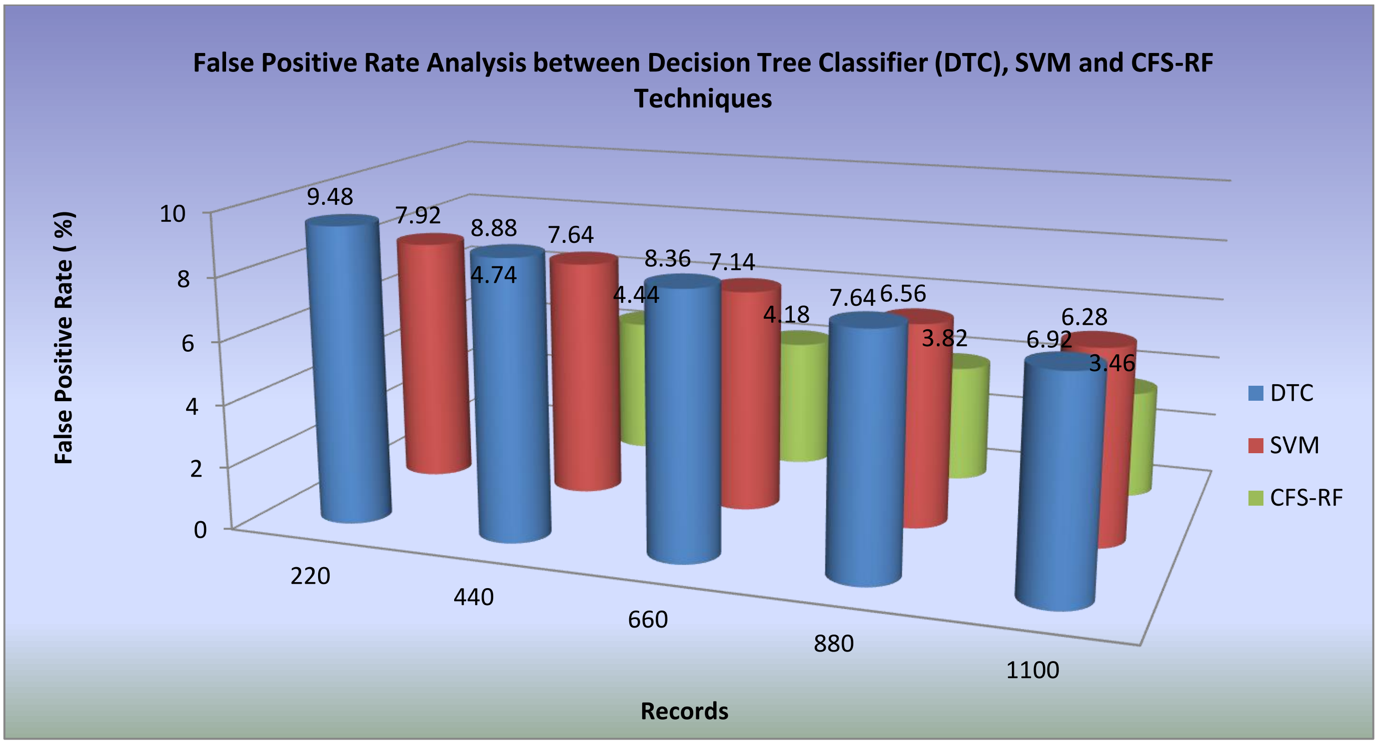


Fig. 11 FPR Analysis between Decision Tree Classifier (DTC), SVM, and CFS-RF Techniques

5.4.2 False Negative Rate (FNR) Analysis

The FNR is calculated as the proportion of falsely consigned negative ASD performance. The FNR analysis among the Decision Tree Classifier (DTC), SVM, and CFS-RF techniques are shown in Table 4 and Figure 12. The SVM-based technique decreases the FNR by 25% more than the DTC technique. The CFS-RF reduces the FNR by 30% more than the SVM technique.

Table 4. False Negative Rate Analysis between Decision Tree Classifier (DTC), SVM, CFS-RF Techniques

Records	DTC (%)	SVM (%)	CFS-RF (%)
220	10.04	7.56	5.02
440	9.32	7.34	4.66
660	8.61	7.04	4.32
880	7.96	6.72	3.98
1100	7.08	6.38	3.52

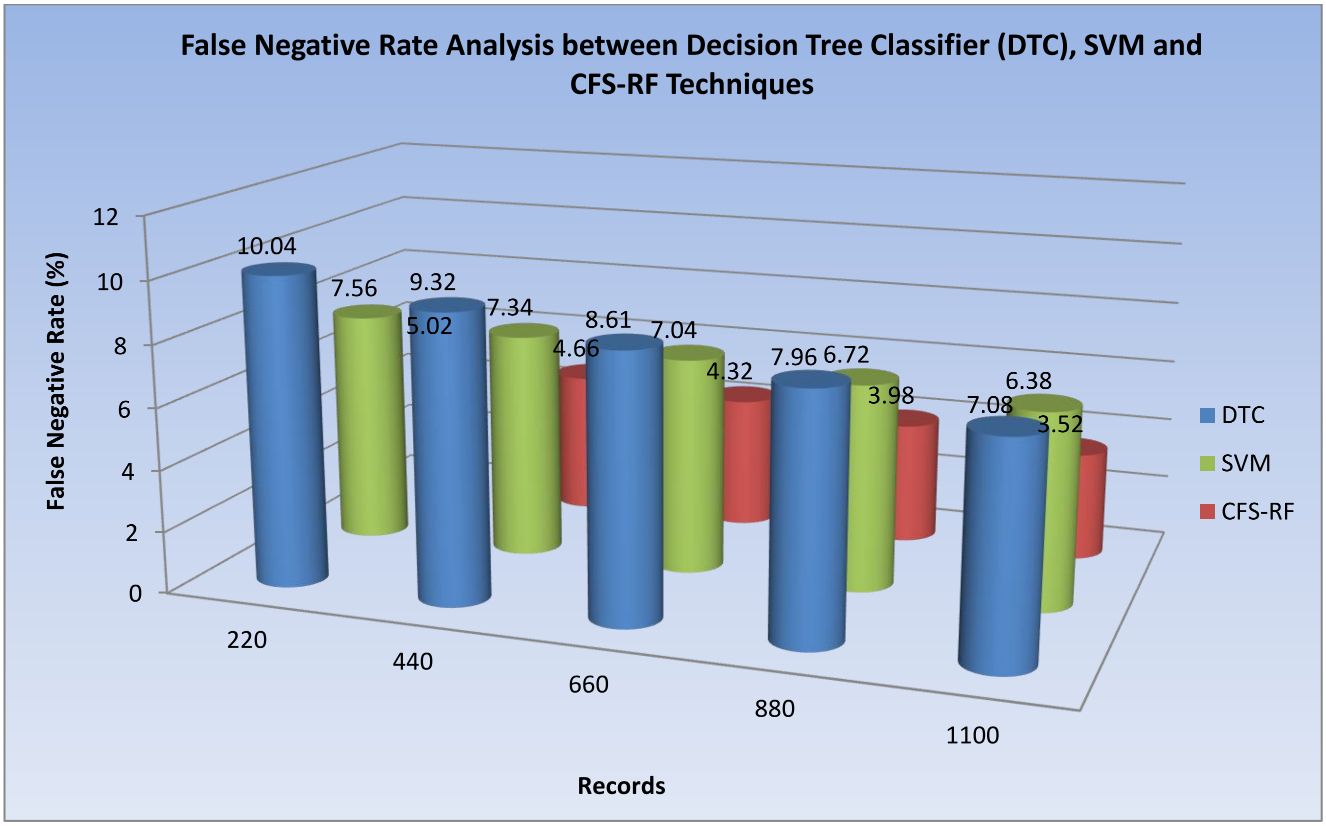


Fig. 12 False Negative Rate Analysis between Decision Tree Classifier (DTC), SVM, and CFS-RF Techniques

5.4.3 Prediction Accuracy Analysis

The prediction accuracy rate is evaluated by fetching the accuracy level of the ASD prediction process. The prediction accuracy analysis between the Decision Tree Classifier (DTC), SVM, and CFS-RF techniques is shown in Figure 13 and Table 5. The SVM-based technique increases the prediction accuracy level by 5% more than the DTC technique. The CFS-RF increases the prediction accuracy level by 10% more than the SVM technique. Figure 14 indicates the comparison of KNN and ANN. The ANN produces good accuracy of 97.68%.

Table 5. Prediction Accuracy Analysis between Decision Tree Classifier (DTC), SVM, and CFS-RF Techniques

Records	DTC (%)	SVM (%)	CFS-RF (%)
220	80.48	84.52	90.24
440	82.19	85.01	90.89
660	83.03	85.82	91.49
880	84.41	86.72	92.21
1100	86.02	87.34	93.03

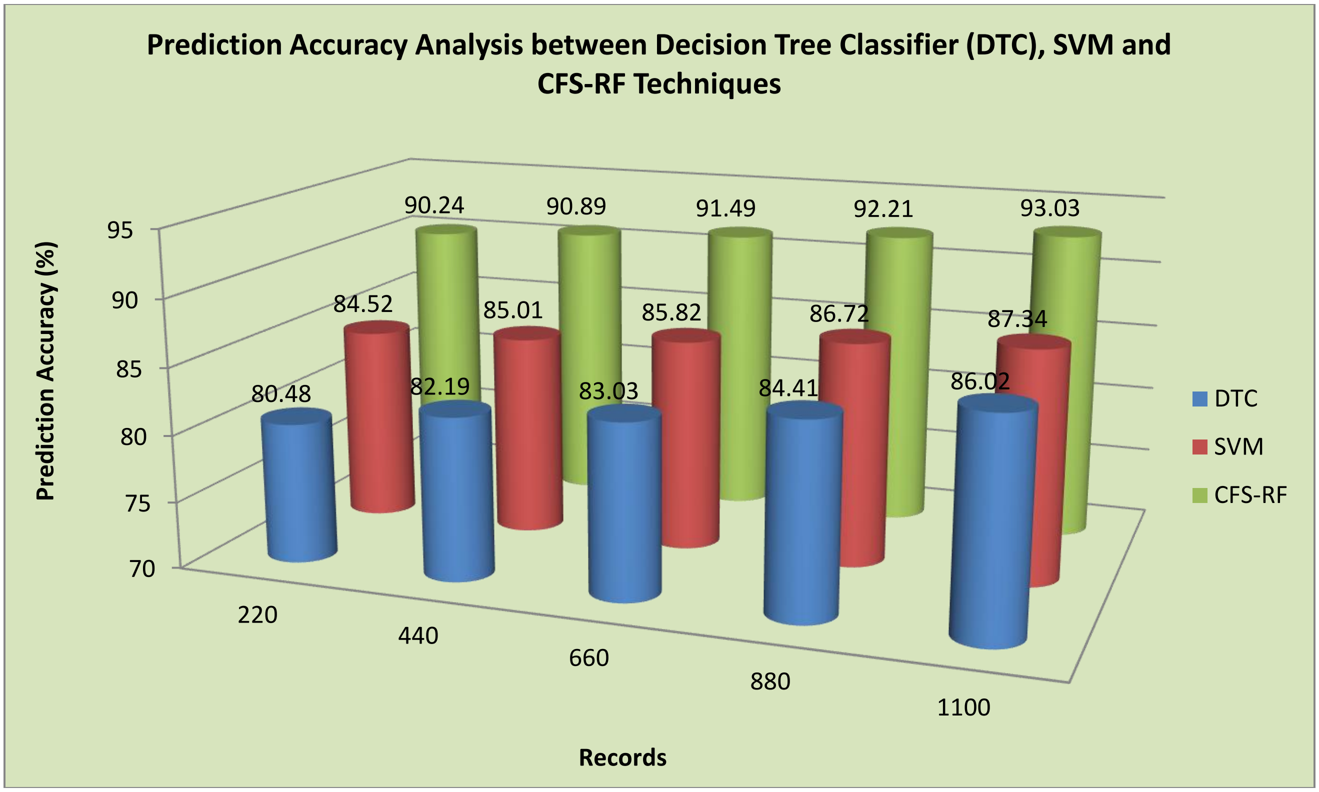


Fig. 13 Prediction Accuracy Analysis between Decision Tree Classifier (DTC), SVM, and CFS-RF Techniques

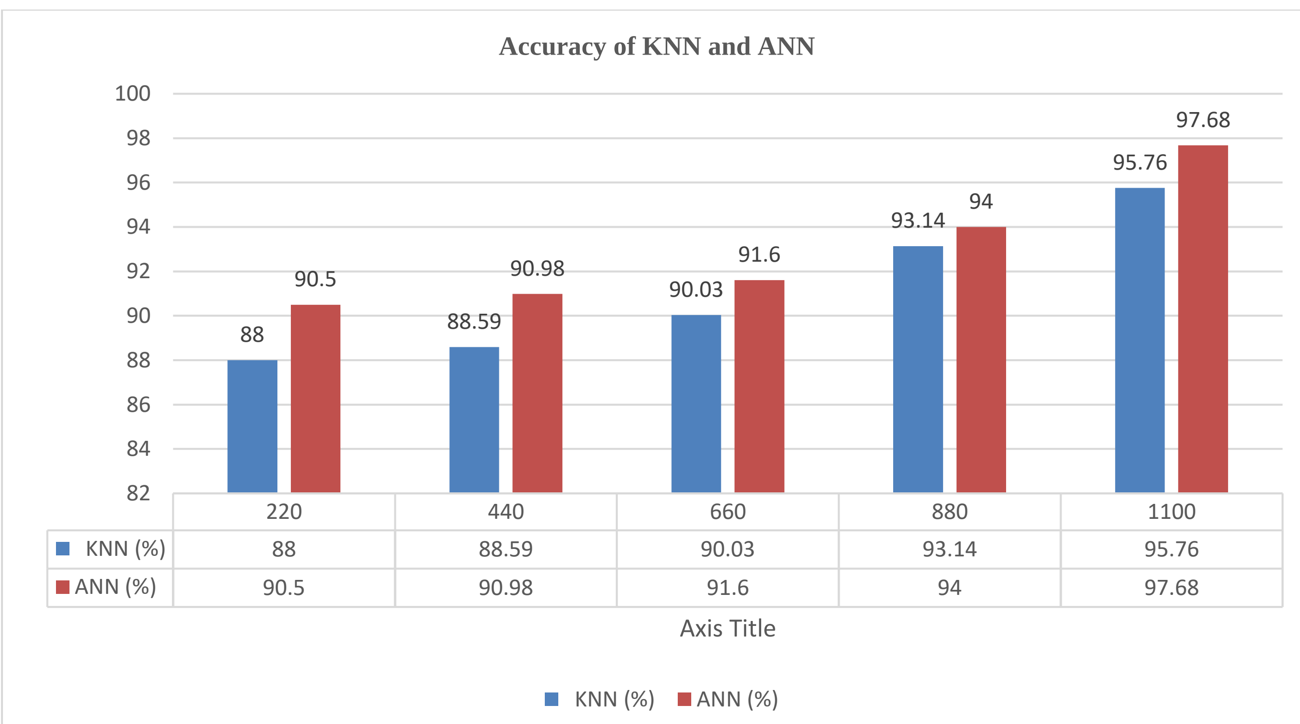


Fig. 14 Prediction Accuracy Analysis between KNN and ANN

Prediction Time Analysis

The prediction time analysis is applied to estimate the duration of the ASD prediction process. The prediction time analysis between the DTC, SVM, and CFS-RF techniques is shown in Table 6 and Figure 15. The SVM-based technique reduces the prediction time by 25% more than the DTC technique. The CFS-RF reduces the prediction time by 30% more than the SVM technique.

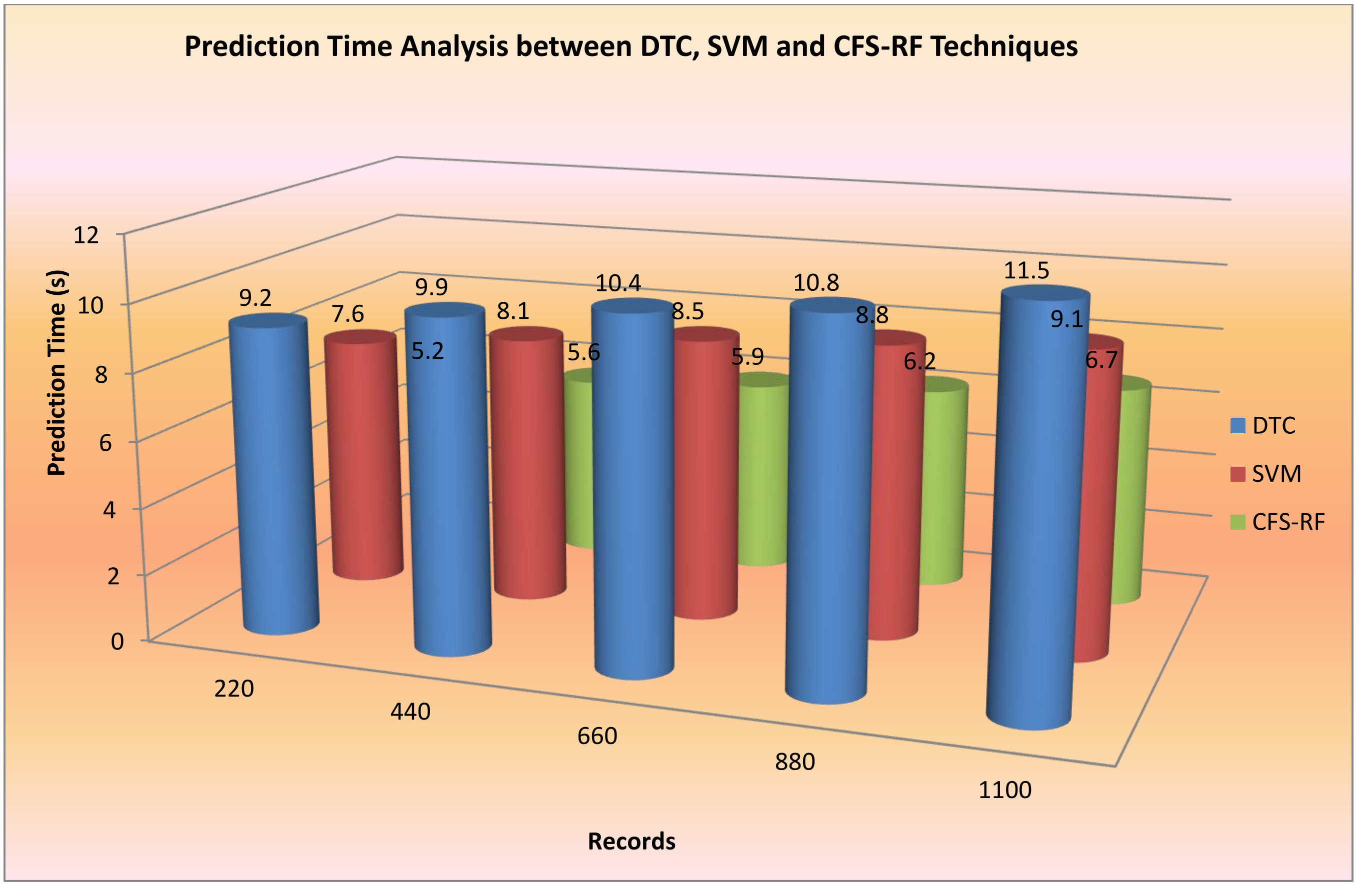


Fig. 15 Prediction Time Analysis between DTC, SVM, and CFS-RF Techniques

6. Conclusion and Future work

The Correlation Feature Selection based Random Forest model increases the accuracy level in the prediction process. The behavioral and age group parameters are integrated into the prediction process. The data dimensionality is managed with feature selection methods. The system solves the device dependency, technician requirements, and delay problems. The ASD mechanism can be enhanced to analyze the adult autism levels. The ASD prediction system can be enhanced with the following features.

- The system can be improved with automated medicine and a treatment suggestion model.
- The system can be improved to analyze the treatment and its impacts on regular patients' activities.
- The ASD prediction process can be upgraded by integrating behavioral features and EEG-based features analysis.
- The ASD prediction process can be implemented as a mobile-based application to conduct an automated self-assessment test.

References

[1] M. Thangamani, R. Vijayalakshmi, M. Ganthimathi, Rajita, P. Malarkodi, S. Nallusamy, Efficient classification of heart disease using K-Means clustering Algorithm, International Journal of Engineering Trends and Technology, 68(12) (2020) 48-53.

[2] N. Suresh Kumar, M. Thangamani, V. Sasikumar S. Nallusamy, An Improved Machine Learning Approach for Predicting Ischemic Stroke, An Improved Machine Learning Approach for Predicting Ischemic Stroke, 68(1)(2021) 111-115.

[3] Thangamani M., Ganthimathi M., Sridhar S.R., Akila M., and Keerthana R. Engineering, Detecting Coronavirus Contact using Internet of things., International Journal of Pervasive Computing and Communications, 16(5) (2020) 447-456. https://doi.org/10.1108/IJPCC-07-2020-0074.

[4] Dilantha Haputhanthri, Gunavaran Brihadiswaran, Sahan Gunathilaka, Dulani Meedeniya, and Sampath Jayarathna, An EEG based Channel Optimized Classification Approach for Autism Spectrum Disorder, Moratuwa Engineering Research Conference (Mercon), (2019).

[5] Surendiran, R., and Alagarsamy,K., 2010. Skin Detection Based Cryptography in Steganography (SDBCS) .International Journal of Computer Science and Information Technologies (IJCSIT), 1(4) (2010) 221-225.

[6] Supriya H S, Sumanth Alva R*, Suprith K P, Vikas Kumar L, and P Kyshan Neheeth, applying supervised learning technique to diagnose autism spectrum disorder (asd), International Journal of Recent Scientific Research , 11(5b) (2020) 38439-38441. DOI: http://dx.doi.org/10.24327/ijrsr.2020.1105.5312.

[7] Inon Wiratsin, L. Narupiyakul, Feature Selection Technique for Autism Spectrum Disorder, Proceedings of the 5th International Conference on Control Engineering and Artificial Intelligence, DOI, 10.1145/3448218.3448241, (2021).

[8] Surendiran, R., 2017. Development of Multi-Criteria Recommender System, SSRG International Journal of Economics and Management Studies (IJEMS) ISSN: 2393 - 9125, 4(1) (2017) 28-33.

[9] Daniel Bone, Somer Bishop, Matthew P. Black, and Shrikanth 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, (2016).

[10] Lucia Billeci, Antonio Narzisi, Alessandro Tonacci, Beatrice Sbriscia-Fioretti, Luca Serasini, Francesca Fulceri, Fabio Apicella, Federico Sicca, Sara Calderoni & Filippo Muratori An integrated EEG and eye-tracking approach for the study of responding and initiating joint attention in Autism Spectrum Disorders, Scientific Reports, 7 (2017) 1-13.

[11] Filippo Muratori, Lucia Billeci, Sara Calderoni, Maria Boncoddo, Caterina Lattarulo, Valeria Costanzo, Marco Turi, Costanza Colombi and Antonio Narzisi, How Attention to Faces and Objects Changes Over Time in Toddlers with Autism Spectrum Disorders: Preliminary Evidence from An Eye-Tracking Study, Brain science, MDPI, 9(344) (2019) 1-11.doi:10.3390/brainsci9120344.

[12] S. Jayarathna, Y. Jayawardana, M. Jaime, and S. Thapaliya, Electroencephalogram (EEG) for Delineating Objective Measure of Autism Spectrum Disorder, in Computational Models for Biomedical Reasoning and Problem Solving, IGI Global, (2019) 34-65.

[13] S. Thapaliya, S. Jayarathna, and M. Jaime, Evaluating the EEG and Eye Movements for Autism Spectrum Disorder, in IEEE International Conference on Big Data (Big Data), (2018) 2328-2336.

[14] Surendiran, R., Similarity Matrix Approach in Web Clustering .Journal of Applied Science and Computations (JASC),5(1) (2018) 267-272.

[15] Narzisi, A. Posada, M. Barbieri, F. Chericoni, N. Ciuffolini, D. Pinzino, M. Romano, R. Scattoni, M.L.Tancredi, R. Calderoni, S., Prevalence of autism spectrum disorder in a large Italian catchment area: A school-based population study within the ASDEU project. Epidemiol. Psychiatr. Sci. (2018).

[16] E. Grossi, C. Olivieri, and M. Buscema, Diagnosis of autism through EEG processed by advanced computational algorithms: A pilot study, Comput. Methods Programs Biomed., 142 (2017) 73-79.

[17] W. J. Bosl, H. Tager-Flusberg, and C. A. Nelson, EEG Analytics for Early Detection of Autism Spectrum Disorder: A data-driven approach, Scientific Reports, 8(1) (2018) 1-20.

[18] Zhong Zhao, Xiaobin Zhang, Wenzhou Li, Xinyao Hu, Xingda Qu, Xiaolan Cao, Yanru Liu, and Jianping Lu, Applying Machine Learning to Identify Autism With Restricted Kinematic Features, IEEE Access, (2019).

[19] J. Baio et al., Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014, MMWR. Surveillance Summaries, 67(6) (2018) 1-23.

[20] Surendiran, R., Rajan, K.P. and Sathish Kumar, M., Study on the Customer targeting using Association Rule Mining. International Journal on Computer Science and Engineering, 2(7) (2010) 2483-2484.

[21] Charlotte Küpper, Sanna Stroth, Nicole Wolff, Florian Hauck, Tanja Schad-Hansjosten, Luise poustka, Veit Roessner, Katharina Schultebraucks & Stefan Roepke, identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and

[22] Adults using machine learning, Scientific report, nature research, 10(4805) (2020). doi.org/10.1038/s41598-020-61607-w.

[23] NaziaJassim, Simon Baron-Cohen, and John Suckling, Meta-analytic evidence of differential prefrontal and early sensory cortex activity during non-social sensory perception in Autism, Neuroscience and Biobehavioural Reviews, Elsevier, 127 (2021) 146-157.

[24] Suman Raj and Sarfaraz Masood, Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques, Procedia Computer Science, Elsevier, 167 (2020) 994-1004.

[25] Judith Charpentier, Marianne Latinus, Frederic Andersson, Agathe Saby, Jean-Philippe Cottier, Frederique Bonnet-Brilhault, EmmanuelleHouy-Durand, and Marie Gomot, Brain Correlates of Emotional Prosodic Change Detection in Autism Spectrum Disorder, Neuro Image: Clinical, Elsevier, 28 (2020) 102512.

[26] Surendiran, R. and Alagarsamy, K., PCA-based geometric modeling for automatic face detection. Int. J. Comput. Sci. Inform. Technol, 1(4) (2010) 221-225.

[27] Jennie Hayes, Tamsin Ford, Hateem Rafeeque, Ginny Russell, Clinical practice guidelines for the Diagnosis of Autism Spectrum Disorder in Adults and Children in The UK: a narrative review, BMC Psychiatry, 13(18) (2018) doi: 10.1186/s12888-018-1800-1.

[28] Mengyi Liao, Hengyao Duan, and Guangshuai Wang, Application of Machine Learning Techniques to Detect the Children with Autism Spectrum Disorder, Hindawi Journal of Healthcare Engineering, 2022 (2022) 1-10, Article ID 9340027.https://doi.org/10.1155/2022/9340027.

[29] Chelsea M. Parlett-Pelleriti, Elizabeth Stevens, Dennis Dixon, Erik J. Linstead, Applications of Unsupervised Machine Learning in Autism Spectrum Disorder Research: a Review, Journal of Autism and Developmental Disorders, (2022) 1-17. https://doi.org/10.1007/s40489-021-00299.

[30] Cooper J. Mellema 1,2, Kevin P. Nguyen 1,2, Alex Treacher 1 & Albert Montillo, Reproducible neuroimaging features for diagnosis of Autism Spectrum Disorder with Machine Learning, Scientific Reports, 0123456789 (2022) 1-13.https://doi.org/10.1038/s41598-022-06459-2.

[31] Yinghan Zhu, Hironori Nakatani, Walid Yassin, Norihide Maikusa, Naohiro Okada, Akira Kunimatsu, Osamu Abe, Hitoshi Kuwabara, Hidenori Yamasue, Kiyoto Kasai, Kazuo Okanoya, and Shinsuke Koike, Application of a Machine Learning Algorithm for Structural Brain Images in Chronic Schizophrenia to Earlier Clinical Stages of Psychosis and Autism Spectrum Disorder: A Multiprotocol Imaging Dataset Study, Schizophrenia Bulletin, 10 (2022) 1-12.

[32] N. Sciaraffa, G. Borghini, P. Aricò, Joint Analysis of Eye Blinks and Brain Activity to Investigate Attentional Demand during a Visual Search Task, Brain Science, MDPI, 11(562) (20210) 1-20. DOI:10.3390/brainsci11050562.

[33] Bram van den Bekerom, Using Machine Learning for Detection of Autism Spectrum Disorder, IEEE, (2017).

[34] Fadi Thabtah, Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment, ACM, (2017).

[35] Rasool Azeem Musa, Mehdi Ebadi Manaa2, and Ghassan Abdul-Majeed Predicting Autism Spectrum Disorder (ASD) for Toddlers and Children Using Data Mining Techniques, Journal of Physics: Conference Series, IOP publishing, doi:10.1088/1742-6596/1804/1/012089, 1804 (2021).

[36] Antoine Frigaux ,Joëlle Lighezzolo-Alnot , Jean-Claude Maleval and Renaud Evrard, Differential diagnosis on the Autism Spectrum: Theorizing an Ordinary Autism, L'evolution ´ psychiatrique, Elsevier, 86(2021) e1–e24.