Measurement: Sensors 27 (2023) 100774

Contents lists available at ScienceDirect

Measurement: Sensors

journal homepage: www.sciencedirect.com/journal/measurement-sensors

The classification of autism spectrum disorder by machine learning methods on multiple datasets for four age groups

Dhuha Dheyaa Khudhur [a,*], Saja Dheyaa Khudhur [b]

[a] Ministry of Education Iraqi Directorate of Education Baghdad Karkh III, Baghdad, Iraq
[b] Computer Engineering Department, Computer Engineering, University of Technology-Iraq, Baghdad, Iraq

ARTICLE INFO

Keywords:
Autism spectrum disorder (ASD)
Machine learning
Logistic regression (LR)
K- Nearest neighbor (K-NN)
Decision tree (DT)
Naïve bayes (NB)
Support vector machine (SVM)
Random forest (RF)

ABSTRACT

The world has seen the advent of numerous illnesses that cannot be medically recognized, such as Autism Spectrum Disorder (ASD). It affects several behavioral domains, including social and linguistic competence and stereotyped and repetitive actions. This illness is a serious neurodevelopmental disorder. Since many other mental illnesses have strikingly similar symptoms to those of ASD, diagnosing ASD can be difficult and time-consuming. Early diagnosis based on different health and physiological characteristics seems feasible with the rising usage of machine learning-based models in predicting many human diseases. This study aims to create a classification model that can predict the likelihood of ASD with the greatest degree of precision. To investigate the potential for predicting and analyzing ASD traits in the Toddler, Child, Adolescent, and adult age groups, we used several supervised Machine Learning (ML) models. These include Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbor (K-NN), Nave Bayes (NB), Logistic Regression (LR), and Random Forest (RF). Four publicly available, distinctive non-clinical ASD screening datasets from Kaggle and the UCI machine learning library are used to test these models. The first dataset includes 1054 instances and 19 toddler-related features. The remaining ones consist of 21 traits and 292, 104, and 704 cases involving children, adolescents, and adults, respectively. After implementing different ML approaches over the pre-processing datasets, the results showed that the DT, LR, and RF classifiers are the dominant models. These dominated models achieve the highest prediction accuracy, among other studied models, of about 100% for all the utilized datasets.

1. Introduction

Recently, the world has witnessed the emergence of many diseases that cannot be diagnosed clinically, including Autism Spectrum Disorder (ASD). This disease is a major neurodevelopmental disorder and impacts various behavioral areas, including social and communicative ability and stereotyped and repetitive behaviors [1]. The disease's cause is unknown; however, it is thought to be connected to biological variables such as genetic abnormalities, brain inflammation, and improper prenatal circumstances. The rapid rise in the number of children identified with ASD emphasizes the need for more study on these inhabitants. In especially, proper clinical procedures are essential [2,3].

A wide range of symptoms characterizes ASD. It impacts how individuals communicate with one another and how they act and educate [4,5]. The symptoms and indications begin at an early age; according to WHO statistics, (ASD) was diagnosed by 0.63% of very young children and continues to expand to adolescents and adults [6]. A person with ASD may suffer from mental problems such as anxiety and misunderstanding, which can impair their capacity to function adequately during various periods of life [7]. As a result, early diagnosis and treatment must always be emphasized [8,9]. One of the most noticeable signs of ASD is the affected person's conduct, which might include unexpected and unsafe behavior inspired by films and cartoons [10,11].

On the other hand, ASD is a developmental disorder of the human brain that affects a person's entire life for a lifetime. It's worth noting that environmental and genetic variables may play a role in developing this condition [9]. Although it is not feasible to entirely heal patients suffering from this condition, its consequences can be mitigated for a period if the signs are discovered early. Scientists have failed to identify the precise origins of ASD, presuming that human genes are at blame for this [11]. A person with ASD is often unable to engage with others and communicate with them; nonetheless, there are particular social interaction and communication issues such as [12,13].

[*] Corresponding author.
E-mail addresses: dhtoota@gmail.com (D.D. Khudhur), Saja.d.khudhur@uotechnology.edu.iq (S.D. Khudhur).

https://doi.org/10.1016/j.measen.2023.100774
Received 29 January 2023; Received in revised form 31 March 2023; Accepted 4 May 2023
Available online 5 May 2023

- Lack of pain sensitivity
- Inability to create correct eye contact
- Inability to respond appropriately to sound
- Lack of desire to cuddle
- Inability to convey gestures
- Lack of engagement with people
- Inappropriate object attachment
- Desire to live alone

Moreover, with the increased use of ML-based models in predicting numerous human diseases, early diagnosis based on various health and physiological parameters appears conceivable. This feature has fuelled our interest in predicting ASD, diagnosing it, analyzing it, and finding ways to treat it [14,15]. Detection of ASD is challenging since there are various other mental diseases with symptoms that are remarkably like those of ASD, making this a challenging undertaking. Moreover, ML is the most common area in finding functional patterns for treating autism patients by using different methods to detect it and find out if the person is affected or not [16]. In addition, ML has been used to detect many diseases and find appropriate solutions for them [15,17].

The contents of this paper are structured as follows: Section 2 includes a review of current research in which specific ASD detection models have been established. The datasets utilized in this study are described in Section 3In addition, Section 4 provides a comprehensive description of all techniques used in this study. In Section 5, the findings of numerous experiments are shown and discussed, and the conclusion is offered in Section 6.

2. Literature survey

ML approach has become one of the most common areas in finding functional patterns to detect many diseases and find appropriate solutions [15,17]. Furthermore, treating autism patients uses different methods to detect and identify patients as affected or not. Azian A. et al. [18] have suggested three ways: Selection Operator (LASSO), Least Absolute Shrinkage, and Chi-square, to test ML approaches that could be used for regression and classification. These techniques are K-Nearest Neighbors (K-NN), Random Forest (RF), and Logistic Regression (LR). The experimental results showed the highest accuracy of the LR model, with 97.541%, among other techniques utilized. Their approach depends on 13 features selected based on the Chi-square selection model.

In [19], the authors suggested a strategy for detecting autism using optimal behavioral sets. They used a binary firefly feature selection wrapper based on swarm intelligence to screen for ASD. The analyzed and categorized dataset contains 21 features obtained from the ML repository. The authors discovered among of 21 characteristics that are studied in the ASD dataset that, only ten characteristics can be classified to distinguish between ASD and non-ASD patients. Their experimental results proved that the system obtained an average accuracy of 92.12%–97.95% from the ASD datasets.

Koushik Ch., and Mir A. I. [20], suggested finding the best way to measure ASD among several measurements carried out in several classifiers, including Support Vector Machine (SVM) and Gaussian Radial Kernel. The obtained results showed the best and the highest accuracy with 95% Using the standard ASD dataset, which is available to all. The authors in Ref. [21] suggested various techniques and methods for detecting and identifying ASD. They used several approaches of ML, such as classifiers and some classifiers based on neural networks. They conducted a comprehensive test to determine the extent to which the proposed system detects ASD. The experiments included three datasets which are (children, adolescents, and adults). The experimental results showed that some of the ML classifiers outperformed the other studied classifiers in terms of accuracy when each was performed in preciseness, F-beta, and recall methods. Furthermore, all three datasets used SHAP approaches to analyze the critical aspects of ASD prediction.

Suman R., and Sarfaraz M. [22], made try to find if there are specific ways to predict the ASD risk. They used Logistic Regression, Nave Bayes, K-NN, Support Vector Machines, Convolutional Neural Networks, and Neural Networks to test the necessary criteria and to analyze and classify ASD. Three ASD datasets were used (children, adults, and adolescents) to test the proposed methods. There are 292 samples for 21 attributes of ASD patients to examine children, 740 samples for 21 attributes of adults, and 104 for 21 attributes of adolescents. After applying the ML methods and extracting the results, it was found that the proposed model obtained the highest accuracy rate in the CNN classifier with 98.30%, 99.53%, and 96.88% for testing and detecting ASD for children, adults, and adolescents, respectively. Nishat MM et al. [23] proposed a model based on ML algorithms to predict ASD and its psychological disorders that significantly affect the Individual's behavior in social life. They used the quadratic discriminant algorithm and linear analysis to detect and analyze the disease and find the appropriate treatment plan to reduce its severity. To build the ML model, they used the data of the University of California, Irvine (UCI) reservoir for analysis and discovery. They comprehensively evaluated the sensitivity, F1 score, accuracy, and Youden index. The experimental results showed the efficiency and effectiveness of the proposed model, as the Quadruple Analysis Algorithm (QDA) showed its high accuracy of 99.77% after adjusting the hyperparameters. Md. Mokhlesur R. et al. [24], the authors made use of ML methods and highlighted essential topics related to autism. In addition, they confirmed the need to identify the best traits of autism, enhance categorization, and maintain top precision. The researchers can create an ML system to produce promising results in detecting ASD by selecting the suitable important autistic traits and minimizing data dimensionality. Depending on the authors, several factors that affect accuracy must be addressed. For example, include an unbalanced and insignificant data set, faulty sampling procedures, and feature redundancy. In another research, Md Delowar H., and Muhammad A. k. et al. [25], tried to identify the most critical characteristics and automate the diagnostic procedure for ASD by utilizing current classification algorithms for better diagnosis. They tested ASD datasets from toddlers, children, teenagers, and adults. They evaluated and identified the best performing classifier among the latest classification techniques to discover the ASD of the above dataset. The results showed the efficiency and superiority of the multilayer perceptron (MLP) classifier, which achieved the highest accuracy of 100% for the four datasets. They also found that the best technique for the four ASD datasets was the "Relief F″ feature selection technique to rank the most important traits. Nurul A. et al. [26] proposed a model based on different ML techniques to identify and analyze the ASD classification. The dataset to be analyzed consists of 16 characteristics, including 703 autistic and non-autistic patients. ML methods were used to predict the state of ASD: support vector machine, K-NN, naïve Bayes (NB), J48, Bagging, and Stacking. These techniques were conducted in a simulated environment using the Waikato environment for knowledge analysis (WEKA) platform. The obtained results showed the accuracy, sensitivity, and superiority of some of the ML classifiers over the other studied classifiers, Stacking, J48, SVM, and Bagging, with an accuracy of 100% and a lower error rate. Al Banna et al. [27], the authors proposed an artificial intelligence-based sensor network system to monitor a patient's condition based on their facial expressions and emotions. The patient care committee is alerted in the event of an error in the patient's behavior. The authors also explained the effectiveness of the artificial intelligence-based model and its use in the Corona pandemic, in addition to the possibility of the system to help parents continue the mental development of their children. Uzma A. S. et al. [28], the authors proposed a wearable sensor based on multiple classification approaches to distinguish autistic gestures. The proposed system identifies, surveils, and categorizes the Individual's gesture. Bluetooth technology allowed these sensors to transmit and classify their data to the server. The authors tested the disease status of 10 affected children using datasets, which consisted of 24 features, and repeated each feature about ten times. Sensor data is categorized using K-NN, DT, neural network, and RF models based on time and frequency-domain

Table 1
Autism spectrum disorder screening datasets description.

No.	Dataset Name	Sources	Feature	No. of Feature	No. of Samples
1	ASD Dataset for Toddler	[29]	Continuous, binary, and categorical	19	1054
2	ASD Dataset for Children	[30]	Continuous, binary, and categorical	21	292
3	ASD Dataset for Adolescent	[31]	Continuous, binary, and categorical	21	104
4	ASD Dataset for Adult	[32]	Continuous, binary, and categorical	21	704

Table 2
List of the common features in all datasets.

No.	Feature Description
1–10	Based on ten screening questions (Q1-Q10 in Table 2)
11	Patient age
12	Sex
13	At birth, the patient had an issue with jaundice
14	Anyone in the family affected by developmental problems
15	Who is responsible for the experiment’s success?
16	The user’s country of residence
17	Is the user familiar with the screening application?
18	Screening test Kind
19	Screening result (Score by Q-chat-10)

Table 3
Ten screening questions to evaluate patients’ features.

ID	Description
Q1	When you call your child’s name, does S/he look at you? (Toddler); s/he detects minor noises while others don’t (child, Adolescent); s/he is always looking for patterns in things (Adult)
Q2	How quick is your ability to make eye contact with your child? (Toddler); s/he is usually inclined to pay attention to the large picture than the mini specifics (child, adult, Adolescent).
Q3	When your child wants something, sometimes a distant toy, does he point to you? (Toddler)Can he focus on other people’s conversations during a family gathering? (child, Adolescent); Is it easy for him to multitask? (Adult)
Q4	Does your child interest in sharing activities with you? (Toddler); can he quickly move between different activities? (child, Adolescent); can s/he get back to what he was doing quickly if something prevents him from following?
Q5	Is your child pretending to care for the toy he owns? (Toddler); in the lesson, s/he might not be able to keep up the conversation with his/her friends (child, Adolescent); when someone starts talking to me, it’s easy for me to read between the lines (Adult).
Q6	Is your child interested in searching with you for lost things? (Toddler); does s/he like to participate in the social chat enthusiastically? (child, Adolescent); when I talk, I can feel that an unwilling person to listen to me (Adult)
Q7	Your child can console any member of the family if s/he is upset? (Toddler); does your child assign the characters’ goals when he reads a story? (Child); when he was a very young child, did he like to play with other kids? (Adolescent); I find it challenging to know the motives that provoke the characters when I read a story (Adult).
Q8	What about the first words that your child uttered? Could you describe them? (Toddler); in preschool, s/he inclined to play fantasy games with friends? (Child); how does it feel to imagine if you were another someone (Adolescent); I feel passion when I get to know things of different kinds (such as plants, cars, animals, etc.) (Adult).
Q9	Does your child use simple cues such as goodbye? (Toddler); once you look at someone’s facial expressions, can you tell what they’re thinking? (Child); is it easy to see social events? (Adolescent); can I tell what is going on in a person’s mind just by looking at their facial expressions? (Adult).
Q10	Does your child always look at unnecessary things carefully? (Toddler); s/he has trouble making new acquaintances (Child, Adolescent); I have difficulties understanding people’s beliefs (Adult)

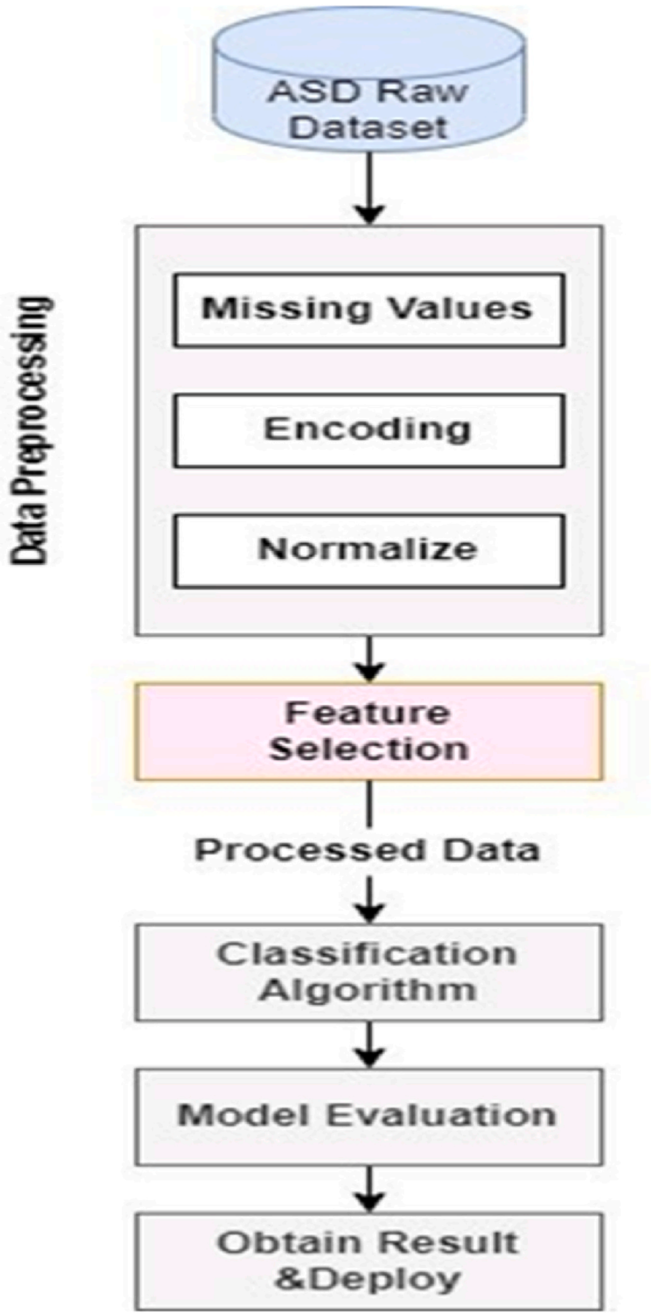

Fig. 1. The proposed system of ASD detection.

features. The obtained results showed the precision of the proposed system with 91% of the rest of the other classifications.

3. Dataset description

Four publicly available ASD screening datasets are used for this analysis. The adopted datasets are for toddler [29], children [30], adolescent [31] and adult [32]. All datasets are composed of 21 features with varying sizes of samples. The highest observation dataset among those is for toddlers, which composes (of 1054). A full overview of the datasets is illustrated in Table 1. The 19 common features in these datasets are utilized for forecasting, as shown in Table 2. Ten behavioral features (Q-Chat-10) and other individual factors help distinguish people with ASD in behavior science. These ten features are represented in the form of questions, as shown in Table 3.

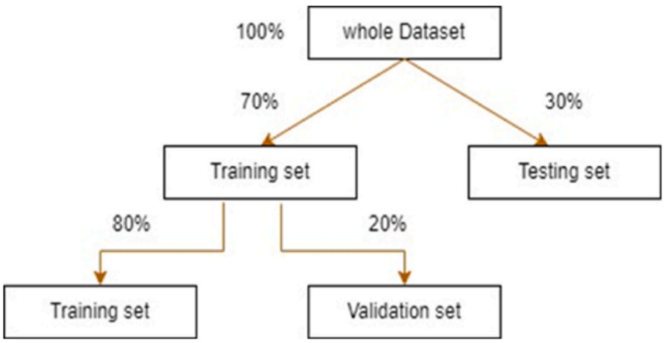

Fig. 2. Training and Testing sets.

4. Proposed system

Fig. 1 illustrates the proposed system, which consists of data pre-processing, feature engineering, model evaluation using specified models, results evaluation, and ASD prediction. The applied models are built using Python 3.9 with data analysis and ML libraries.

4.1. Data pre-processing

It is one of the techniques that can convert raw information into a format that can be used and understood. Sometimes real-world data is incomplete and unformatted because it can have many nulls and errors. Several data pre-processing methods are utilized, such as handling the missed value, data normalization, and encoding.

4.2. Training and testing mode

In the training and testing process of all the adopted ML models, the entire datasets have been split into two main sections, as shown in Fig. 2, using 70:30 ratios. Training data has been divided into two sections for cross-validation reasons. With 80:20 ratios, as training and validation set.

A. Support Vector Machine (SVM)

It is one of the ML algorithms that can be applied for classification and regression issues. It is usually used in classification issues for its effectiveness and ability to obtain excellent accuracy in most data. The idea of this algorithm is based on finding the best separation level between the classes by making the margin as large as possible. The main idea behind the SVM model is the separation of the classes based on determining the decision boundaries [33].

B. Logistic Regression (LR)

LR is one of the ML approaches. It is used to analyze binary dependent variables and to predict the probability of an event occurring for the values of variables that can be related to or explained by that event. The output value is either 0 or 1. LR is a multi-response used in the case of nominal or normal variables. It can be expressed by a sigmoidal function [34].

Table 4
Evaluation results for ASD screening data for Toddler.

Classifier	Class	precision	recall	F1-score
DT	ASD, Not ASD traits	1 1	1 1	1 1
K-NN	ASD, Not ASD traits	0.9076 0.9794	0.9516 0.9597	0.9291 0.9694
LR	ASD, Not ASD traits	1 1	1 1	1 1
NB	ASD, Not ASD traits	0.9833 0.9801	0.9516 0.9932	0.9672 0.9866
RF	ASD, Not ASD traits	1 1	1 1	1 1
SVM	ASD, Not ASD traits	0.9833 0.9801	0.9516 0.9932	0.9672 0.9866

C. Decision Tree (DT)

It is one of the ML algorithms which can be used in classification and regression problems. However, it is most commonly used to solve classification difficulties. Internal nodes contain dataset attributes. In this tree-structured classifier, the branches are represented as decision rules, and each leaf node represents the conclusion. Moreover, DT asks a question and divides the tree into subtrees based on the answer (Yes/No) [35].

D. Naïve Bayes (NB)

This approach is one of the supervised ML and is based on the principle of probability. This approach is characterized by processing speed and forecasting efficiency. NB is based on the statistical concept, which calculates the probability of a specific outcome and shows less training time compared to the SVM and ME models [36].

E. K- Nearest Neighbor (K-NN)

K-NN is one of the ML approaches. This model is the simplest, does not require complex mathematical equations, and is used in regression and classification issues. This approach only needs to know how to calculate the distance between the data and the presence of similar data nearby. The 'K' component denotes the number of seed point that is to be chosen. It should be carefully picked to limit the chance of error [37].

F. Random Forest (RF)

The RF approach is the foundation of The Decision Tree (DT). DTs use questions and responses to limit the range until the confidence level is high enough to produce a single forecast. Individual DT predictions may not be accurate. However, combining many DTs into a single model improves the accuracy of the forecast. The merging of several DTs refers to RF, one of the ML models used in classification and regression [38].

5. Result and discussion

This section illustrates and discusses the evaluation metrics and model evaluation results for the selected dataset. The evaluation results depend on how the models are trained and measured in terms of precision, recall, and f1-score. by using the classification report and confusion matrix.

The performance evaluation of a predictive model is essential for determining how effectively a model performs in achieving a goal. Performance assessment measures are utilized on the test dataset to assess the classification model's efficacy and performance. It's critical to use the right metrics to assess performance, such as precision, accuracy, confusion matrix, and recall. These metrics depend on the True Positive (TP), True Negative (TN), False Positive (FP) False Negative (FN) as the main parameters. The performance metrics are calculated using the formulas below.

The percentage of accurately detected positives to all expected positives is known as precision. In terms of mathematics, it is computed using (1):

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

FP and TP refer to the number of incorrect and correct positive predictions.

Table 5
Evaluation results for ASD screening data for Children.

Classifier	Class	precision	recall	F1-score
DT	ASD, Not ASD traits	1	1	1
		1	1	1
K-NN	ASD, Not ASD traits	0.9545	1	0.9767
		1	0.9555	0.9772
LR	ASD, Not ASD traits	1	1	1
		1	1	1
NB	ASD, Not ASD traits	0.8936	1	0.9438
		1	0.8888	0.9411
RF	ASD, Not ASD traits	1	1	1
		1	1	1
SVM	ASD, Not ASD traits	1	1	1
		1	1	1

Table 6
Evaluation results for ASD screening data for Adolescent.

Classifier	Class	Precision	recall	F1-score
DT	ASD, Not ASD traits	1	1	1
		1	1	1
K-NN	ASD, Not ASD traits	1	0.8571	0.9230
		0.9615	1	0.9803
LR	ASD, Not ASD traits	1	1	1
		1	1	1
NB	ASD, Not ASD traits	0.8	0.5714	0.6666
		0.8888	0.96	0.9230
RF	ASD, Not ASD traits	1	1	1
		1	1	1
SVM	ASD, Not ASD traits	1	0.8571	0.9230
		0.9615	1	0.9803

Table 7
Evaluation results for ASD screening data for Adult.

Classifier	Class	Precision	recall	F1-score
DT	ASD, Not ASD traits	1	1	1
		1	1	1
K-NN	ASD, Not ASD traits	0.9440	0.9743	0.9589
		0.92	0.8363	0.8761
LR	ASD, Not ASD traits	1	1	1
		1	1	1
NB	ASD, Not ASD traits	0.9870	0.9807	0.9839
		0.9464	0.9636	0.9549
RF	ASD, Not ASD traits	1	1	1
		1	1	1
SVM	ASD, Not ASD traits	0.9397	1	0.9689
		1	0.8181	0.9

The number of accurate predictions made across all accurate samples is known as recall. It is calculated using (2):

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

where FN refers to the number of incorrect negative predictions (False Negative).

The harmonic means of Recall and Precision is the *F*1-*score*. The F1 score is a superior performance statistic for unbalanced data than the accuracy metric. This score is computed using (3).

$$\text{F1} - \text{score} = 2\text{ x } \frac{Precision\ x\ Recall}{Precision + Recall} \tag{3}$$

The experimental evaluation results of several ML models with the features selected for the utilized ASD screening data have been demonstrated and shown in Tables 4–7. In this, the Q-chat-10, gender, age, jaundice, and result score features are utilized in the evaluation process of training and testing the adopted ML models. The evaluation of various ML models over the four datasets notes that the dominant models over others are the DT, LR, and RF.

The classification report for the utilized classification models is illustrated in Table 8. From the obtained results, confusion matrix, all datasets except the child dataset suffered from the unbalancing distribution of the classes. Due to the reality that the tree-based algorithms are robust against unbalancing [39], the DT and RF obtain the highest accuracy, among others.

6. Comparison and conclusion

In this study, Autism Spectrum Disorder (ASD) was detected utilizing multiple ML models on four publicly distinct non-clinically ASD screening datasets provided by the Kaggle and UCI machine learning repository. These datasets are linked to four age groups, toddlers, children, adolescents, and adults. Several performance assessment criteria were utilized to examine the performance of the models constructed for ASD identification.

The comparison results for the comparing process with other recent studies [12,21,22] on this problem show that our models obtain better over all the utilized classifiers after addressing the missing values in the Toddler's autism spectrum compared to the study [24].

As for the study [21], a better result was obtained for the LR, and NB classifiers, while a better result was obtained for the K-NN, LR, and SVM classifiers in improving and treating the missing values compared to the study [22]. The Comparison results with [21,22 and [14] are illustrated in Tables 9 and 10, respectively.

Credit author statement

Dhuha Dheyaa Khudhur: Paper written, related work, Simultion, out put figures. **Saja Dheyaa Khudhur**: Referencess arangement, English editing.

Declaration of competing interest

The authors declare that they have no known competing financial

Table 8
Classification report of the classification models with ASD screening data for the four datasets.

		DT	LR	SVM	NB	K-NN	REF
Child Dataset	Accuracy	100%	100%	100%	94.2%	97.7%	100%
	Confusion Matrix	$\begin{bmatrix}42 & 0\\0 & 45\end{bmatrix}$	$\begin{bmatrix}42 & 0\\0 & 45\end{bmatrix}$	$\begin{bmatrix}42 & 0\\0 & 45\end{bmatrix}$	$\begin{bmatrix}42 & 0\\5 & 40\end{bmatrix}$	$\begin{bmatrix}42 & 0\\2 & 43\end{bmatrix}$	$\begin{bmatrix}42 & 0\\0 & 45\end{bmatrix}$
Adult dataset	Accuracy	100%	100%	95.2%	97.6%	93.8%	100%
	Confusion Matrix	$\begin{bmatrix}156 & 0\\0 & 55\end{bmatrix}$	$\begin{bmatrix}156 & 0\\0 & 55\end{bmatrix}$	$\begin{bmatrix}156 & 0\\10 & 45\end{bmatrix}$	$\begin{bmatrix}153 & 3\\2 & 53\end{bmatrix}$	$\begin{bmatrix}152 & 4\\9 & 46\end{bmatrix}$	$\begin{bmatrix}156 & 0\\0 & 55\end{bmatrix}$
Adolescent dataset	Accuracy	100%	100%	96.8%	87.5%	96.8%	100%
	Confusion Matrix	$\begin{bmatrix}7 & 0\\0 & 25\end{bmatrix}$	$\begin{bmatrix}7 & 0\\0 & 25\end{bmatrix}$	$\begin{bmatrix}6 & 1\\0 & 25\end{bmatrix}$	$\begin{bmatrix}4 & 3\\1 & 24\end{bmatrix}$	$\begin{bmatrix}6 & 1\\0 & 25\end{bmatrix}$	$\begin{bmatrix}7 & 0\\0 & 25\end{bmatrix}$
Toddler dataset	Accuracy	100%	100%	96.8%	98.1%	95.7%	100%
	Confusion Matrix	$\begin{bmatrix}62 & 0\\0 & 149\end{bmatrix}$	$\begin{bmatrix}62 & 0\\0 & 149\end{bmatrix}$	$\begin{bmatrix}59 & 3\\1 & 148\end{bmatrix}$	$\begin{bmatrix}59 & 3\\1 & 148\end{bmatrix}$	$\begin{bmatrix}59 & 3\\6 & 143\end{bmatrix}$	$\begin{bmatrix}62 & 0\\0 & 149\end{bmatrix}$

Table 9
Comparison results with [21,22] on ASD screening data for Children, Adolescents, and Adults datasets.

		Children			Adolescents			Adults		
		[21]	[22]	Our work	[21]	[22]	Our work	[21]	[22]	Our work
Classifier	DT	**100**	–	**100**	**100**	–	**100**	**100**	–	**100**
	K-NN	–	88.13	**97.7**	–	80.95	**96.8**	–	**95.75**	93.8
	LR	97	98.30	**100**	88	85.71	**100**	98	96.69	**100**
	NB	73	**94.91**	94.2	66	**90.47**	87.5	34	96.22	**97.6**
	RF	**100**	–	**100**	**100**	–	**100**	**100**	–	**100**
	SVM	–	98.30	**100**	–	95.23	**96.8**	–	**98.11**	95.2

Table 10
comparison results [12] on ASD screening data for the Toddlers dataset.

		Toddlers	
		[12]	Our work
Classifier	DT	**-**	**100**
	K-NN	90.52	**95.7**
	LR	97.15	**100**
	NB	94.79	**98.1**
	RF	81.52	**100**
	SVM	93.84	**96.8**

interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] J. Kang, X. Han, J. Song, Z. Niu, X. Li, The identification of children with autism spectrum disorder by SVM approach on EEG and eye-tracking data, Comput. Biol. Med. 120 (2020), 103722, 103722.

[2] V.S. Padala, K. Gandhi, P. Dasari, Machine learning: the new language for applications, IAES Int. J. Artif. Intell. 8 (4) (2019) 411.

[3] D. Eman, A.W.R. Emanuel, Machine learning classifiers for autism spectrum disorder: a review, in: 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2019.

[4] F. Thabtah, N. Abdelhamid, D. Peebles, A machine learning autism classification based on logistic regression analysis, Health Inf. Sci. Syst. 7 (2019) 12, https://doi.org/10.1007/s13755-019-0073-5.

[5] K.D. Cantin-Garside, Z. Kong, S.W. White, L. Antezana, S. Kim, M.A. Nussbaum, Detecting and classifying self-injurious behavior in autism spectrum disorder using machine learning techniques, J. Autism Dev. Disord. 50 (11) (2020) 4039–4052.

[6] N.A. Mashudi, N. Ahmad, N.M. Noor, Classification of adult autistic spectrum disorder using machine learning approach, IAES Int. J. Artif. Intell. 10 (3) (September 2021) 743–751, https://doi.org/10.11591/ijai.v10.i3.pp743-751.

[7] U. Erkan, D.N.H. Thanh, Autism Spectrum Disorder detection with machine learning methods, Current Psychiatry Research and Reviews 15 (4) (2019) 297–308.

[8] K. Shahrukh Omar, P. Mondal, N. Shahnaz Khan, R. Karim Rizvi, N. Islam, A machine learning approach to predict autism spectrum disorder, in: Proceedings of the 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), IEEE, Cox'sBazar, Bangladesh, Feb 2019, pp. 1–6.

[9] M. Panda, D.P. Mishra, S.M. Patro, S.R. Salkuti, Prediction of diabetes disease using machine learning algorithms, IAES Int. J. Artif. Intell. 11 (1) (2022) 284.

[10] G. Murat, A novel machine learning model to predict autism spectrum disorders risk gene, Neural Comput. Appl. 31 (10) (2019) 6711–6717.

[11] M.K. Hanif, N. Ashraf, M.U. Sarwar, D.M. Adinew, R. Yaqoob, Employing machine learning-based predictive analytical approaches to classify autism spectrum disorder types, Complexity 2022 (2022) 1–10.

[12] K. Vakadkar, D. Purkayastha, D. Krishnan, Detection of autism Spectrum Disorder in children using machine learning techniques, SN Comput Sci 2 (5) (2021) 386.

[13] F. Abdali-Mohammadi, M.N. Meqdad, S. Kadry, Development of an IoT-based and cloud-based disease prediction and diagnosis system for healthcare using machine learning algorithms, IAES Int. J. Artif. Intell. 9 (4) (2020) 766.

[14] Y. Karunakaran, Babiker hamdan, and sathish, "early prediction of autism spectrum disorder by computational approaches to fMRI analysis with early learning technique, December 2 (4) (2020) 207–216, 2020.

[15] S.D. Khudhur, D.D. Khudhur, IgG-IgM antibodies based infection time detection of COVID-19 using machine learning models, TELKOMNIKA 20 (2) (Apr. 2022) 340.

[16] N.A. Ali, Autism spectrum disorder classification on electroencephalogram signal using deep learning algorithm, IAES Int. J. Artif. Intell. 9 (1) (2020) 91.

[17] M.S. Croock, S.D. Khuder, A.E. Korial, S.S. Mahmood, Early detection of breast cancer using mammography images and software engineering process, Telkomnika (Telecommunication Computing Electronics and Control) 18 (4) (2020).

[18] A.A. Abdullah, S. Rijal, S.R. Dash, Evaluation on machine learning algorithms for classification of Autism Spectrum Disorder (ASD), J. Phys. Conf. Ser. 1372 (1) (2019), 012052.

[19] R. Vaishali, R. Sasikala, A machine learning based approach to classify Autism with optimum behavior sets, Int. J. Eng. Technol. 7 (4) (2018) 1–6.

[20] K. Chowdhury, 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.

[21] M. Masum, I. Nur, M.J. Hossain Faruk, M. Adnan, H. Shahriar, A Comparative Study of Machine Learning-Based Autism Spectrum Disorder Detection with Feature Importance Analysis, 2022, p. 3 [Online].Available: https://www.researchgate.net/publication/357861486.

[22] S. Raj, S. Masood, Analysis and detection of autism spectrum disorder using machine learning techniques, Procedia Comput. Sci. 167 (2020) 994–1004, https://doi.org/10.1016/j.procs.2020.03.399.

[23] M.M. Nishat, et al., Detection of autism spectrum disorder by discriminant analysis algorithm, in: Lecture Notes on Data Engineering and Communications Technologies, Springer Singapore, Singapore, 2022, pp. 473–482.

[24] M.M. Rahman, O.L. Usman, R.C. Muniyandi, S. Sahran, S. Mohamed, R.A. Razak, A review of machine learning methods of feature selection and classification for autism Spectrum Disorder, Brain Sci. 10 (12) (2020) 949.

[25] M.D. Hossain, M.A. Kabir, A. Anwar, M.Z. Islam, Detecting autism spectrum disorder using machine learning techniques, Health Inf. Sci. Syst. 9 (1) (2021), https://doi.org/10.1007/s13755-021-00145-9.

[26] N.A. Mashudi, N. Ahmad, N.M. Noor, Classification of adult autistic spectrum disorder using machine learning approach, IAES Int. J. Artif. Intell. 10 (3) (September 2021) 743–751, https://doi.org/10.11591/ijai.v10.i3.pp743-751.

[27] M.H. Al Banna, T. Ghosh, K.A. Taher, M.S. Kaiser, M. Mahmud, A monitoring system for patients of autism spectrum disorder using artificial intelligence, in: Brain Informatics, Springer International Publishing, Cham, 2020, pp. 251–262.

[28] U.A. Siddiqui, et al., Wearable-sensors-based platform for gesture recognition of autism spectrum disorder children using machine learning algorithms, Sensors 21 (10) (2021) 3319.

[29] Fadi, Autism Screening for Toddlers, Kaggle, 2018 [Online]. Available: https://www.kaggle.com/fabdelja/autism-screening-for-toddlers. (Accessed 1 June 2022).

[30] Fadi Fayez Thabtah, Autistic spectrum disorder screening data for children. https://archive.ics.uci.edu/ml/machine-learningdatabases/00419/, 2017. (Accessed 2 June 2022), 2017.

[31] Fadi Fayez Thabtah, Autistic spectrum disorder screening data for adolescent. https://archive.ics.uci.edu/ml/machine-learningdatabases/00420/, 2017. (Accessed 30 May 2022).

[32] Fadi Fayez Thabtah, Autistic spectrum disorder screening data for adult. https://archive.ics.uci.edu/ml/machine-learningdatabases/00426/, 2017. (Accessed 29 May 2022).

[33] D. Mustafa Abdullah, A. Mohsin Abdulazeez, Machine learning applications based on SVM classification A review, Qubahan Academic Journal 1 (2) (Apr. 2021) 81–90.

[34] Y. Zheng, T. Deng, Y. Wang, Autism classification based on logistic regression model, in: IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering vol. 2021, ICBAIE, Mar. 2021.

[35] I.L. Cohen, M.J. Flory, Autism spectrum disorder decision tree subgroups predict adaptive behavior and autism severity trajectories in children with ASD, J. Autism Dev. Disord. 49 (4) (Dec. 2018) 1423–1437.

[36] R. Blanquero, E. Carrizosa, P. Ramírez-Cobo, M.R. Sillero-Denamiel, Variable selection for naïve Bayes classification, Comput. Oper. Res. 135 (105456) (2021), 105456.

[37] M.M. Haque, et al., Informing developmental milestone achievement for children with autism: machine learning approach, JMIR Med. Inform. 9 (6) (2021), e29242.

[38] N. Azmi, et al., RF-based moisture content determination in rice using machine learning techniques, Sensors 21 (5) (2021) 1875.

[39] S.D. Khudhur, H.A. Jeiad, A content-based file identification dataset: collection, construction, and evaluation, Karbala int. j. mod. sci. 8 (2) (2022) 63–70.