Toddler ASD Classification Using Machine Learning Techniques

Ashima Sindhu Mohanty, Krishna Chandra Patra
Sambalpur University, Odisha, India

Priyadarsan Parida (✉)
GIET University, Odisha, India
priyadarsanparida@giet.edu

Abstract—At present era, Autism Spectrum Disorder (ASD) has become one of the severe neurologically developed disorders throughout the world and early recognition can substantially get rid of this problem. The proposed work is based on the analysis of unbalanced ASD toddler dataset from UCI data repository. The work in this paper is performed in three stages. In first stage, the original data is preprocessed through converting the categorical attributes to numeric values by the process of frequency encoding followed by standardization of numeric attributes. In the second stage, the dimension of input is reduced using Principal component analysis (PCA). At the end, the classification of ASD Toddler data is performed through different machine learning classification models in two stages viz. through training parameter ε and through k-fold cross validation (k=10). The experimentation yields very high classification performance in comparison with other state-of-art approaches.

Keywords—ASD, Quantitative Checklist of Autism, Standardization, PCA, Machine Learning, Performance Parameters

1 Introduction

Neuro-developmental disorder reflects mental illness where the nervous system is affected. ASD is such a disorder where the social interaction, communication and behavior of an individual are in concern. It is characterized by repeated activities and aimless imaginary thoughts[1]. In Asia, within the common population, the average prevalence of ASD was nearly 1.9/10,000 in the year 1980, which got increased to approximately 14.8/10,000 by the year 2008[2]. It is observed that the early signs of ASD are marked during the initial 6-18 months of a toddler's life span. Further the signs are followed by developmental regression like loss in verbal, social and communication ability with abnormal motor development between 18 and 36 months of the child's life span[3]. Upon all individuals including adult (17 years and above), adolescent (12 to 16 years), children (4 to 11 years) and toddlers (up to 36 months), ASD diagnosis can be implemented. But it is not possible using conventional medical tests like blood test

or body check-ups[4]. There are many sort of ASD screening approaches involving structured questionnaires and interviews to be faced by patients, the procedure which is lengthy and subjective [5]. ASD data sets are developed for all category of individuals and stored in UCI data repository for research purpose. Many researchers successfully classified ASD classes on all category of individuals except the case of toddler. One of the reasons for not considering the toddler dataset in the studies is that major number of toddler cases have been found not associated with ASD which made the entire data set unbalanced.

This paper focuses on classification of ASD from no ASD traits using Machine Learning (ML)[6][7] among toddlers in the age group of 12-36 months. The proposed plan uses Qualitative Checklist for Autism in Toddlers (Q-CHAT-10)[8] based up on many behavioral independent attributes such as age, gender, ethnicity, etc in the data set. The dataset used in this paper was formulated by Fadi Fayez Thabtah using ASDTests mobile application for screening ASD mental disorder in toddlers[9]. The entire method adopted in this paper possibly, can help medical professionals in putting up attention towards individuals with ASD symptoms for further assessment.

2　　Literature Review

When the subject was idle, images of brain were taken by Functional Magnetic Resonance Imaging (fMRI)[10]. From the Autism Brain Imaging Data Exchange (ABIDE)[11], the analysis procured 1035 fMRI occurrences and analyzed the pattern for diagnosis of ASD. For understanding and classifying the distinguished features of neuro-images in autistic brains, deep learning techniques were implemented. It was found that compared to conventional diagnosis methods, neuro-imaging patterns of the brain distinguished ASD with improved accuracy.

In further study, from the subject's brain, the node variability was achieved. Various machine learning models[12]got trained by the obtained variability on resting fMRI ASD and no ASD data of ABIDE. Machine learning (ML) algorithms like Naive Bayes (NB)[13], Random Forest (RF)[14] as well as Support Vector Machines (SVM)[15] were applied on 147 cases. The result was achieved with 60-65 percent performance.

A total of 2925 Social Responsive Scale (SRS) samples got derived from Simons Simplex Collection 15.0, Boston Autism Consortium & AGRE[16]. The test was carried out on 2925 data dividing them into 10 folders each comprising of 65 samples with 10 percent of both ADHD and ASD data[17]. Following feature selection method based on minimal redundancy-maximal relevance [mRMR][18], six ML algorithms-SVM, Linear Discriminant Analysis (LDA)[19], Categorical Lasso (CL) [20] Decision Tree (DT) [21], RF and Logistics Regression Model (LR) [22] were tested using package Scikit-learn package[23]. The investigation showed optimum result for classification of ASD and ADHD test samples by incorporation and fusion of SVM, LDA, CL and LR.

The author in[24] designed a mobile-based ASD screening tool called as ASDTest for testing all category of individuals based upon Q-CHAT and Autism Spectrum Quotient (AQ-10) screening questionnaire[8] in which response to each question scores

a point. The author collected 1452 instances using the app for all categories of individuals but left the toddler cases as they made the entire data unbalanced. The remaining 1100 instances comprised of adult, adolescent and child data sets respectively with 21 numbers of features in each data set. Following feature extraction using wrapping filtering, two ML algorithms: LR and NB were used for classifying the ASD data. Adult data showed the maximum performance with the implementation of LR model.

To improve the performance of detecting ASD class within children, the author utilized fuzzy data mining models[25]. The data set was obtained from UCI data repository collected by ASDTest app developed by the author in[24]. It consists of 509 instances with a distribution of 252 NO ASD and 257 ASD traits and 21 numbers of features. Along with FURIA, a fuzzy data mining algorithm, JRip, RIDOR[26] and PRISM[27] algorithms have been also used to generalize the overall performance. In terms of accuracy as well as sensitivity, classification model of FURIA suppressed other models. But the specificity rate of FURIA went down in comparison with JRIP algorithm set.

Instead of analyzing single category of individual, the author in[28] focused on early detection of ASD in child, adolescent and adult based on supervised learning. The data was gathered from UCI repository from the ASDTest app designed by the author in[24]. Following pre-processing, the author performed the classification on the data sets thereby employing K-Nearest Neighbor (KNN)[29][30], SVM and RF classifier. The pre-processed data was partitioned into the training data expressed in α percent within a range of 50 to 90 percent as well as the testing data expressed as (100 - α) percent. The performance parameters were calculated based up on 5 number of experiments performed with $\alpha = 50$ to 90 percent under two cases: missing and complete data. The RF classifier showed maximum performance for both adult and adolescent data sets up on both complete and missing data.

The author in[31] predicted the chance of ASD in child, adolescent and adult by applying four ML algorithms: NB, LR, SVM, KNN as well as Artificial Neural Network (ANN). Apart from the ML algorithms, Convolution Neural Network (CNN) also got used to predict ASD. The datasets got collected from the UCI Repository gathered by the author in[24] via the ASDTest app. The author pre-processed the raw data and then partitioned the pre-processed data as training and testing dataset in the ratio of 80:20 respectively. Without any dimensionality reduction, all 21 features were used to predict ASD. CNN outperformed rest of the models in all categories of individual's data sets. The category of adult data set yielded maximum result.

3   Data Collection

Dr. Fadi Fayez Thabtah developed a data set by using ASDTest to screen ASD in individuals including toddlers, children, adolescents and adults. The autism screening data set for toddlers got published in July, 2018 at UCI data repository. It is in activation from November, 2018 till date with six unique contributors. The data set is based upon QCHAT-10 toddler ASD screening tool[8]. It incorporates 1054 number of instances

with 17 independent variables excluding the case number and one dependent variable as the output which indicates the ASD class. Out of 1054, 326 numbers of cases are with no ASD class and remaining 728 numbers of cases are with ASD class [9]. Due to such distribution of ASD instances, the respective data set is unbalanced. So, prior to the investigation on the data set, the research aimed to balance the data set by the process of standardization.

4 Proposed Work

The workflow of the proposed method is shown in Fig.1. The complete method is divided into four parts viz. preprocessing of original ASD data, dimension reduction of original data, classification using different machine ML models and performance evaluation of all the ML models. The training is completed in two ways i.e. through training parameter ε [28] and 10-fold cross validation[5]. Both the approaches calculate the performance parameters for classification and compare with the existing state-of-art methods.
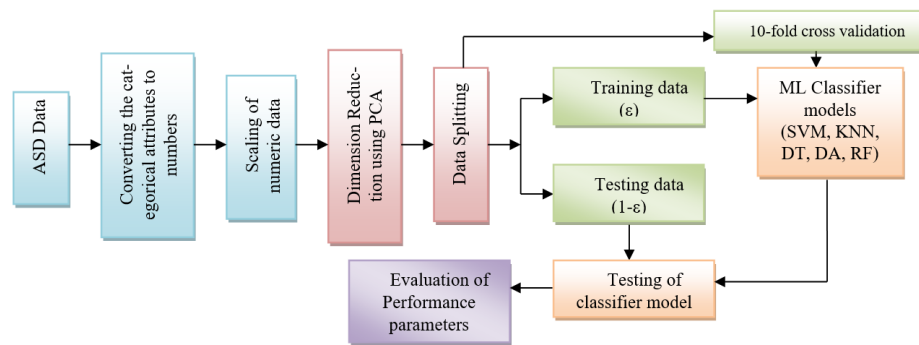

Fig. 1. Work Flow of proposed Method

4.1 Preprocessing

Numeric transformation of categorical data: The original toddler dataset contains the categorical input attributes like "Ethnicity", "Sex", "Jaundice", "Family_mem_ASD", "Who_completed_the_test" and the categorical output "Class/ASD Traits". It is very much essential to convert these categorical data into numeric values before feeding to the classifier. Among the categorical attributes, "Ethnicity" is assigned with dummy variable transformation where for a particular ethnicity the numeric value is assigned as 1 and under the same case other values are assigned as 0. Other input attributes like "Sex", "Jaundice", "Family_mem_ASD", and "Who_completed_the_test" are assigned the numbers based on the frequency of repetitions in the table. The output class "Class/ASD Traits" is assigned logical values i.e. 1 is assigned for "yes" and 0 is assigned for "no". Table.1 shows numeric

transformation rule for all categorical attributes. Table.2 represents the sample set of numeric data after transformation where the number of input attributes increased from 17 to 27 through numeric conversion rules.

Table 1. The numeric transformation of categorical data

Sex		Ethnicity		Jaundice		Family_mem_ with_ASD		Who_completed_the_test		Class/ASD Traits	
String	*Value*	*String*	*Value (element specific)*	*String*	*Value*	*String*	*Value*	*String*	*Value*	*String*	*Value*
'f'	319	'All'	1	'Yes'	288	'Yes'	170	'Family member'	1018	'yes'	1
'm'	735	-	-	'No'	766	'No'	884	'Self'	4	'no'	0
-	-	-	-	-	-	-	-	'Others'	3	-	-
-	-	-	-	-	-	-	-	'Health care'	29	-	-

Table 2. Sample of numeric data set after transformation

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	Age_Mons	Qchat-10-Score	Sex	Ethnicity											Jaundice	Family_mem_ with_ASD	Who_completed_ the_test	Class/ASD Traits
													'Hispanic'	*'Latino'*	*'NativeIndian'*	*'Others'*	*'Pacifica'*	*'WhiteEuropean'*	*'asian'*	*'black'*	*'middleeastern'*	*'mixed'*	*'southasian'*				
0	0	0	0	0	0	1	1	0	1	28	3	319	0	0	0	0	0	0	0	0	1	0	0	288	884	1018	0
1	1	0	0	0	1	1	0	0	0	36	4	735	0	0	0	0	0	1	0	0	0	0	0	288	884	1018	1
1	0	0	0	0	0	1	1	0	1	36	4	735	0	0	0	0	0	0	0	0	1	0	0	288	884	1018	1
1	1	1	1	1	1	1	1	1	1	24	10	735	1	0	0	0	0	0	0	0	0	0	0	766	884	1018	1
1	1	0	1	1	1	1	1	1	1	20	9	319	0	0	0	0	0	1	0	0	0	0	0	766	170	1018	1

4.2 Scaling of numeric data

After numeric transformation of original dataset it is seen that some transferred attributes like "Sex", "Jaundice", "Family_mem_with_ASD", and "Who_completed_the_test" have high numeric values and they exceed the range of output class. Hence they are standardized by mean and standard deviation approach using Eq. (1) [32].

$$x_{new} = \frac{x - \mu}{\sigma} \tag{1}$$

Where, x_new represents the standardized value of the attribute, x represents the original value, μ is the mean of all values of an attribute and σ is standard deviation of all values of an attribute. Table.3 represents the sample of numeric dataset after the process of standardization.

Table 3. Sample of numeric data set after standardization

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	Age_Mons	Q chat-10-Score	Sex	'Hispanic'	'Latino'	'Native Indian'	'Others'	'Pacifica'	'White European'	'asian'	'black'	'midll eeastern'	'mixed'	'South asian'	Jaundice	Famly_mem_with ASD	Who_completed_the test	Class/ASD Traits
																		Ethincity									
0	0	0	0	0	0	1	1	0	1	0	-0761	-1.517	0	0	0	0	0	0	0	0	1	0	0	-1.630	0.438	0.187	0
1	1	0	0	0	1	0	0	0	1	1	-0.417	0.658	0	0	0	0	0	1	0	0	0	0	0	-1.630	0.438	0.187	1
1	0	0	0	0	1	1	0	1	1	1	-0.417	0.658	0	0	0	0	0	0	0	0	1	0	0	-1.630	0.438	0.187	1
1	1	1	1	1	1	1	1	1	1	0	1.646	0.658	1	0	0	0	0	0	0	0	0	0	0	0.612	0.438	0.187	1
1	1	0	1	1	1	1	1	1	1	-1	1.302	-1.517	0	0	0	0	0	1	0	0	0	0	0	0.612	-2.279	0.187	1

Dimension reduction using PCA: In previous steps it is noticed that the total number of input attributes exceed the original number of attributes. Hence it is essential to reduce the dimension of data with little affecting the output of the system. For this purpose, in this paper, Principal Component Analysis (PCA)[33] is used.

PCA is the most popular linear dimensionality reduction technique used for reduction of dimensions of input data to a lower linear subspace. PCA is used to construct lower-dimensional representation of the input data, which describes as much variance in the data as possible. In this method, we have applied the standardized input having 27 attributes to PCA for its dimension reduction and only six principle components with variances of 20.8%, 12.5%, 12.05%, 10.81%, 9.90%, and 9.16% are considered as the transferred features which affect most to the output class. The Table.4 represents a sample table of principle components with the output.

Table 4. Samlpe table describing only the principle components and output class.

PC1	PC2	PC3	PC4	PC5	PC6	ASD Traits
-1.399	-0.917	-1.061	0.117	0.653	-1.616	0
-0.236	1.111	-0.869	0.492	1.337	-0.876	1
-0.508	1.023	-0.745	0.454	1.423	-0.845	1
2.223	-0.494	0.570	0.262	0.042	0.768	1
1.497	-1.013	-1.064	-0.773	-2.648	-0.101	1

4.3 Data splitting

The pre-processed data obtained from PCA is partitioned into two sets: training and testing data set through two methods.

In the first approach the complete data is converted into training and testing samples through a training parameter ε with two values 0.8 and 0.7 respectively. To prepare an effective trained model in ML, more number of samples is required for training. Hence for the said purpose, training parameters of 0.7 and 0.8 are set. The ML models are trained by training data set and the same classifiers are tested with the testing samples through testing parameter (1- ε).

In the second approach 10-fold cross validation procedure is used for training the ML models. In this model, the complete data is splitted into ten numbers of groups. The model is trained on nine groups and gets tested on remaining one group, thereby followed by ten times repetition of the procedure by randomly partitioning the training dataset. During the entire process of data splitting, random shuffling is surely to be performed so that each class should be available in each partition.

4.4 Machine learning models

Decision tree: DT[21] is a supervised learning algorithm for solving statistical classification and regression problems. In this paper, the DT creates a training model which is used for predicting target variables (YES/NO). The algorithm partitions data into subsets containing instances with similar values.

To build the DT, we have to evaluate two types of entropy utilizing the dataset as follows:

a) Entropy evaluation with single attribute is mathematically given by Eq. (2),

$$E(T) = \sum_{i=1}^{c} - p_i \log_2 p_i \tag{2}$$

Where, E is entropy, T is output class, pi is the probability of ith class and c is the number of output class.

b) Entropy evaluation with two attributes is mathematically given by Eq. (3),

$$E(T, X) = \sum_{c \in X} P(c) E(c_1, c_2) \tag{3}$$

Where T is the output class, X is one of the inputs, P(c) is probability of output for a particular value of input, c_1 is the number of counts for class1 with respect to input X_1 and c_2 is the number of counts for class1 with respect to input X_1.

c) Repeat step a) and step b) for each input and calculate the total sum of (2) and total sum of (3) to get the information gain given by the formula in Eq. (4),

$$G(T, X) = E(T) - E(T, X) \tag{4}$$

d) Continue the same process till the entropy reaches to 0. At this point the leaf node is found where all the data are classified.

Discriminant analysis: DA[19] finds a set of prediction equations that is based on independent attributes for classifying individuals into groups. Two possible objectives in a DA: To find a predictive equation for classifying new individuals and to interpret the predictive equation for understanding the relationships between the variables. In the model for DA, using multivariate normal distribution each class (Y) generates data (X).

The basic purpose is minimization of expected classification cost as shown in Eq. (5) which is given by,

$$\hat{y} = \arg\min_{y=1,2,3...N} \sum_{n=1}^{N} \hat{P}(n \mid x) C(y \mid n) \tag{5}$$

Where $\hat{y}$ is predicted classification, N is total number of classes, $\hat{P}(n \mid x)$ is posterior probability of class n for observation x as well as $C(y \mid n)$ is cost of classifying an observation as y for its true class tobe n.

K-Nearest neighbor: kNN[29][30] classification model is implemented in numerous areas like data mining, prediction, pattern recognition, as well as other areas of applied sciences. The unclassified data is identified thereby testing the closeness of k in the dataset. By determining Euclidean distance, the nearest neighbour is determined. Let k is total number of samples, m_i is i^{th} input, n_i is output for i^{th} input then, Eq. (6) represents mathematically the Euclidean distance,

$$d(m, n) = \sqrt{\sum_{i=1}^{k} (m_i - n_i)^2} \tag{6}$$

Support vector machine: SVM[15]is utilized for regression as well as classification problems with high accuracy. It utilizes a hyperplane which classifies the data samples in a N-dimensional space. SVM focusses on maximizing the distance between data samples of both classes.

$$\min_{w,b,\xi} \{ \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \}, \text{ constraints to}$$

$$n_i (w^T \phi(m_i) + b) \geq 1 - \xi_i, \xi \geq 0,$$

where, (m_i , n_i), i=1,…..,l is an instance-label pair,

$m_i \in \Re^p$, $n \in \{1,-1\}^l$, ϕ is the mapping function,

C > 0 is the penalty parameter.
The Kernel function,

$$K(m_i, m_j) \equiv \phi(m_i)^T \phi(m_j) \qquad (7)$$

Random forest: RF classification algorithm[14] is based upon numerous classification trees in which every tree yields a classification result. It's a collection of decision trees.

It can be represented by,

h(m, θ_k), for k=1,….., K.

where, θk is the identically distributed random input samples.
The general error for enough number of trees is mathematically given by Eq. (8),

$$e \leq \frac{c(1-o^2)}{o^2} \qquad (8)$$

Where, c resembles correlation between two trees as well as o is strength metric.

4.5   Evaluation of performance parameters

To outline the performance of implemented ML algorithms, five evaluation parameters viz. accuracy(Acc), sensitivity(SN), and specificity (SP), F-measure(F1-Score) as well as Area under Curve (AUC) are utilized for classifying ASD and no ASD test instances over the Toddler data set. Eq. (9-13) represent the expressions for performance parameters. The confusion matrix describes the performance of the classification models up on a set of test data with known true values [5].

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \qquad (9)$$

$$SN = \frac{TN}{TP + FN} \qquad (10)$$

$$SP = \frac{TN}{TN + FP} \qquad (11)$$

$$F1 - Score = \frac{2 \times TP}{(2 \times TP + FP + FN)} \qquad (12)$$

$$AUC = \frac{(1 + \dfrac{TP}{TP + FN} - \dfrac{FP}{FP + TN})}{2}$$

$\hspace{12cm}$ (13)

5 Result and Discussion

The proposed work is implemented in MATLAB 2016a environment utilizing a PC with Intel Corei3 1.99GHz processor and 4GB RAM. In this paper followed by preprocessing, training of the data set is done in two ways:

a) Through the training parameter (ε) and testing parameter (1-ε).
b) Through 10-fold cross validation.

Two values of training parameters(ε_1 = 0.8 and ε_2 = 0.7) are set[28] and for each ε, the values of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) are found from the confusion matrix. For **ε** = 0.8, the number of test instances in the data set is 20 percent of the total number of toddler instances i.e. 210 whereas for **ε** = 0.7, the number of test instances is 30 percent of the total number of toddler instances i.e. 316. The peformance parameters are evaluated. Table.5 shows the statistics of performance parameters through training and testing parameters.

Table 5. Performance parameters through training and testing samples

ML Algorithms	ε	Acc	SN	SP	F1-score	AUC
SVM	0.8	0.995	1.000	0.993	0.993	0.994
	0.7	0.997	1.000	0.995	0.995	0.996
KNN	0.8	0.981	0.968	0.986	0.968	0.984
	0.7	0.987	0.989	0.987	0.978	0.987
DT	0.8	1.000	1.000	1.000	1.000	1.000
	0.7	0.991	0.979	0.995	0.984	0.993
DA	0.8	0.971	0.917	1.000	0.957	0.986
	0.7	0.968	0.917	1.000	0.957	0.984
RF	0.8	0.986	0.962	1.000	0.981	0.993
	0.7	0.984	0.979	0.986	0.974	0.985

From Table.5 it is observed that for **ε** = 0.8, DT outperformed other classification models for classifying Toddler ASD class with the maximum performance. The maximum performance of DT in terms of accuracy is being followed by SVM with an accuracy of 0.997 for **ε** = 0.7. Further SVM yielded maximum sensitivity for **ε** = 0.8 as well as **ε** = 0.7 as FN rate came out to be zero resembling there was not a single instance who got wrongly classified as no ASD class. DA produced maximum specificity for **ε** = 0.8 as well as **ε** = 0.7 because of zero FP rate resembling presence of no instance who got wrongly classified as ASD class, while RF gave maximum specificity for **ε** = 0.8 with no instance being wrongly classified as ASD class. Following DT, SVM resulted in F1-score of 0.995 and 0.993 **ε** = 0.7 and 0.8 respectively. The performance of SVM

is next to the performance of DT in case of AUC. Overall from the above analysis it is concluded that DT due to its maximum performance is perferable for classification of standardized Toddler ASD data.

Table.6 represents the result from 10-fold cross validation. The final result is calculated from summation of results from each fold upto tenth fold.

Table 6. Performance parameters through 10-fold cross validation

ML Algorithms	Acc	SN	SP	F1	AUC
KNN	0.986	0.981	0.989	0.978	0.987
SVM	0.999	1.000	0.998	0.998	0.998
DT	0.992	0.984	0.995	0.987	0.994
DA	0.971	0.918	0.998	0.955	0.985
RF	0.996	0.993	0.997	0.993	0.996

From Table.7 it can be concluded that SVM performed effectively for Toddler ASD classification with a performance of more than 99 percent. At the same time DA also yielded a specificity rate of 99.86 percent. For a comparison, the performance of DA classifier is the lowest among all classifiers but still acceptable. It is due to the fact of higher FN rate where 26 number of instances are wrongly classified as no ASD class. Similarly specificity rate of KNN is the lowest among all classifiers due to the fact of higher FP rate where 12 number of instances are wrongly classified as ASD class.

Table.7 represents the results of various state of art methods in adolescent, adult and child ASD datasets and compared with the results found in this paper for Toddler ASD dataset.

The performances of ML algorithms can also be visualized through the Receiver operating characteristics (ROC) curves from Fig.2 using 10-fold cross validation method for classification. From the Fig.2 it is observed that DA gives an Area under curve(AUC) value of 0.9851, KNN yields an AUC of 0.9879, DT gives AUC value of 0.9941, the AUC for random forest is 0.9963 whare as SVM yields the better result as compared to other methods i.e. AUC of 0.9988.
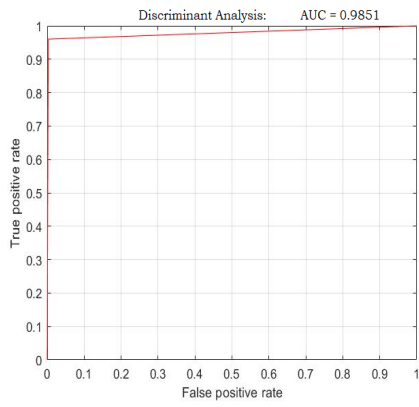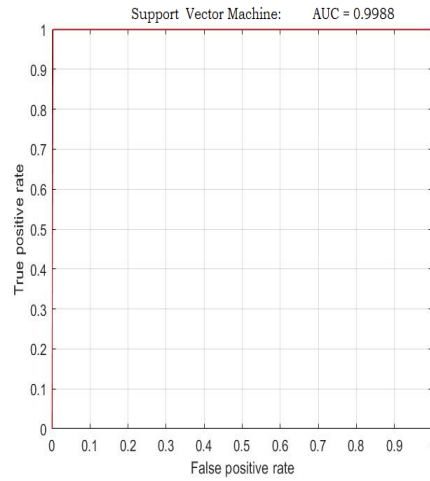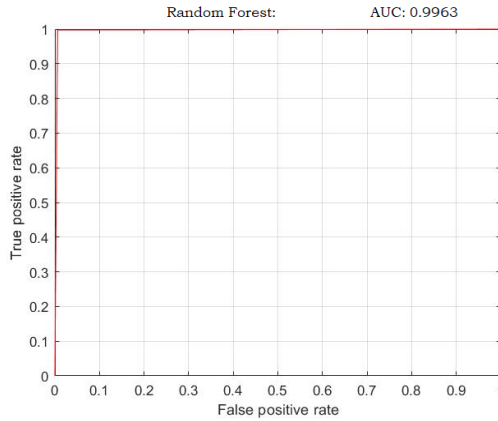
(a) Discriminant Analysis
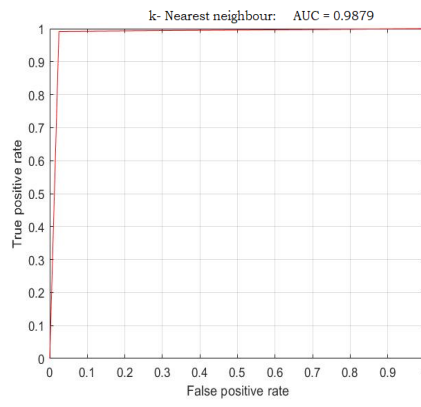
(b) K-nearest neighbour
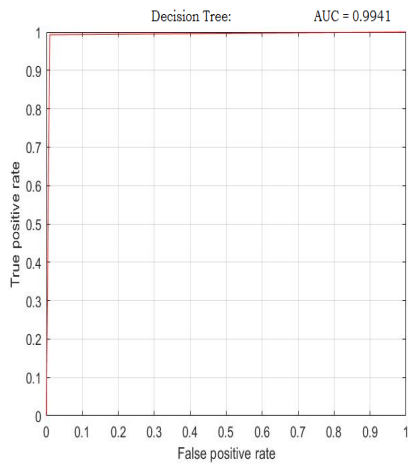
(c) Decision Tree

(d) Support Vector Machine

(e) Random Forest

Fig. 2. ROC curves of proposed machine learning algorithms using 10-fold cross validation Method

Table 7. Performance comparsion of the proposed approach with other approaches.

Author	Individual's ASD Dataset	Method used	Acc	SN	SP	F1	AUC
[24]	Adult	LR	0.998	0.999	0.997	-	-
[25]	Child	FURIA	0.913	0.914	0.880	-	-
		JRIP	0.882	0.882	0.928	-	-
[28]	Adult	RF (Complete & Missing data) $\alpha = 70$ and 80 percent	1.0	1.0	1.0	1.0	1.0
	Adolescent	RF (Complete & Missing data) $\alpha = 70$ and 80 percent	1.0	1.0	1.0	1.0	1.0
[31]	Adult	CNN	0.995	0.993	1.0	-	-
Proposed Approach	Toddler (training & testing data)	SVM ($\varepsilon = 0.7$)	0.997	1.000	0.995	0.995	0.996
		DT ($\varepsilon = 0.8$)	1.0	1.0	1.0	1.0	1.0
	Toddler (10-fold validation)	SVM	0.999	1.0	0.998	0.998	0.998

6 Conclusion

The proposed work emphasizes on early ASD detection in toddlers. In this work, using PCA, there is a dimension reduction based on contribution of minimal benefit in the number of attributes followed by use of ML classifier models to detect ASD in toddler dada set. The evaluation parameters yielded clinically acceptable results using ML classifiers. The acceptance level of standard performance is 80 percent and in the analysis carried out in this investigation, all the evaluation parameters produced more than 90 percent performance. Generally, in most of the research analysis, toddler data

set is dropped because of its unbalanced nature making the detection of disorder difficult in that category. In this study, toddler data set is analyzed thereby making it successful to detect ASD in toddlers. The study will be further enhanced in future thereby extending the investigation on rest category of individuals in addition with toddler one.

7   References

[1] Mohanty1 A S, Patra2 D K C and Akshya Kumar Sahoo3 a Brief Instigation to Autism: Review Article Int. J. Res. Sci. Innov. IV 29–32

[2] Sun X and Allison C 2010 A review of the prevalence of Autism Spectrum Disorder in Asia Res. Autism Spectr. Disord. 4 156–67

[3] Al Backer N B 2015 Developmental regression in autism spectrum disorder. Sudan. J. Paediatr. 15 21–6

[4] Johnson C P and Myers S M 2007 Identification and Evaluation of Children with Autism Spectrum Disorders Pediatrics 120 1183–215

[5] Thabtah F 2019 Machine learning in autistic spectrum disorder behavioral research: A review and ways forward Informatics Heal. Soc. Care 44 278–97. https://doi.org/10.1080/17538157.2017.1399132

[6] Puttinaovarat S, Pruitikanee S, Kongcharoen J and Horkaew P 2021 Machine Learning Based Emergency Patient Classification System Int. J. Online Biomed. Eng. 17 133. https://doi.org/10.3991/ijoe.v17i05.22341

[7] Akundi S H, R S and PM M 2020 Big Data Analytics in Healthcare using Machine Learning Algorithms: A Comparative Study Int. J. Online Biomed. Eng. 16 19. https://doi.org/10.3991/ijoe.v16i13.18609

[8] Anon https://www.autismresearchcentre.com/. Autism Research Centre - University of Cambridge

[9] Thabtah F 2018 https://www.kaggle.com/fabdelja/autism-screening-for-toddlers/version/1

[10] Christensen D L, Baio J, Braun K V N, Bilder D, Charles J, Constantino J N, Daniels J, Durkin M S, Fitzgerald R T, Kurzius-Spencer M, Lee L-C, Pettygrove S, Robinson C, Schulz E, Wells C, Wingate M S, Zahorodny W and Yeargin-Allsopp M 2016 Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012 MMWR. Surveill. Summ. 65 1–23. https://doi.org/10.15585/mmwr.ss6503a1

[11] Nielsen J A, Zielinski B A, Fletcher P T, Alexander A L, Lange N, Bigler E D, Lainhart J E and Anderson J S 2013 Multisite functional connectivity MRI classification of autism: ABIDE results Front. Hum. Neurosci. 7. https://doi.org/10.3389/fnhum.2013.00599

[12] Leroy, G., Irmscher, A., Charlop-Christy, M., Kuriakose, S., Pishori, A., Wurzman, L., Stephanie B 2006 Data mining techniques to study therapy success. Int. Conf. Data Min. 1–7

[13] Langseth H and Nielsen T D 2006 Classification using Hierarchical Naïve Bayes models Mach. Learn. 63 135–59. https://doi.org/10.1007/s10994-006-6136-2

[14] Breiman L. 2001 Random forests Mach Learn.

[15] Cortes C and Vapnik V 1995 Support-vector networks Mach. Learn. 20 273–97

[16] Fischbach G D and Lord C 2010 The Simons Simplex Collection: A Resource for Identification of Autism Genetic Risk Factors Neuron 68 192–5. https://doi.org/10.1016/j.neuron.2010.10.006

[17] la Iglesia-Vaya M de, Molina-Mateo J, Jose M, S. A and Marti-Bonmati L 2013 Brain Connections – Resting State fMRI Functional Connectivity Novel Frontiers of Advanced Neuroimaging (InTech). https://doi.org/10.5772/52273

[18] Hanchuan Peng, Fuhui Long and Ding C 2005 Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy IEEE Trans. Pattern Anal. Mach. Intell. 27 1226–38. https://doi.org/10.1109/tpami.2005.159

[19] Bone D, Goodwin M S, Black M P, Lee C-C, Audhkhasi K and Narayanan S 2015 Applying Machine Learning to Facilitate Autism Diagnostics: Pitfalls and Promises J. Autism Dev. Disord. 45 1121–36. https://doi.org/10.1007/s10803-014-2268-6

[20] D'Angelo G M, Rao D and Gu C C 2009 Combining least absolute shrinkage and selection operator (LASSO) and principal-components analysis for detection of gene-gene interactions in genome-wide association studies BMC Proc. 3 S62. https://doi.org/10.1186/1753-6561-3-s7-s62

[21] Quinlan J R 1986 Induction of decision trees Mach. Learn. 1 81–106

[22] Barros A J D and Hirakata V N 2003 Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. BMC Med. Res. Methodol. 3 21. https://doi.org/10.1186/1471-2288-3-21

[23] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Müller A, Nothman J, Louppe G, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M and Duchesnay É 2012 Scikit-learn: Machine Learning in Python. https://doi.org/10.1145/2786984.2786995

[24] Thabtah F 2019 An accessible and efficient autism screening method for behavioural data and predictive analyses Health Informatics J. 25 1739–55. https://doi.org/10.1177/1460458218796636

[25] Al-diabat M 2018 Fuzzy Data Mining for Autism Classification of Children Int. J. Adv. Comput. Sci. Appl. 9

[26] W. CW 1995 Fast effective rule induction Proceedings of the 12th International Conference on Machine Learning pp 115–123

[27] Cendrowska J 1987 PRISM: An algorithm for inducing modular rules Int. J. Man. Mach. Stud. 27 349–70. https://doi.org/10.1016/s0020-7373(87)80003-2

[28] Erkan U and Thanh D N H 2020 Autism Spectrum Disorder Detection with Machine Learning Methods Curr. Psychiatry Res. Rev. 15 297–308. https://doi.org/10.2174/2666082215666191111121115

[29] NS. A 1992 An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat 43(3) 175–85

[30] Badr H 2021 Building Recommendation Systems Using the Algorithms KNN and SVD Int. J. Recent Contrib. from Eng. Sci. IT 9 71–80

[31] Raj S and Masood S 2020 Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques Procedia Comput. Sci. 167 994–1004. https://doi.org/10.1016/j.procs.2020.03.399

[32] Grus J 2015 Data Science from Scratch. Sebastopol, CA: O'Reilly. 99, 100

[33] Hotelling H 1933 Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. 24 417–41

8 Authors

Ashima Sindhu Mohanty is pursuing her PhD in the Department of Electronics, SUIIT, Sambalpur University, Odisha, India. Her research interest includes Biomedical Data Analysis and Artificial Intelligence.

Krishna Chandra Patra has completed his Ph.D. in Electronics from Delhi University. He is presently working as an Assistant Professor in the Department of Electronics, SUIIT, Sambalpur University, Odisha, India. His Area of research is Advanced Communication Techniques and Photonics.

Priyadarsan Parida has completed his Ph.D. in Electronics and Telecommunication Engineering from Veer Surendra Sai University of Technology, Burla, Odisha. He is presently working as Associate Professor in the Department of Electronics and Communication Engineering, GIET University, Odisha, India. His area of research is Signal Processing and Computer Vision**.** Apart from it he is serving as reviewer for various international journals of repute. priyadarsanparida@giet.edu